THE OPERATIONAL RESEARCH SOCIETY

The OR Society
10th Simulation Workshop

SW21

22-26 March 2021

Proceedings

EDITORS: **Dr Masoud Fakhimi,** University of Surrey
Dr Tom Boness, Operational Research in Health Ltd
Dr Duncan Robertson, Loughborough University

In association with

informs. SIMULATION

Publication kindly sponsored by

Headline sponsor:
dse, anylogic, anyLogistix

Premium sponsors:
Loughborough University, SIMUL8, Taylor & Francis

Sponsors:
dstl The Science Inside, Saker Solutions Part of the ITI Group

INTRODUCTION TO THE WORKSHOP

Welcome to the 2021 Operational Research Society Simulation Workshop (SW21)

This year, we meet again at the 10th Simulation Workshop. Due to the Coronavirus pandemic, SW20 was postponed to 2021 and we have had to move to an online format. As a result, we have scheduled the programme across five, shorter days. We have kept all the successful features of previous conferences and have added some new features which we hope will contribute to a successful virtual conference experience.

The programme includes a range of exciting talks, panels, tutorials and more, bringing you the latest updates in the simulation field and research. There will also be social activities for delegates to interact during the networking breaks and at the end of the conference day such as Virtual Pub Quiz, Relaxation and Meditation sessions to name a few. There will be a group modelling competition with prizes to be awarded to the winning team(s) and for the first time we have introduced a best paper award. Also, the editors of the Journal of Simulation (Professors Christine Currie, John Fowler and Loo Hey) will give advice on how to write winning simulation papers and Professor Sally Brailsford will discuss diversity in the simulation community. We hope that these, combined with our keynotes, tutorials, contributed papers and posters will provide a stimulating programme.

To mark the tenth anniversary of the Simulation Workshop, a paper by Professors Stewart Robinson and Simon J.E. Taylor, the founders of the simulation workshop is included in the proceedings. The paper narrates the history of the simulation workshop. We hope you will enjoy reading it.

The conference programme includes 35 contributed papers, 15 posters, 7 tutorials, a panel discussion and 2 keynote speakers on a range of simulation topics, including simulation for Covid-19.

We are delighted to welcome Professor Susan M. Sanchez from the Naval Postgraduate School and Professor Young-Jun Son from the University of Arizona as our keynote speakers. Susan will be discussing data farming methods and opportunities and challenges for further research in the area. Son will be talking about a planning and control framework for effective and efficient surveillance and crowd control based on dynamic data-driven adaptive multi-scale simulation.

There will be a panel session on the final day of the workshop. This year, the panel discussion will focus on the relationship between artificial intelligence and simulation, a topic that will no doubt generate a lively debate.

We continue the successful tutorials. Prestigious names in the field will talk about verification and validation, optimisation, agent-based simulation, system dynamics, hybrid simulation and facilitated simulation and text analytics for simulation.

We also continue the poster competition. In the Lightning Poster Plenary, each poster delegate is given 2 minutes to pitch their poster. Delegates will have the opportunity to discuss the poster during the dedicated poster sessions.

Following the workshop, there will be a special issue in the Journal of Simulation (JoS) dedicated to work submitted and presented at the conference. We invite SW21 authors, both academics and practitioners, to extend their work and submit to the SW21 special issue. The deadline for submission of papers is June 2021. So far JOS has published an exciting range of papers with a focus on the practice and application of simulation. OR Society members have free access to JOS online. For those with institutional libraries, do remember to ask the library to subscribe to JOS. We would encourage you all to think about submitting a paper to JOS. Even if you cannot make the deadline for the special issue, why not revise and extend your SW21 paper for a regular JOS issue?

We would like to thank every single person that supported us in putting together SW21, including the authors and speakers, the sponsors, session chairs and reviewers. We also thank those, whose hard work has made SW21 possible, especially The OR Society Events Team, Dr Masoud Fakhimi, Dr Tom Bonnes, Dr Duncan Robertson (Programme Chairs), Dr Lucy Morgan (Poster Chair), Dr Durk-Jouke van der Zee (Publicity Chair), Dr Christina Philips (Social Media Chair) and Professor John Fowler (International Liaison Chair). Finally, we are very grateful for the guidance and advice that Professor Stewart Robinson and Professor Simon J.E. Taylor have given us in organising the workshop.

Take care, stay safe and enjoy the conference!

Anastasia Anagnostou and Antuela Tako

Conference Chairs

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.

CONTENTS

Contents

Contents

Agent Based Simulation

Conceptual Modelling

Hybrid Simulation

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.001

CELEBRATING 20 YEARS: THE STORY OF THE SIMULATION WORKSHOP

Prof. Stewart Robinson

School of Business and Economics
Loughborough University
Loughborough, LE11 3TU
s.l.robinson@lboro.ac.uk

Prof. Simon J.E. Taylor

Department of Computer Science
Brunel University London
Uxbridge, UB8 3PH
simon.taylor@brunel.ac.uk

ABSTRACT

This year marks the 10th Operational Research Society Simulation Workshop (SW). It is over 20 years since the planning commenced for the very first SW which took place in Birmingham in March 2002. To mark this occasion, this article presents a short history of the conference including the events that led up to the first conference, details of the first SW, a summary of all 10 SWs, and how all this led to the creation of the Journal of Simulation. As founders of the conference series we provide our personal reflections on these events.

Keywords: Simulation Workshop, Conference, History

1 INTRODUCTION

March 2021 sees the 10th Simulation Workshop. This, of course, has been delayed by a year due to the Covid-19 pandemic. It is, in fact, over 20 years since the planning for the inaugural workshop in 2002 started. We set-up this series of conferences following some years of running successful events through the Operational Research Society's Simulation Study Group. More correctly, we ventured to run a conference in 2002, and then one in 2004, and before we knew it the run of conferences had become a series.

On the occasion of the 10th Simulation Workshop this paper sets out the beginnings of the conference, its history and our guiding philosophy for these events. There have been some light hearted and challenging moments along the way. What we hope to show is that two early career academics, if they put their minds to it, can create something our of nothing. As athletes often say on winning an event, 'if I can do it, anyone can.' We would encourage our colleagues to think in the same way and simply have a go. Doing so even helped us become senior academics.

In the paper we set out how we teamed-up to lead the Simulation Study Group, the developments that led to the first Simulation Workshop, the key facts, figures and events over the last 20 years, and how all this led to the Journal of Simulation. Of course, it is quite possible that we have omitted some things of importance, or misreported some of the events that took place; all such errors are probably a result of our age and certainly not deliberate attempts to misrepresent history!

2 THE BEGINNINGS: THE SIMULATION STUDY GROUP

The story of the Simulation Workshop started for us sometime around 1996. Stewart was a junior member of staff, indeed a Teaching Fellow on a temporary contract, at Aston University. In his annual review with the Head of the Operations and Information Management Group, Professor Colin Lewis asked him if he was involved with the OR Society's Simulation Study Group. To Stewart's shame he admitted that he had never heard of the group. Professor Lewis pointed him to the inside cover of the Journal of the Operational Research Society where it not only told him of the existence of the group, but revealed that it was being led by a certain Ray Paul and David Balmer. Somewhat embarrassed by his lack of knowledge of both the activities of 'his' Society and his chosen research field, he emailed Professor Paul to ask about the group. (At this point Stewart had known Ray for around four years,

but only ever met him on US soil.) Stewart had expected to find the Study Group was a hive of activity that he had been missing out on and that it was about to transform his career. To his surprise, he discovered that the group had not met for some years. At least that explained his ignorance of the existence of the group.

Ray went on to say that the group need new leadership from someone ‘like you’. For those who have worked with Ray, you will know that it is very hard to say ‘no’ when he makes a statement like that. Indeed, he went on to say that he had just employed a new lecturer called Simon Taylor and he thought we would do an excellent job of running the Simulation Study Group together. Simon had been a Research Fellow in the Centre for Parallel Computing at the University of Westminster working on Distributed and High Performance Simulation. Ray had had a similar conversation with him about Stewart with the ‘like you’ and ‘no’ options! Neither had any idea about who each other was, but Ray’s convincing arguments sounded reasonable… and that is when we (Stewart and Simon) started to work together.

One of our earliest discussions with each other about the study group took place at the Winter Simulation Conference in San Diego in December 1996. We visited a bar in downtown San Diego, discussed ideas for the group and rather than forget them, wrote them on a menu we ‘borrowed’ from the bar. That menu (shown in figure 1) guided our thoughts over the next couple of years. Our other memories of that evening were that it was probably the last time Stewart was asked to prove I was over 21 by a bar attendant and having to beat a hasty retreat onto a San Diego tram as a man with a large stick was threatening us!

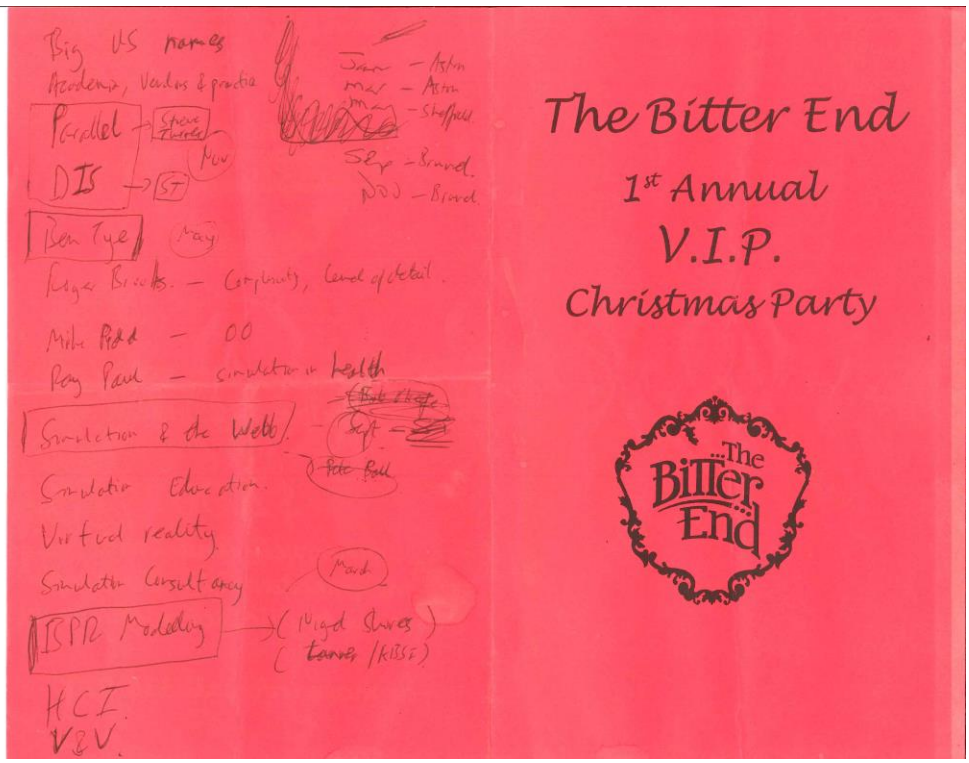


Figure 1 *Study Group Plans on a San Diego Bar Menu*

We organised the first of the new Simulation Study Group meetings on 29 January 1997 at Aston University in Birmingham with two speakers: John Salt and Stewart Robinson. The title of the event was ‘Simulation Should be Simple and Fun: Some Do's and Don'ts of Simulation’, which borrowed heavily from John’s keynote talk at the 1993 Winter Simulation Conference. To our surprise around 20-30 people turned out for a two hour early evening event. One notable attendee was Sally Brailsford who had travelled from Southampton University; this was the first time that we had met her. We enthusiastically went on to organise five further meetings during 1997, although one did not run because of poor attendance.

One thing we learnt from these meetings was that attendance at short one or two speaker events was generally quite low. So in 1998 we changed the approach and moved to a format of running two, one day meetings a year; something we continued over the next few years. The first such event was held on 24 June 1998 at Brunel University under the title 'Simulation Software: Present and Future'. The five speakers for the day were: Vlatka Hlupic (Brunel University), Steve White (British Airways), Mike Pidd (Lancaster University), Tony Waller (Lanner Group) and Ray Paul (Brunel University). Attendance at these one-day meetings was generally between 30 and 40 with the record attendance of over 70 being at a joint meeting with the UK Simulation Society organised by Susan Howick from the University of Strathclyde in April 2000: 'Discrete Event Simulation and System Dynamics: Never the Twain Shall Meet?' The paradox being that modellers from both camps did actually meet in what was an early foray into comparing these two worlds and how they might work together.

A particular facet of the study group was the involvement of both academics and practitioners. Indeed, many of the meetings included speakers from both worlds. To note that due to a change in OR Society naming conventions, the Simulation Study Group was renamed the Simulation Special Interest Group; the name it still carries today.

3 THE FIRST SIMULATION WORKSHOP (2002)

It was around the turn of the century when Brian Lehaney, then Chair of the OR Society's Events Committee, suggested to us that we should run a simulation conference. This was presumably as a result of the successful Study Group events we had been running. Stewart's first reaction was to say 'no'. Simon's was 'yes!!!' After all, there is a huge difference between running occasional one day events, for which there was no charge, and running a full blown conference. Stewart cannot remember what persuaded him to go ahead with the conference, but he suspects that it was Simon's more optimistic outlook on life! So somewhere early in the year 2000 we decided we would give it a go. And so what became known as the 'Simulation Study Group Two-Day Workshop', or 'Simulation Workshop' for short, was scheduled for 20-21 March 2002. The front cover of the proceedings for the original event is shown in Figure 2.

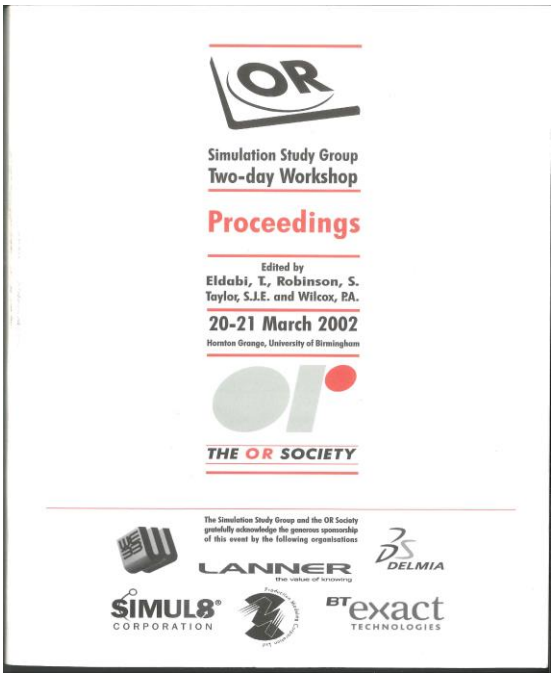


Figure 2 *Proceedings Cover for the First Simulation Workshop*

We had some discussion about whether to call the event a conference, but in the end we went for the 'workshop' title. This reflected our view that the event should create an environment where academics and practitioners, from those just starting out to the well-established, shared and discussed ideas about simulation.

Over the two years of planning for the 2002 conference we developed a number of guiding principles that set the tone for future Simulation Workshops. First, we were insistent that the conference should be held in hotel type accommodation and not using student halls. If we wanted practitioners, senior academics and international visitors to attend, we knew they would expect a better quality of accommodation. We had also been tainted by an earlier experience of a conference using student accommodation; Simon woke to find a faulty spring had caused his mattress to attach itself to his knee! Stewart had slept on a similarly uncomfortable bed in the same facility, but had managed to avoid being physically assaulted by his mattress!

Second, we wanted to hold the conference in a facility that was largely, if not fully, dedicated to our delegates. We wanted people to know that if they spoke to someone else at dinner or in the bar, they were almost certainly attending the Simulation Workshop as well. The focus should be on an intensive two days for sharing and discussing ideas.

Third, we decided that all presentations should have published and reviewed papers associated with them (although there were two presentation only papers at the first conference). In doing so we followed the practice of the Winter Simulation Conference at the time. It had the benefit of helping to assure the quality of the papers presented and created an archival record of the conference. For the academics it had the added advantage of developing a first draft of a future journal paper. This requirement was relaxed a little for later Simulation Workshops when 'practice papers' were introduced; these only needed to be three pages in length. For more recent conferences this has been relaxed further and the Simulation Workshop does now have some presentation only papers. Of course, the technology for publishing the proceedings has changed over the years, moving from paper only, to CD and recently to a memory stick.

Fourth, we wanted an international keynote speaker. For 2002 Professor Richard Nance from Virginia Tech agreed to give the keynote address. Professor Nance, a leading international figure in simulation research spoke on the 'Simulation Research Agenda: Hope, Hype and Hyperbole'. In doing so he identified some of the grand challenges facing simulation at the time.

Fifth, we wanted strong representation from the simulation practitioner community. One way in which we aimed to achieve this was to invite companies to sponsor the event for a relatively low fee. In doing so they were highlighted in the conference advertising, documentation and proceedings. They were also given space in the exhibition area during the conference. Something which has become a focal point for gathering during refreshment breaks. Although we do not have detailed evidence our sense is that we have achieved a close to 50:50 split of academic and practitioner attendance.

Finally, we wanted to keep the registration fee to below £300. Something we maintained for many years and only recently has it crept above this level due, of course, to the ongoing impact of inflation.

And so we launched the workshop in the spring of 2001 with an announcement and a call for papers. Everything had to be designed from scratch: the call for papers, the author instructions, the review process, the proceedings layout, the programme for the conference, the planning timetable, etc. We also had to find a location for the event. With some guidance from Chris Barrett, the OR Society's conference organiser, we chose Hornton Grange on the edge of the Birmingham University campus. It provided good quality en-suite accommodation and a dedicated conference facility which could provide an exhibition/refreshment area and two rooms for presentations.

We arrived at Hornton Grange on 20th March 2002 with no clue as to how well the next two days would progress. My main memory of that first conference was the opening session. We had 64 expectant delegates (we think) in the main meeting room waiting for the conference to start. Just before we opened the proceedings we remember a distinct air of 'what on earth have we done!'

4 TEN SIMULATION WORKSHOPS: 2002-2021

Suffice it to say that the 2002 'Simulation Study Group Two-Day Workshop' seemed to go rather well. So much so that the delegates suggested we should do it again. On further investigation the

general view was that doing this on a biennial (every two years) basis would be better than doing it every year; partially to keep our sanity in organising the conference, but also to make it more of an event and to ensure the world of simulation had moved sufficiently forward between conferences.

After a six month break, we went ahead and started planning for a 2004 conference, again to be held at Hornton Grange. The '2004 Operational Research Society Simulation Workshop' used the shortened title of SW04 for the first time; a naming convention that has stuck ever since.

Table 1 *Summary of Simulation Workshop Conferences*

Year	Location	Chair(s)	Key committee roles	Papers	Attendance
2002	Hornton Grange, Birmingham	Stewart Robinson Simon Taylor	Tillal Eldabi, Pauline Wilcox	26	63
2004	Hornton Grange, Birmingham	Stewart Robinson Simon Taylor	Les Oakshott, Sally Brailsford	37	85
2006	Ashorne Hill Conference Centre, Royal Leamington Spa	Stewart Robinson Simon Taylor	Sally Brailsford, Jeremy Garnett	33	76
2008	Abbey Hotel, Redditch	Stewart Robinson Simon Taylor	Kathy Kotiadis, Christine Currie	34	84
2010	Abbey Hotel, Redditch	Stewart Robinson Simon Taylor	Murat Gunal, Benny Tjahjono, Sally Brailsford, Antuela Tako	31	73
2012	Abbey Hotel, Redditch	Benny Tjahjono	Cathal Heavy, Stephan Onggo, Durk-Jouke van der Zee, Thomas Monks	32	74
2014	Abbey Hotel, Redditch	Benny Tjahjono	Cathal Heavy, Stephan Onggo, Durk-Jouke van der Zee, Thomas Monks	28	60
2016	Ettington Chase Hotel, Warwickshire	Tom Monks Christine Currie	Anastasia Anagnostou, Katy Hoad, Martin Kunc, Anastasia Gogi	28	71
2018	Ettington Chase Hotel, Warwickshire	Tom Monks Christine Currie	Anastasia Anagnostou, Rudabeh Meskarian, Duncan Robertson, Masoud Fakhimi, Tom Boness	21	69
2021*	Burleigh Court, Loughborough	Anastasia Anagnostou Antuela Tako	Masoud Fakhimi, Duncan Robertson, Tom Boness, Lucy Morgan, Durk-Jouke van der Zee, John Fowler	n/a	n/a

** The 10th Simulation Workshop was originally planned for March 2020, but the UK went into lockdown due to the Covid-19 pandemic a matter of a few days before the event was scheduled. As a result, the conference was postponed and then rescheduled as a virtual event for March 2021.*

SW04 still holds the record for the best attended conference, with 85 delegates, although this was nearly beaten in 2018 when 84 delegates registered for SW18. Indeed, this was the only time that we had to resort to running three parallel sessions during part of the conference in order to fit in all the presentations over the two days. The number of delegates presented us with a problem as the main conference room at Hornton Grange was not large enough to seat all the attendees. We may have sneaked some extra chairs into the conference room! It was for this reason that we moved the 2006

conference to a different location (Ashorne Hill Conference Centre near Royal Leamington Spa). SW04 was the only conference at which we did not have a panel discussion; instead we opted for a plenary session given by Brian Hollocks entitled 'Still simulating after all these years. Reflections on 40 years in simulation.' Brian submitted this as a standard paper, but we thought it was of sufficient interest to be delivered as a plenary, and it proved to be so.

Tables 1-3 summarise the details of every SIMULATION WORKSHOP conference from 2002 to 2021. Table 1 provides details of the conference location, chairs, committee members, the number of papers and attendance. Table 2 lists the keynote talks and table 3 gives details of the panel discussions at each conference.

Simon and Stewart chaired the first five conferences, eventually handing the reins over to Benny Tjahjono who, at the time, was at Cranfield University. Benny ably chaired SW12 and SW14 after which Christine Currie and Tom Monks co-chaired the next two conferences. Anastasia Anagnostou and Antuela Tako took over for this one, SW21. Key committee roles covered the delivery of the programme and proceedings, publicity and bringing together the poster session. Although the posters were initially aimed at encouraging PhD student participation, they have become an opportunity for other delegates to display their ideas before they are ready for a full paper. More recently there has been a prize for the best poster.

The number of papers has varied over the years, but has remained between 20 and below 40. More recent conferences have had more keynote and plenary events, as well as tutorial sessions on the day before the main conference starts. Attendance has varied from 60 to 85. One notable feature is that the number of attendees has always been at least double the number of papers, which suggests there is very high interest in the conference without attendees feeling the need to make a presentation.

Table 2 *Keynote Talks*

Year	Keynote Speaker	Title
2002	Richard Nance (Virginia Tech)	The simulation research agenda: hope, hype and hyperbole (or whence, wherefore and whither?)
2004	Paul Fishwick, University of Florida	Modelling: taking it to the next level
2006	John Morecroft (London Business School)*	Representation and simulation – an information feedback view
2008	Brian Hollocks (Bournemouth University)	Intelligence, innovation & integrity – K D Tocher and the dawn of simulation
2010	Charles Macal (Argonne National Laboratory)	The future of agent-based modeling and simulation
2012	Shane Henderson (Cornell University)	Real-time ambulance-fleet control via an amalgam of simulation, optimization, and statistics
2014	Barry Nelson (Northwestern University)	Why good simulations go bad
2016 (a)	Alexander Verbraeck (Delft University of Technology)	Data driven simulation
2016 (b)	Sally Brailsford (University of Southampton)	Hybrid simulation: the best thing since sliced bread, or just a fad?
2018 (a)	John Fowler (Arizona State University)	Personal reflections on the evolution of simulation over the last 20 years
2018 (b)	Russell Cheng (University of Southampton)	Visual representation of simulation results
2021 (a)	Young-Jun Son (University of Arizona)	A DDDAMS-based surveillance and crowd control via UAVs and UGVs
2021 (b)	Susan Sanchez (Naval Postgraduate School, Monterey)	Data farming: the meaning and methods behind the metaphor

* Stephen Chick (INSEAD) was originally lined-up as keynote speaker but could not join us due to injury.

From the beginning we sought to bring in a keynote speaker from overseas. This was not only beneficial in terms of bringing in some of the leading researchers in the field of simulation, but also a means of connecting them with the UK simulation community. We are very grateful to the willingness of these speakers to participate in and support the SIMULATION WORKSHOP conferences. They have certainly brought to our attention some current and exciting topics in simulation.

Table 3 *Panel Discussions*

Year	Chair	Panellists	Topic
2002	Stewart Robinson	Dick Nance Mike Pidd Ray Paul Simon Taylor	Model reuse
2004	*No panel discussion*		
2006	Mike Pidd	Not recorded	Simulation in health
2008	Michael Pidd	Stephen Chick Mark Elder Shane Kite Ray Paul	Simulation optimisation: the best thing since sliced bread
2010	Michael Pidd	Peer Olaf Siebers Charles Macal Jeremy Garnett Dave Buxton	DES is dead, long live ABS!
2012	Sally Brailsford	Stewart Robinson Shane Henderson Claire Cordeaux Shane Kite	The practice of simulation: useful, in theory? The theory of simulation: practically useless?
2014	Simon Taylor	Barry Nelson Mark Elder Ken McNaught Christine Currie	Simulation analytics: the future of simulation?
2016	Christine Currie	Stewart Robinson Simon Taylor John Fowler Sally Brailsford	Celebrating 10 years of the Journal of Simulation
2018 (a)	Kathy Kotiadis	Sally Brailsford Antuela Tako Stewart Robinson Christina Phillips Mark Elder	Discussing the challenges of stakeholder involvement and how to overcome them
2018 (b)	Peer Olaf Siebers	Peer Olaf Siebers Antuela Tako Dave Buxton Tom Monks Kim Warren	Model development strategies: from a copy/paste mentality to truly innovative approaches
2021	Simon J E Taylor	TBC	Artificial Intelligence and Simulation: Friend or Foe?

In terms of keynotes, 2006 was probably the most memorable from a conference organisers point of view. Stephen Chick (INSEAD) was originally lined-up as the keynote speaker. However, he contacted us only a couple of weeks before the conference, very apologetically, to say that he had broken his leg playing ice hockey and so was unable to attend. John Morecroft stood in at the very last minute and gave a very memorable talk on system dynamics. We remain grateful to John for rescuing

us and also to Steve who subsequently wrote-up and published his never delivered keynote address as a paper in the inaugural issue of the Journal of Simulation (Chick, 2006).

With the exception of 2004, we have always held a panel discussion on the second day of the conference. These have led to lively debates on current issues in simulation. One of the most memorable was in 2010 following Charles Macal's keynote on agent-based simulation (ABS). The title of the discussion was 'DES [discrete-event simulation] is dead, long live ABS!' In response to the proposition that DES was no longer worthwhile, Sally Brailsford later responded with her paper 'Discrete-event simulation is alive and kicking!' (Brailsford, 2014).

The Simulation Workshop started as a DES conference, but has increasingly encompassed system dynamics, ABS and hybrid simulation. The keynotes from John Morecroft (2006) and Charles Macal (2010) had a significant impact in introducing the DES community to these alternative simulation approaches.

We used the 2008 conference to celebrate the 50th anniversary of the first simulation software. The General Simulation Program (GSP) was developed by K D Tocher at the United Steel Companies in the UK in 1958. Having worked with Tocher, we asked Brian Hollocks to give the keynote address; an excellent history of Tocher's contribution to the simulation field. A highlight was having Tocher's widow, Charlotte, and their two children in attendance. It took Stewart six months to track the family down. Charlotte, then aged 90, gave a memorable speech about her husband's work. It included a story of the press being very excited to interview Tocher because of his work on 'computer stimulation!' At SW08 we announced the launch of the K D Tocher Medal for the best paper in the Journal of Simulation, making the inaugural award in 2010. Charlotte returned to SW10 to give the award to Kathy Kotiadis.

We have always sought to ensure the conference gains international recognition. Apart from international keynote speakers, we have benefitted from international participants from Europe, Asia, the USA and as far away as Australia. Over time we have also gained 'in-cooperation' agreements from the following international societies: the INFORMS Simulation Society, ACM SIGSIM and the Society for Modeling and Simulation International.

Finally, in the long line of memorable happenings, we never expected a Simulation Workshop (SW20) to get postponed due to a global pandemic and for the conference to go fully on-line as it will in 2021. Certainly the technology would not have been available in the early years to support such a virtual event. Thanks to the dedication of the conference team it is great to know that the 'show will go on.'

5 AND SO TO THE JOURNAL OF SIMULATION

The story of the Journal of Simulation is closely aligned with the Simulation Workshop and a late night conversation with Ray Paul. During SW04 Ray suggested that, based on the success of the conference, the growing UK simulation community and the need for more journals on simulation, Simon and Stewart should start-up a new Operational Research Society journal. Our reactions were 'no' and 'yes!!!' (the reader can guess who said what!) Stewart's reservations were the huge difference between running a conference and setting-up and editing a journal from scratch. Stewart cannot remember what persuaded him to go ahead with the journal, but he suspects once again it was Simon's relentlessly optimistic attitude! (If you think you have read this before, you have, at the start of section 3 in relation to setting-up a conference.) This once again shows the impact that Ray Paul has had on us and on simulation in general.

And so the Journal of Simulation was born. After two-and-a-half years of negotiation and work, we launched the first issue of JOS at a special event in December 2006 during the Winter Simulation Conference in Monterey, California.

6 CONCLUSION

So 'what have we done?' First and foremost it is not what we have done; the success of the Simulation Workshop has been the result of many contributions from those that have helped organize the events, through key speakers, special guests, presenters, panelists, to attendees. One of the greatest facets of the Simulation Workshop has been the way that it brings a community of academics and

practitioners together every two years. As we had originally envisioned, the conference has a workshop feel, where participants share and discuss ideas freely. It has become the centre piece of a very active UK simulation community. But more than that, an opportunity to showcase that community's work to the wider world and for the wider world to input to the development of simulation in the UK.

We were once asked if we would like to see the conference become much bigger. Afterall, that would surely be a measure of its success. After reflecting on this for a while, we realized that the size of the conference (60-90 delegates) was the reason for its success. Attendees are pretty much able to meet and talk with every other attendee during the two days, which is the bedrock of creating a community. So growing larger may not signify success.

One thing we do not know is whether anyone has attended all ten conferences. Simon and Stewart have both missed one Simulation Workshop (Stewart in 2010 and Simon in 2012). What we do know is that it has been hard work, it has been rewarding and above all it has been fun!

ACKNOWLEDGMENTS

We would like to acknowledge the invaluable input of Professor Ray Paul. Without his insistence that we could actually do these things the Simulation Study Group, the Simulation Workshop conferences and the Journal of Simulation would never have existed. We would also like to thank all the teams involved with each successful Simulation Workshop. We also acknowledge the incredible support of Chris Barrett and her successor Hilary Wilkes as OR Society conference organisers (especially Hilary for delving into the ORS dungeons to help us to get a complete set of all the Simulation Workshop materials). Thanks for putting up with us!

REFERENCES

Brailsford S C (2014). Discrete-event simulation is alive and kicking! *Journal of Simulation* **8(1)**: 1-8.
Chick S E (2006). Six ways to improve a simulation analysis, *Journal of Simulation* **1(1)**: 21-28.

AUTHOR BIOGRAPHIES

STEWART ROBINSON is Professor of Management Science and Dean at the School of Business and Economics, Loughborough University, UK. His research interests are in the practice and use of simulation models where he focuses on conceptual modelling, model validation, output analysis and alternative simulation methods (discrete-event, system dynamics and agent based). His home page is www.stewartrobinson.co.uk.

SIMON J E TAYLOR is Professor of Computer Science, Vice Dean and Research Director of the Department of Computer Science, Brunel University London, UK. His research interests are in the advancement of academic and industrial simulation techniques and methods through novel digital technologies and in the development of digital infrastructures and simulation in Africa. He leads the Modelling & Simulation Group (modsimg.com). Find him at https://www.brunel.ac.uk/people/simon-taylor, https://orcid.org/0000-0001-8252-0189 and @simonatbrunel.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.002

DATA FARMING: THE MEANINGS AND METHODS BEHIND THE METAPHOR

Susan M. Sanchez

Naval Postgraduate School
Monterey, California, USA
ssanchez@nps.edu

ABSTRACT

Data farming captures the notion of purposeful data generation from simulation models. The ready availability of computing power has fundamentally changed the way simulation and other computational models can be used to provide insights to decision makers. Large-scale designed experiments let us grow the simulation output efficiently and effectively. We can explore massive input spaces, use statistical and visualization techniques to uncover interesting features of complex response surfaces, and explicitly identify cause-and-effect relationships. Nonetheless, there are many opportunities for research methods that could further enhance this process. I will begin with a brief overview of key differences between physical and simulation experiments, as well as current data farming capabilities and their relationship to emerging techniques in data science and analytics. I will then share some thoughts about opportunities and challenges for further improving the state of the art, and transforming the state of the practice, in this domain.

Keywords: Simulation, Experimental Design, Visualization

1 INTRODUCTION

We live in a world bombarded by data. The term ‘data mining’ is ubiquitous in the literature, while ‘data analytics’ and ‘data science’ have skyrocketed in popularity in recent years. Much of this digital dust is collected automatically—by our communication technology, sensors in the environment, cookies placed on websites, wireless devices comprising the internet of things, and more. Some of these observational data sources can be used to characterize input distributions for stochastic simulation models, either by fitting distributions from which pseudo-random numbers are generated, by bootstrapping samples from the empirical distributions, or used in data-driven simulation models to affect real-time system intervention and control. Yet this type of data is observational by nature, and so has limitations. Simulation output data, by contrast, is available only after the simulation is run, where a ‘data farming’ metaphor is more appropriate. Consider this description: “Real-world farmers cultivate the land to maximize their yield. They manipulate the environment to their advantage by using irrigation, pest control, crop rotation, fertilizer, and more. Small-scale designed experiments can help them to determine whether these treatments are effective. Similarly, data farmers manipulate simulation models to their advantage—but using large-scale designed experimentation. This allows them to learn more about the simulation model's behavior in a structured way. In this fashion, they ‘grow’ data from their models, but in a manner that facilitates identifying the useful information. For large-scale simulation experiments, this often results in data sets that, while big, are far smaller than what would be needed to gain insights if the results were observational (i.e., obtained using *ad hoc* or randomly generated combinations of factor settings). Data generated prospectively from designs is also better, in the sense that it lets us identify root cause-and-effect relationships between the simulation model input factors and the simulation output.” (Sanchez 2018).

Simulation is not the only community to use the data farming metaphor. As a noun, data farm may refer to a large bank of connected computers used to process and store data, host web services,

provide access for scientific computing, and more. As a verb, data farming has been used as a metaphor for dealing with big data in non-simulation contexts: see, e.g., Kusiak (2006) for enhancing industrial data for decision-support purposes, or Mayo et al. (2016) for improving patient outcomes in healthcare settings. These data farming approaches attempt to improve the collection, storage, maintenance, and retrieval of observational data so it is faster and easier to harvest insights. While some effort has been made to address causality from observational datasets (Pearl 2009), we can distinguish the simulation data farming view as one of generating and analyzing inferential big data, in contrast to methods for curating and analyzing observational big data.

Schruben (2017) asserts that "model is a verb" for simulation professionals. Likewise, data farming is a verb from the simulation perspective we use in this paper.

1.1 Background Terminology and Notation

Factors are inputs (or functions of inputs) to a simulation model that are purposefully varied at different levels when growing the data from a simulation experiment. An *experiment design* for k factors is an n by k matrix or table where each column specifies the levels or settings for a single factor and each row specifies the combination of factor settings to be used. We refer to the rows as *design points*, they might also be called runs or trials in other literature.

Features are characteristics of the *response surface* that maps the inputs to the simulation outputs. A statistical or analytical model of our simulation model's I/O behavior is called a *metamodel* because it is a model of a model. Many types of metamodels are possible, including partition trees (also known as classification and regression trees), multiple regression metamodels, logistic regression metamodels, Gaussian process metamodels, and more.

Flexibility is an important consideration when embarking on a data farming study because the types of designs used to grow the data will affect the types of metamodels we can fit, and the types of questions we can answer. In the data farming context, we are proponents of 'thinking big' in terms of the number and types of factors, the number of outputs and breadth of their response surface behaviors, and the types of analysis tools and methods that can be applied to the output data.

2 MEANINGS AND METHODS

In the rest of this paper, we will focus on data farming (the verb) as a metaphor for simulation studies. We will describe the meanings of several subcomponents of this metaphor, and present some practical data farming methods, with the goal of encouraging the readers to incorporate data farming into their future simulation studies.

2.1 Cross Fertilization

In our experience, data farming is most effective when it is a collaborative effort (NATO 2014). Stakeholders in the problem domain help ground the data farming effort and ensure that it does, in fact, address questions and provide insights that are useful and interesting to decision makers. Simulation modelers bring a variety of expertise. At early stages of a simulation study, their conceptual modelling skills may help scope the project so the simulation model is neither overly simplified nor overly complex for its intended purposes. They catch logical misconceptions that might invalidate the results or interpretation, such as a user who does not realize that different random number seeds lead to different results, or that modelling a queue as capacitated vs. uncapacitated will yield different results. Problem domain experts are key players in the model validation process.

One practical piece of advice is to ask all stakeholders to jot down a few key expectations, such as "What do believe the three most important factors will be? How will they affect the response?" Done early, this may lead to discussions that help frame the conceptual model and make sure there is a common understanding of its component, especially if the stakeholders have different backgrounds and expertise. Done before running the experiment, this helps ensure that the factors, their ranges or settings, and the experimental design used will be suitable for addressing the initial questions—although it is better to think of experimentation and analysis as an iterative process instead of a single event. Done after the data have been generated but before conducting analysis, this may help clarify

whether or not the results are surprising. Ultimately, when a surprising result is found, it should either lead to a bug being fixed (model verification) or intuition being changed (model validation).

2.2 Sowing the Seeds

Design of experiments (DOE) can be viewed as sowing the seeds for successful data farming, and brings tremendous capabilities to simulation studies. There are several reasons for this. First, experimentation is a straightforward way of establishing cause-and-effect. By purposefully varying factors using a good design, we can observe what (if any) effects they have on the responses—at least within the context of our simulation model. Varying multiple factors simultaneously is the only way to reveal interactions effects, varying factors at many levels in a space-filling design provides analysis flexibility, and using a good experimental design is absolutely required. What constitutes a bad design? A one-factor-at-a-time design is bad because it does not reveal any interactions. A design with high correlations among factors is bad design because it means that factor effects are confounded, so there is no unique way to determine which factors impact the response. A design that cannot be executed in the time required is a bad design because it means the decision maker cannot leverage insights from the study. A design that ignores factors simply to reduce the number of design points is bad because it drastically limits the potential insights that could be gained.

There are many good experimental designs, but some are more suitable for physical experiments or deterministic computer experiments than for stochastic simulation experiments. Here are a few that we recommend, use often, and are readily available for you to use in your next data farming experiment:

- Nearly orthogonal Latin hypercubes (NOLHs),
- Nearly orthogonal-and-balanced (NOAB) designs,
- Resolution V fractional factorials (R5FFs),
- Resolution V central composite designs (R5CCDs), and
- Frequency based designs (R5FBDs).

More details of these designs and their characteristics and applicability appear in the Appendix.

As a practical tip, follow the links in the Appendix to download the software and run a data farming experiment. The tutorial paper by Sanchez, Sanchez, and Wan (2020) discusses both design and analysis considerations in more depth.

2.3 Pest Control

In Section 2.1 we described how stakeholders' predictions of which factors will be most important can be helpful in verification and validation (VandV) efforts. A large-scale sensitivity analysis is a much broader and more rigorous way of stress-testing a simulation model.

This debugging effort also reinforces the view that model is a verb. We should not separate the process of modelling and experimentation, they enhance each other. It is better to continually experiment as you go along and build a model, catching at least some of the bugs earlier, than waiting until the end. Experimentation can also help the modeller avoid adding unnecessary model detail if it becomes clear that variation in certain model subcomponents is dampened by the system, so additional complexity is not warranted. For example, if varying a deterministic setup time for a station in a job shop between 15 minutes and 30 minutes does not yield a noticeable difference in overall throughput, then it would not be worthwhile to expend effort to create a stochastic setup time that varies over that same range.

At any stage, a practical way of proceeding is to begin with a baseline design point. Set the ranges for each quantitative factor a small percentage (say, 5% or 10%) above and below the baseline (if the baseline is at the lowest or highest level of interest, expand the range in only one direction). Run designed experiments regularly during the model-building process. This also means the model you're making will be data farmable, which will save you the time of having to restructure the finished model or create a data farming wrapper to facilitate experimentation. It also means you will easily identify situations where the model behaves strangely or stops working. We have often found it possible to diagnose and track down errors by using such a method. For example, in one experiment we varied thirty simulation inputs that we had previously left unchanged, and found that the

simulation failed to run part of the time. A few splits of a partition tree isolated the problem to an interaction between two factors that could lead to a buffer overflow. The bug could then be corrected.

2.4 Harvesting Efficiently

Automation is a key enabler of data farming, since there are many repetitive tasks. A little work up front makes life much easier down the road.

If you are just getting started on data farming, you may find it helpful to use some of the run control scripts in the `datafarming` Ruby gem described in Section 2.2. These are scripts that allow you to run any simulations that can be run from the command line (such as simulations written in python, Matlab, R, java, or similar languages) and .

When you are ready, parallel computing can easily be leveraged for purposes of data farming. Each run (a single replication of a single design point) is a self-contained simulation that can be sent off to a core, with the data consolidated once all runs are complete. Software such as HTCondor at https://research.cs.wisc.edu/htcondor/ [accessed 1 March 2021] can be used to farm jobs out to multiple cores, either on a single multicore machine or on a computing cluster. SESSL at http://sessl.org [accessed 1 March 2021] is another software language set up to facilitate experiments for a variety of simulation modelling platforms (Ewald and Uhrmacher 2014; Warnke and Uhrmacher 2018). For more about the nuts and bolts of data farming, see Sanchez and Sanchez (2017). If your models are set up to be data farmable from the start, running the data farming experiments is straightforward—and you will never want to go back to manual experimentation.

2.5 Maximizing Yield

By maximizing yield, we mean gaining as much knowledge and insight as we can from our simulation study to inform decision makers. This may be insight about the simulation model's behavior itself, or about a real-world situation that we are simulating. If the model's intended use is to assist decision makers on important and complex questions, then we should 'think big' in terms of the insights that might be gained. Decision makers attempting to address complex problems are not likely to be interested in answers to simple questions. Given the time and effort that can be spent to conceptualize and implement a simulation model, make sure that effort is put to good use. Data farming is a way to make your simulation model work for you!

Think of robustness as you plan your data farming experiments (Sanchez and Sanchez 2020). Robustness is a structured way to guard against making unwarranted assumptions. Factors in your data farming experiment can be differentiated as decision factors, noise factors, and artificial factors. Decision factors are those that can be controlled in the real-world setting for which the simulation is based. Noise factors are those that either cannot be controlled, or can be controlled only at great cost or difficulty, in the real world. Artificial factors are specific to the simulation environment, such as the warm-up period for steady-state simulations; choices of random number generators, seeds, or streams; run lengths; time intervals for discrete-time simulations; and more. Including artificial factors in a data farming experiment may yield insights about using simulation for real-time control. Including both decision and noise factors makes it more likely that recommended solutions will work well for a broad range of situations that might arise in practice, even if these are not optimal solutions for any particular setting. A robustness perspective can also be used to ascertain whether certain model assumptions, such as input distribution shapes, lead to substantively different recommendations. The current combination of computational power and modelling platforms and paradigms helps simulation modelers to reduce so-called 'Type III errors' of solving the wrong problem (Mitroff and Featheringham 1974). Seeking robust solutions aids this process.

2.6 Reaping the Benefits

Once we have generated an inferential big data set from our data farming experiment, what do we do with it? We have found that just as "having" big data from the internet meant that companies found new and exciting things to do with it, having big data from simulation experiments offers the opportunity for new and interesting ways of looking at the results (Elmegreen, Sanchez, and Szalay 2014). These include a wide variety of metamodeling and visualization techniques.

Theoretically, every one of the inputs should affect at least one response in some way. If not, there is something wrong with either the conceptual model (e.g., we have added unnecessary detail or left out connections) or its implementation (the code contains bugs). However, even if all factors have some 'true' effect, that does not mean they are all equally important. Data farming can help us identify the factors or interactions that are key drivers of performance over the region of factor exploration. Consequently, when constructing metamodels we may end up excluding factors or terms that are statistically significant—either because they are dwarfed by other factors or terms that have much stronger effects, or because their effects, while statistically significant, are not of practically interest given the region of interest for this particular experiment.

Some features are best revealed by graph-analytic techniques: see, e.g., Feldkamp Bergmann and Strassburger (2015, 2020), Matković, Gracanin, and Hauser (2018), or Sanchez (2020) for examples drawn from simulation experiments. Past data farming studies have helped save lives, time, money, and the environment; improve algorithms; and facilitate thoughtful discussions around modelling human behaviors and interactions.

2.7 Serving the Community

Our metaphor involves farming, not gardening. In the real world, both might be used to grow vegetables (or herbs, or flowers)—but a garden is a small plot intended for private use, while a farm is a larger enterprise that grows crops for others. This sense of distributing results to a large community of stakeholders, rather than simply generating the insights for ourselves, is important. The sense of scale also matters. We have over a trillion times the computing power at our fingertips than was used to first put a man on the moon (Lucas et al. 2015). How are we leveraging this power? Are our methods of building and analyzing simulation models keeping pace?

3 CONCLUDING THOUGHTS

Going forward, there are many opportunities for advancing the theory, the practice, and the applications of simulation. This work can be worthwhile, rewarding, fascinating, and fun! We hope the data farming metaphor helps researchers think broadly about how their talents and interests might grow the capability in one or more of these areas, and anticipate the needs that practitioners will face in the future. We hope this metaphor resonates with practitioners, allowing them to reap immediate benefits by using a data farming approach for their next simulation study. We hope that the breadth and depth of insights that can be gleaned will help decision makers in the public and private sectors turn to simulation as a means of obtaining useful, robust, and actionable recommendations to address the complex problems they face.

Our global simulation community has opportunities to make differences in all these dimensions. The COVID pandemic of the past year is but one striking example of how useful and important it can be to gain insights from modelling and simulation. Virtual experiments have helped facilitate timely decision making for numerous types of systems at a variety of levels, from procedures for administering tests and vaccines, to creating new layouts and patient flows for specific healthcare facilities, to policy recommendations at local, regional, or national levels intended to contain and halt the spread of the disease. The pandemic response also makes it clear that modelling and simulation are not enough. Our simulation community must continue to strengthen its ties and outreach to other communities—sharing with them, listening to them, and learning from them—to reach our full potential and help address the major challenges our world now faces.

ACKNOWLEDGMENTS

Thanks to the many people I have had the pleasure of discussing with and working on these concepts over the years. Department of Defense (DoD) Distribution Statement: Approved for public release; distribution is unlimited. The views expressed in this document are those of the author and do not necessarily reflect the official policy or position of the Navy, DoD, or the U.S. Government.

APPENDIX

There are many potential classes of designs available. Here are a few that we recommend, use regularly, and are readily available for you to use yourself. Most are components of the `datafarming 1.4.0` Ruby gem that can be found at https://rubygems.org/gems/datafarming [accessed 1 March 2021]; the `README` file has instructions for installing and running the gem on Windows, MacOS, or Linux systems. See https://bitbucket.org/paul_j_sanchez/datafarmingrubyscripts/src/master/ [accessed 1 March 2021] to view or download the source code.

- ***Nearly orthogonal Latin hypercubes (NOLH).*** These are space-filling designs suitable for quantitative factors that are continuous-valued or discrete-valued with many levels. The maximum absolute pairwise correlation between any two columns is less than 0.05. Several base design sizes (maximum factors *k*, number of design points *n*) can be used: those currently coded are (7,17), (11,33), (16,65), and (22,129) (Cioppa and Lucas 2007); (29,257) (Hernandez, Lucas, and Carlyle 2012), and (100,512) (Vieira et al. 2013). A shift-and-stack approach can generate larger designs with improved space-filling behavior for any *k* between 2 and 100.
- ***Nearly orthogonal-and-balanced (NOAB) designs***. These are suitable 'as is' for quantitative factors and discrete-valued factors with 2 to 11 levels, with maximum absolution pairwise correlation of 0.0347 between any two columns. Nearly-balanced means that the levels of any particular discrete-valued factor appear in roughly equal numbers of design points. A customizable 512-dp (design point) NOAB allows the analyst to create a design involving up to 20 *m*-level factors (m = 2, 3, …, 11) and 100 continuous-valued factors. With a little extra care, the discrete-valued columns can be used for qualitative factors as well. Also, the entire design can be shifted-and-stacked if the shift-and-stack is applied separately to each m+1 groups of columns: one group for the *m*-level factors) (m = 2, 3, …, 11) and one group for up to 50 continuous-valued quantitative factors.
- ***Resolution V fractional factorials (R5FF).*** These orthogonal designs are suitable for any mix of two-level factors, either qualitative or quantitative. They are not space-filling, but they have the property that all main effects, all quadratics, and all two-way interactions can simultaneously be estimated. Design sizes are powers of two. The design generators can be stored efficiently, and result in design sizes that are powers of two. Some examples are 2^2=4 dps for k=2, 2^{20-11}=512 for k=20, 2^{50-38}=4096 for k=50, and $2^{120-105}$=32768 for k=100. Applying shift-and-stack to these designs does not improve space-filling, but it does increase the number of corner points sampled.
- ***Resolution V central composite designs (R5CCD).*** These orthogonal designs are suitable for quantitative factors, and they have the property that all main effects and all two-way interactions can simultaneously be estimated. They augment the R5FFs with one center and 2*k* star points. This results in three levels per factor if the star points are placed on the faces of the hypercube, or five levels per factor if all non-center points are an equal distance from the center (a rotatable CCD). The improved space-filling behavior provides greater metamodel flexibility. Metamodels with quadratic terms can be fit from the output data for both types of CCDs. Metamodels could contain cubic or quartic terms for the rotatable CCDs.
- ***Frequency based designs (R5FBD).*** These orthogonal designs are suitable for quantitative factors, and have the property that all main effects, all quadratics, and all two-way interactions can simultaneously be estimated. Factor levels can be viewed as oscillating sinusoidally at carefully selected frequencies as a function of the design point. R5FBDs have a smaller proportion of dps in the interior of the sampling region than NOLHs, but a larger proportion than R5FFs. Some examples of design sizes are 13 for k=2, 1673 for k=20, 17761 for k=50, and 115434 for k=100. Applying a shift-and-stack approach to these designs improves their space-filling behavior.

Another straightforward way of creating a design that contains both qualitative and quantitative factors is to create two separate designs D_1 (for k_1 factors in n_1 dps) and D_2 (for k_2 factors in n_2 dps) and then *crossing* them, obtaining a design for k_1+k_2 factors in n_1n_2 dps. A crossed design is typically much larger than a single combined design (such as a NOAB) so a combined design is usually preferred if either k_1 or k_2 is large.

Other websites for obtaining data farming software and designs include

- The `datafarming` Ruby gem has self-documenting scripts for design generation, design scaling, and data farming run control. The `README` file has instructions for installing and running the gem on Windows, MacOS, or Linux systems. [https://rubygems.org/gems/datafarming accessed 31 January 2021].
- Source code for the `datafarming` Ruby gem can be viewed or downloaded from [https://bitbucket.org/paul_j_sanchez/datafarmingrubyscripts/src/master/ accessed 31 January 2021].
- The Naval Postgraduate School's *SEED Center for Data Farming* website at has downloadable spreadsheets (such as the customizable 512-dp NOAB) and links to other software [https://harvest.nps.edu accessed 31 January 2021].
- The R package FrF2Large also has code for generating the R5FF designs [https://rdrr.io/cran/FrF2/man/FrF2Large.html accessed 31 January 2021].

There are several other R packages that create designs. Many commercial statistical software packages, and some simulation modelling platforms, also have design-generating capabilities.

REFERENCES

Cioppa T M and Lucas T W (2007). Efficient nearly orthogonal and space-filling Latin hypercubes. *Technometrics* **49(1)**: 45-55.

Elmegreen B E, Sanchez S M, and Szalay A (2014). The future of computerized decision making. In: Tolk A, Diallo S D, Ryzhov I O, Yilmaz L, Buckley S, and Miller J A (eds). *Proceedings of the 2014 Winter Simulation Conference.* IEEE, Piscataway, New Jersey, pp 943-949.

Ewald R and Urhmacher A (2014). SESSL: A domain-specific language for simulation experiments. *ACM Transactions on Modeling and Computer Simulation* **24(2)**: 11:1-11:25.

Feldkamp N, Bergmann S, and Strassburger S (2015) Visual analytics of manufacturing simulation data. In: Yilmaz L, Chan W K V, Moon I, Roeder T M K, Macal C, Rossetti M D (eds) *Proceedings of the 2015 Winter Simulation Conference*. IEEE, Piscataway, New Jersey, pp 779–790.

Feldkamp N, Bergmann S, and Strassburger S (2020). Knowledge discovery in simulation data. *ACM Transactions on Modeling and Computer Simulation* **30(4)**: 24:1-24:25.

Hernandez A S, Lucas T W, and Carlyle M (2012). Constructing nearly orthogonal Latin hypercubes for any nonsaturated run-variable combination. *ACM Transactions on Modeling and Computer Simulation* **22(4)**: 20:1-20:17.

Lucas, T W, Kelton W D, Sanchez P J, Sanchez S M, and Anderson B L (2015). Changing the paradigm: Simulation, now a method of first resort. *Naval Research Logistics* **62**: 293-305.

Pearl J (2009). *Causality: Models, Reasoning, and Inference (2nd ed.).* Cambridge University Press, New York, NY.

Kusiak A (2006). Data farming: Concepts and methods. In: Triantaphyllou E and Felici G (eds). *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques*. Springer US: Boston, Massachusetts, pp 279–304.

Matković K, Gracanin D, and Hauser H (2018). Visual analytics for simulation ensembles. In: Rabe M, Juan A A, Mustafee N, Skoogh A, Jain S, and Johansson B (eds). *Proceedings of the 2018 Winter Simulation Conference.* IEEE: Piscataway, New Jersey, 321-335.

Mayo C S, Kessler M L, Eisbruch A, Weyburne G, Feng M, Hayman, J A, ... Ten Haken R K (2016). The big data effort in radiation oncology: Data mining or data farming? *Advances in Radiation Oncology* **1**: 260–271.

Sanchez S M (2018) Data farming: Better data, not just big data. In: Rabe M, Juan A A, Mustafee N, Skoogh A, Jain S, and Johansson B (eds). *Proceedings of the 2018 Winter Simulation Conference.* IEEE: Piscataway, New Jersey, pp 425–439.

Sanchez P J and Sanchez S M (2019). Orthogonal second-order space-filling designs with insights from simulation experiments to support test planning. *Quality and Reliability Engineering International* **35(3)**: 854-867

Sanchez S M (2020). Data farming: Methods for the present, opportunities for the future. *ACM Transactions on Modeling and Computer Simulation* **30(4)**: pp 22:1-22:30.

Sanchez S M and Sanchez P J (2017). Better big data via data farming experiments. In: Tolk A, Fowler J., Shao G, and Yücesan E (eds). *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences.* Springer International Publishing AG: Cham, Switzerland, pp 17-26.

Sanchez S M and Sanchez P J (2020). Robustness revisited: Simulation optimization viewed through a different lens. In: Bae K-H, Feng B, Kim S, Lazarova-Molnar S, Zheng Z, Roeder T, and Thiesing R (eds). *Proceedings of the 2020 Winter Simulation Conference.* IEEE: Piscataway, New Jersey, pp 60-74.

Schruben, L W (2017). Model is a verb. In: Tolk A, Fowler J., Shao G, and Yücesan E (eds), *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences.* Springer International Publishing AG: Cham, Switzerland, pp 17-26.

Vieira H, Sanchez S M, Kienitz K H K and Belderrain M C M (2013). Efficient, nearly orthogonal-and-balanced, mixed designs: An effective way to conduct trade-off analyses via simulation. *Journal of Simulation* **7(4)**: 264-275.

Warnke T and Uhrmacher A M (2018). Complex simulation experiments made easy. In: Rabe M, Juan A A, Mustafee N, Skoogh A, Jain S, and Johansson B (eds). *Proceedings of the 2018 Winter Simulation Conference.* IEEE: Piscataway, New Jersey, pp 410-424.

AUTHOR BIOGRAPHY

SUSAN M. SANCHEZ is a Distinguished Professor in the Operations Research Department at the Naval Postgraduate School, with a joint appointment in the Graduate School of Defense Management. Her research interests include experimental design, data-intensive statistics, and robust selection, with applications to simulation experiments, manufacturing, health care, and military operations. She is Co-director of NPS's *SEED Center for Data Farming*, which was established to advance the use of simulation experiments and efficient designs to provide decision makers with timely insights. https://harvest.nps.edu

Proceedings of the Operational Research Society Simulation Workshop 2020 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.

A DDDAMS-BASED SURVEILLANCE AND CROWD CONTROL VIA UAVS AND UGVS

Professor Young-Jun Son

The University of Arizona
Department Head of Systems and Industrial Engineering
Tucson, Arizona, 85721, USA
son@sie.arizona.edu

ABSTRACT

In this talk, we first introduce a dynamic data driven adaptive multi-scale simulation (DDDAMS) based planning and control framework that we have developed for effective and efficient surveillance and crowd control via unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs). The framework is composed of integrated planner, integrated controller, and decision module for DDDAMS. The integrated planner, which is designed in an agent-based simulation (ABS) and Unity-based game engine, devises best control strategies for each function of 1) crowd detection, 2) crowd tracking, and 3) UAV/UGV motion planning. The integrated controller then controls real UAVs/UGVs for surveillance tasks via 1) sensory data collection and processing, 2) control command generation based on strategies provided by the decision planner, and 3) control command transmission via radio to the real system. The decision module for DDDAMS enhances computational efficiency of the framework via dynamic switching of fidelity of simulation and information gathering. Finally, we will share the results of our field demo, which successfully integrated a fast running simulator, a real-time simulator, and the real system (viz. UAVs, UGVs, and crowd).

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.003

THE BASIC PRINCIPLES OF SYSTEMS THINKING AND SYSTEM DYNAMICS

Dr Brian Dangerfield

dangerfiel@gmail.com

ABSTRACT

This tutorial paper presents the basics of System Dynamics (SD) modelling, together with an introductory section on Systems Thinking, specifically influence (or causal loop) diagrams. The coverage of SD starts with how a stock-flow diagram is used to commence the conceptualisation process by the creation of a spinal flow(s). Auxiliary variables (contained in the information system) and model parameters are then deployed in equations which explain the various flow rates depicted in the spinal flows. By this means a full working model emerges with a web of information overlaid onto the spinal flows. The paper concludes with a fully worked (but simple) example in the domain of workforce modelling.

Keywords: Systems Thinking; System Dynamics; Influence Diagram; Stock-flow Diagram.

1 INTRODUCTION

This tutorial paper introduces the basics of the system dynamics simulation methodology, together with the adjunct field of systems thinking which emerged subsequently. The field of system dynamics was initially known as Industrial Dynamics which reflected its origins in the simulation of industrial supply chain problems. The first paper published by the founder of the field, Jay W Forrester, appeared in 1958 (Forrester, 1958) and it was a precursor to what proved to be a hugely influential book: Industrial Dynamics (1961). Forrester sought to apply concepts of control engineering to management type problems and was very probably influenced by the earlier work of Arnold Tustin (1953). Forrester argued that the field of Operations/Operational Research (OR) at that time was not focused on the sort of problems that sought to inform policy (top-level) issues in an organisation. By its very definition OR was restricted to operational problems. Forrester saw a niche for a methodology which could tackle strategic issues more appropriately addressed to the success or failure of an organisation, as well as prominent national and international policy issues. See Forrester (2007) for his personal recollections of the history of the field.

The characteristics of system dynamics simulation models can be listed as follows:

- They address issues by considering aggregates (of products, people etc) and not individual entities (as in discrete event simulation) or individual agents (as in agent based modelling)
- They primarily reflect the dynamics of a system as having endogenous causes: change over time comes from within the system boundary due to information feedback effects and component interactions, although the initial stimulus for those dynamics may be exogenous. For more on the endogenous perspective see Richardson (2011).
- They carefully distinguish between resource flows and the information flows which cause those resource flows to increase or run down. This is a fundamental (and powerful) feature of the methodology which means such models can be used to design and evaluate information systems as well as the more usual focus on resource systems.

- The flows are assumed to be continuous and are governed by what are in effect ordinary differential equations. System dynamics models belong to the broader category of continuous simulation models.
- Although flow rates are included, SD models are primarily concerned with the behaviour of stocks or accumulations in the system. These are described by integral equations. Forrester has famously stated that differentiation does not exist in nature, only integration. Mathematical models characterised by differential equations must be solved in order to determine the stock values; system dynamics puts stock variables to the forefront.[1]
- They do not ignore soft variables (such as morale or reputation) where these are known to have a causative influence in the system.

Before addressing some of these characteristics in greater detail it is sensible to offer an overview of the adjunct field of systems thinking. This is sometimes described as qualitative system dynamics for its provenance is based upon diagramming or mapping techniques, primarily influence diagrams (ID) or causal loop diagrams (CLD). It was not until the 1970's, nearly fifteen years after the publication of Forrester's early industrial models, before such diagrams started to appear. Their origins can be traced back to Maruyama (1963); Goodman's text (1974) portrays some seminal examples.

2 SYSTEMS THINKING

The use of diagramming techniques in the analysis of a system has a long history going back to the block diagrams of control and electrical engineering. However, the qualitative strand associated with system dynamics emphasises the feedback loops present in the system. Feedback is an essential building block of system dynamics whereby information about the current state of the system is used to regulate controls on the resource flows and it underscores the endogenous point of view. For instance, if stocks of manufactured goods are beginning to over-accumulate, it is necessary to either cut back on production throughput or inaugurate a sales drive or both.

These mapping techniques are not mandatory in a system dynamics study. On the other hand there are those who argue such methods, of themselves, have the capacity to generate insight and can help form a consensus for policy change in a problem system. See for instance the testimony from Merrill *et al* (2013) concerning a health application. They state: *"As a tool for strategic thinking on complicated and intense processes, qualitative models can be produced with fewer resources than a full simulation, yet still provide insights that are timely and relevant"*. Books have appeared which focus exclusively on such mapping techniques, for instance Ballé (1994) and Sherwood (2002), to the exclusion of formal simulation models which are described in section 3 below. Whether such diagrams alone can be considered advantageous in the overall practice of system dynamics has long been the subject of debate in the field. An exchange between Coyle and Homer and Oliva occupied many pages of the *System Dynamics Review* in 2000-2001. See Coyle (2000; 2001) and Homer and Oliva (2001).

Although the proponents of the need for formal system dynamics models remain implacable, some authors and organisations have prospered in the propagation of systems thinking techniques. Pegasus Communications has for many years published the magazine *The Systems Thinker* and Peter Senge's reputation as a managerial thought leader was founded on his book *The Fifth Discipline* (1990) and its associated *Fieldbook* (1994). It was these sources that, primarily, introduced 'behaviour over time' sketch graphs together with the notion of system archetypes as additional tools in the armoury of systems thinking.

2.1 Behaviour Over Time Graphs

Consider the chart in figure 1. It represents some (hypothetical) data for two local authorities showing the percentage of girls who were classed as overweight in 2012.

[1] Research has shown that even well-educated people find it difficult to infer the behaviour of a stock variable given known behaviour of the flows affecting that stock.

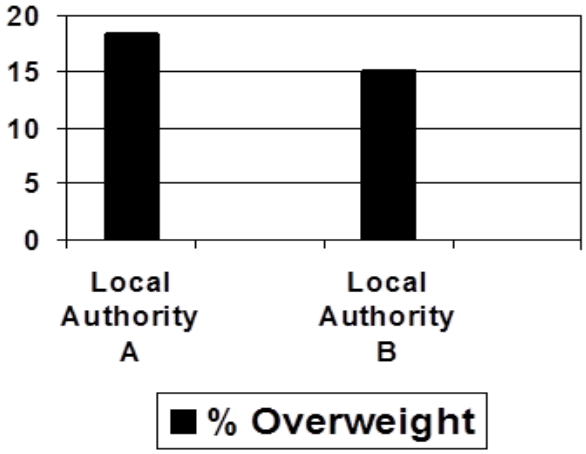


Figure 1 *Prevalence of overweight girls aged 10 – 15 years in two local authorities in 2012*

It is a static graph and, as such, conveys limited and what could be misleading information. At first, an examination of the data would appear to suggest that local authority A has a more serious public health situation on its hands than local authority B. However, re-framing the situation using a behaviour over time graph paints an altogether different picture (see figure 2). It is clear that local authority B is more in need of a public health intervention. Consideration of the dynamics in a system is vitally important.

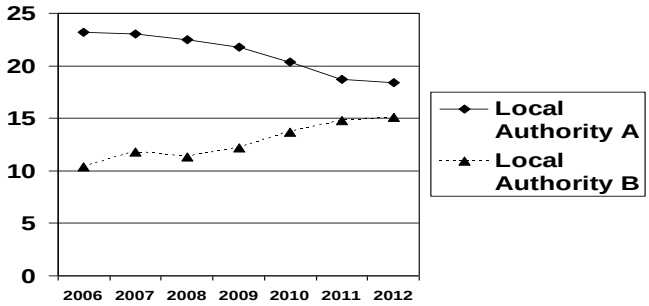


Figure 2 *Behaviour over time graph of the prevalence (%) of overweight girls, 2006-2012*

2.2 Principles of Influence (or Causal Loop) Diagrams

This section examines the building blocks of the mappings which have come to constitute the heart of systems thinking – the diagrams known as influence diagrams (ID) or causal loop diagrams (CLD). There is no counterpart to an ID or CLD in discrete event simulation. There one progresses to the development of an activity cycle diagram as the initial framework on which the computer simulation model is constructed. That is to say the field of discrete event simulation does not offer an optional diagramming phase which, of itself, is capable of generating insight.

Some practitioners have expressed the view that, in certain instances, an intervention based on systems thinking diagrams is sufficient to unearth the insight necessary to achieve a profound effect on system performance. The argument is bound up with project resources: models as mappings absorb less costs and can still produce insights which are timely and relevant.

A simple example of an influence diagram is given in figure 3. Here we see the basic process underlying a firm's organic growth. As average profits increase they are re-invested to the future benefit of the organisation (positive links).

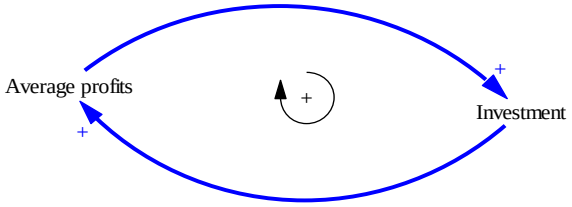


Figure 3 *Simple positive feedback loop (Note: the loop descriptor in the middle should flow in the same direction as the loop, in this case clockwise)*

Other examples of positive links are: sales per unit time of a durable product increase the customer base; revenues received increase the cash balance; students enrolling on a course increase the total student population.

Note that the + sign not only implies that an increase in one variable causes an increase in another but also, alternatively, that a reduction in one variable causes a reduction in another. In the example in figure 3 a reduction in average profits engenders a reduction in investment. A more obvious example is when rumours of a firm's financial health lead prospective customers to decline to engage with it.

Let us now consider a negative loop. The underlying influence created by such a loop is one of a controller. If movement occurs in the dynamics in one direction then a countervailing force pushes against that momentum to establish the original (or a new) equilibrium. The entire discipline of control engineering is concerned with how negative loops can be represented as physical controllers in machinery of all types, for example the auto-pilot in modern aircraft and the thermostat in a heating system.

Figure 4 shows an example of a simple negative loop taken from the domain of stock control. The very word 'control' reflects the nature of what is going on. As stock levels increase then replenishment ordering is cut, or vice versa (negative link). The change in the flow of orders directly affects the stock level and thus completes the loop. Other examples of negative links are: a perceived reduction in the numbers of a particular workforce will lead to an increase in employee recruitment; an increase in spending on wages will lead to a fall in an organisation's cash balances.

In selecting the sign to place on a given arrowhead (establishing link polarity)[2] it is important not to take into account other influences that may be simultaneously operating. The Latin maxim of *ceteris paribus*, so common in elementary economics texts, needs to be adhered to: i.e. let other factors remain constant. Therefore, the only consideration in assigning link polarity is: what effect will a change in the variable at the tail of the arrow have on the variable at its head?

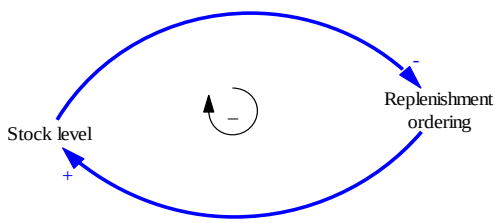


Figure 4 *Simple negative feedback loop*

Two mutually connected negative relationships create a positive loop. Consider figure 5 where an increase in staff turnover (in a close working team) will lead to a fall in morale which in turn will lead to a further increase in staff turnover.

[2] In recent years some authors have replaced the use of + and – by s (same) and o (opposite).

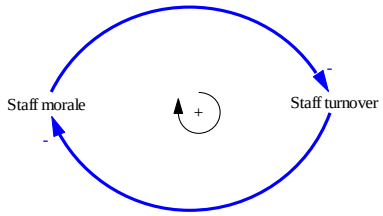


Figure 5 *Two mutually causative negative relationships create a positive loop*

In determining the loop (as opposed to link) polarity there are two methods available. One can enter the loop at any given point, and start with, say, an increase in that variable and trace around the effect. If one returns to that point with a further increase then the loop is positive, but if the initial increase has resulted in a decrease then the loop is negative. An arguably easier approach is to add up the number of negative links in the loop: if the number is zero or is even, then the loop is positive and if it is an odd number then the loop is negative. The loop polarity is the algebraic product of the number of negative signs, e.g. three negatives multiplied together yield a negative result, hence a negative loop.

2.3 From Diagrams to Behaviour

The determination of loop polarity is not merely an exercise for its own benefit but rather serves as a precursor to being able to infer the behaviour mode of the loop if it were to be 'brought to life'. Loop dynamics differ between negative and positive loops so it is essential to determine loop polarity. A positive loop produces dynamics which reinforce an initial change from an equilibrium point and so underpin growth and decay behaviour patterns. A pure positive loop in growth mode will produce exponentially increasing behaviour. A negative loop, on the other hand, will generate equilibrating behaviour such that any shift away from an initial equilibrium point will produce a compensating force driving it back towards that point (or indeed a new equilibrium). Introducing a delay into a negative loop will induce an oscillation in the behaviour. It is this knowledge which can aid in model conceptualisation when time-series data is available. After smoothing out any noise which may be present, an oscillatory behaviour pattern is indicative of a system dominated by a negative loop or loops; one which exhibits growth or decay would suggest that a positive loop is at work somewhere. An oscillatory behaviour associated with a trend up or down would suggest the need for a model conceptualisation based around a combination of negative and positive loops.

In order to further develop this idea of behaviour generated by different feedback loops it is necessary to move away from the single loop examples above to a more realistic real-world situation where multiple loops are at play. For instance, the example at figure 6 portrays a simple product diffusion model where initial sales generate further growth through 'word-of-mouth' effects but this growth is ultimately curtailed by market limitations of one form or another. Because this system structure also underpins the dynamics of an epidemic in a closed population (e.g. passengers on a cruise liner) the variables named for the diffusion example have been duplicated by the equivalent epidemic variables: the same system structure can underpin quite widely different situations!

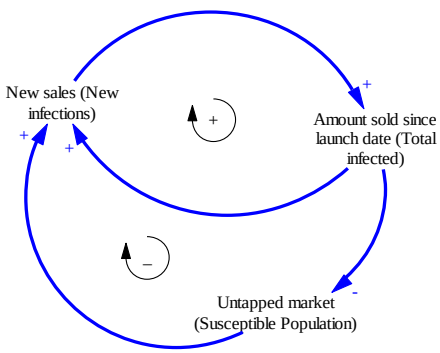


Figure 6 *Influence diagram showing two loops and two different examples: diffusion dynamics and epidemics underpinned by the same system structure*

Also brought out by figure 6 is the associated concept of loop dominance. As the structure plays out over time, the positive loop is dominant initially – that is to say it has the control of system behaviour in the early stages while the word-of-mouth effects are at play. Ultimately the market limits begin to take over. There are fewer and fewer people who do not have this product and so the capability of making further new sales is diminishing by each passing week. Now the negative loop assumes dominance in system behaviour and growth slows. Figure 7 shows the resultant behaviour: s-shaped growth where the transition from growth to market maturity coincides with the switch in loop dominance. Technically this is at the point of inflection on the cumulative curve, a point where the sales per unit time (not shown) reach a peak and start to fall.

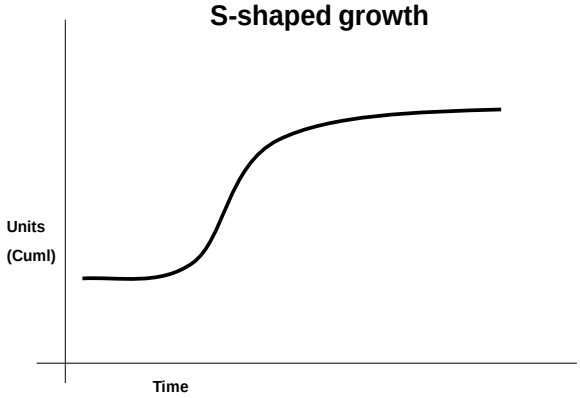


Figure 7 *S-shaped (or sigmoidal) growth generated by coupled positive and negative loops*

3 SYSTEM DYNAMICS

It is now appropriate to move forward and to consider the conceptualisation and formulation of a formal system dynamics simulation model. As mentioned earlier there is no essential requirement to preface the creation of an SD model with an influence diagram. There are those who argue that an influence diagram can aid in the definition of system content (and model boundary) but there is no direct linkage between such a diagram and the formal simulation model. This is in contrast to the stock-flow diagram: here the stocks (levels) and flows (rates) need to be explicitly present in the equation listing for the model.

3.1 Principles of Stock-flow Diagramming

The stock-flow diagram in system dynamics is the counterpart to the activity cycle diagram (ACD) in discrete event simulation. Although the flows may not result in a cycling of resources as such (which is common in DES), each diagram is there to underpin the formal model and the quantitative expressions which define its constituent elements.

System dynamics flow rates are depicted by a tap-like symbol which indicates a device which can control the flow, equivalent to policy controls in the real world. A stock is represented by a rectangle and here there exists an unfortunate misalignment in the DES and SD diagramming conventions. In DES a rectangle is reserved for an activity – an active state. A stock in SD is a ‘dead’ state, equivalent to a queue in DES. Figure 8 is an example of what might be part of the stock-flow diagram underpinning an SD model of a nation’s education system.

It is important to note that the boundary of the flow at each edge of the system is represented by a cloud-like symbol. Consideration of the resource beyond these points is outside the scope of the model. Also, the stocks and flows must alternate along the sequence. The incoming flow adds to a stock while an outgoing one drains it. Only one resource can be considered along any process flow. So, for example, what starts as a flow of material (or product) cannot suddenly be transformed into a flow of finance. Thus, separate flow lines need to be formulated for the various different resources being considered in the model. ‘Resource’ can be taken to be a product class, financial flow, human resources, orders, capital equipment and so on. Clearly the more resource flows being considered the more complex the model and the more equations it will comprise.

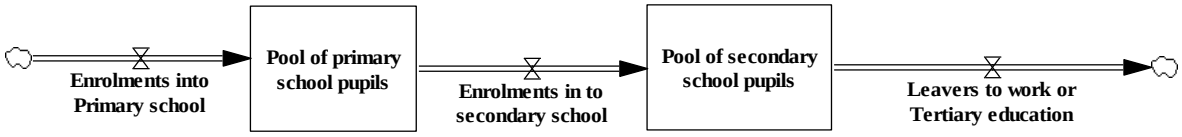


Figure 8 *Example of a single flow process in a stock-flow diagram*

Additional arrangements are possible. One can have an inflow to a stock without an outflow (or vice versa). A flow structure may call for more than one inflow and/or outflow. In certain cases the flow might actually form a cycle. This can happen, for instance, if one is modelling a manufacturing recycling process often described as 'reverse logistics' or a 'closed loop supply chain'. Although such flow arrangements do constitute a loop or cycle they are in no circumstances a feedback loop. As will be described later, a feedback loop is based on *information* feedback.

3.2 Model Purpose and Model Conceptualisation

Getting started can be the greatest difficulty in the creation of a useful SD model. One starts with the proverbial blank sheet of paper. Experience over many years has taught the author that two fundamental aspects of SD model conceptualisation are: firstly, being able to write in one sentence the purpose of the model and, secondly, ensuring the stock-flow representation is 'right'. This latter term is deliberately placed within inverted commas because no model can ever be perfectly correct and represent the ultimate truth, but it is meant to suggest that a great deal of thought needs to go into deciding which resource flows to include, and how to structure those flows as bald stocks and flows with no consideration of any other variables or constants at this juncture – these can be usefully termed the spinal stock-flow structures (see example in figure 8). Where a client is involved they need to 'buy into' that raw stock-flow diagram and the written definition of model purpose before any further model formulation work is undertaken. Several iterations of this first conceptualisation are typically necessary. The above advice also underlines the point made earlier about influence diagrams – they are not always necessary as a precursor to formal model creation. For this task the stock-flow diagram reigns supreme.

A particularly useful precept, first expressed by Forrester in *Industrial Dynamics* (1961), is to define the level (stock) variables. These would still be visible if the system metaphorically stopped (e.g. employees in a factory; cash in the firm's bank accounts). Next, consider what might be flowing into and/or out of those stocks. These flows would, of course, *not* be visible if the system 'stopped'. All the time it is necessary to remember that a number of different spinal stock-flow modules may be required in order to fully conceptualise the model in line with the agreed model purpose.

3.3 Adding Auxiliaries, Parameters and Information Links to the Spinal Stock-flow Structure

In order to flesh out the spinal stock-flow structure it is necessary to embellish it with other explanatory variables (called auxiliaries), together with parameters. In general one follows the oft-restated mantra: rates (flows) affect levels (stocks) via resource flows, while levels (stocks) affect rates via information (feedback) links. The sequence is:

Resource flows >> System state
System state >> Information to management
Information to management >> Managerial action
Managerial action >> Resource flows

This is the essential expansion of the concept of the feedback loop which is illustrated in figure 9.

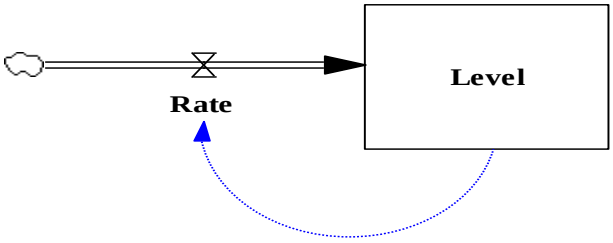


Figure 9 *A simple feedback loop in stock-flow symbolism*

In general more complexity will be required and other variables, which are neither stocks nor flows, of necessity have to be introduced – these are termed auxiliaries. These reflect variables which, in a business model, lie in the managerial planning and information system. Thus, any variable which is intended to represent something planned, desired, a target, or a management goal would be modelled by an auxiliary variable. Consider the augmented stock-flow diagram in figure 10. Here the concept of a desired workforce has been added to explain the recruitment rate on the spinal flow. It would seem intuitive that recruitment policy might be explained by a comparison between the desired workforce and what one currently possessed.

The level of sophistication can increase, however. To jump ahead a little there is another item which would need to be added, namely the adjustment time for eliminating any discrepancy between the desired and actual workforce. *Workforce adjustment time* would be a parameter and would mimic the average time to advertise and recruit new people or to give notice of redundancy and fire them if business conditions dictated it. Moreover, it might be necessary to have two different parameter values if the average time constant were thought to be different for recruitment and firing processes. Additionally, there may be a need to introduce other auxiliary variables in order to better define the *desired number of employees*. In fact, chains of auxiliaries are often created in order to effect a proper definition for the flow variable. System dynamics models tend to reflect real-world causes and effects very closely and this is one of the reasons why it is such a powerful methodology and why the total number of variables and parameters can rapidly escalate over and above the original number of variables on the spinal flows.

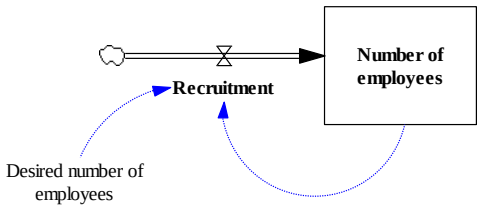


Figure 10 *Auxiliary variable and information links added to the spinal flow*

3.4 Equation Writing and Dimensional Checking

Undoubtedly for many the most challenging task in SD model formulation is the composition of the equations for the rates and auxiliary variables. In modern SD software the stock variables are automatically created because the system can 'see' what is flowing into and/or out of a stock. These integration equations take the form:

Stock value at current time t= old value of stock at t-dt + dt*(Inflows – Outflows) [dt is 'delta time' an infinitesimally small slice of time in mathematical calculus; approximated by a fraction of the time unit on a computer.]

System dynamics simulations exhibit a constant time advance (unlike DES) and through this process the equations describing the flow rates (which are, in effect, differential equations) are converted to difference equations and solved to yield the values of the stocks as in the example above. In the earlier SD literature the time increment was termed dt to reflect the 'with respect to' element commonly seen in differential calculus; TIME STEP is often employed nowadays. Its value is normally restricted to a binary fraction ($1/2^n$, for n=0,1,2…) because of the way computers handle real numbers;

by this means the greatest accuracy is achieved in determining the value for the system (reserved) variable *Time*. Clearly at the beginning of the simulation an 'old' value is needed to initialise the stock and this is termed an *initial value*. All stocks must have an associated initial value declared in order for the time advance process of the simulation to get started.

However, whilst formulation of the integration equations can be left to the software, this is not the case with rate and auxiliary equations. Here the user needs to compose the expression based upon the known informational influences evident in the developing stock-flow diagram. To this end it is recommended that the influence links are entered on the diagram *before* building the equation.

Structuring the equation is unavoidably bound up with units (or dimensional) checking. Most with a background in the physical sciences and engineering will know that any equation describing a real-world process needs to have the units balanced on each side of the '=' sign. Thus, if the units on the left are $/yr then the expression on the right side needs to algebraically decompose to $/yr. Further, if any terms on the right side are added or subtracted then each individual term needs to have the same units as the variable on the left side.

In the integration equation above, the 'dt*' element on the right side is necessary in order for the units to balance since the flows will be in terms of units/time. The dt term is a time interval and so we have time*units/time= units and the entire expression is units= units + units – units.

For the formulation of rate and auxiliary equations the user needs to think in terms of the units involved. If the variable concerned is expressed in terms of units/mths then the expression on the right side needs to also be units/mths. Thinking along these lines can actually aid in the formulation of the expression. You should know what units the rate or auxiliary is measured in; the right side needs to duly conform.

Let's consider some simple examples:

(1) The Accounts Payment Rate (APR) is known to be influenced by the value of Accounts Payable (AP) and a Delay in Making Payment (DMP).

APR= AP/DMP and $/mths = $/mths [This describes a flow of funds used to settle accounts.]

(2) The annual Out-Migration Rate (OMR) from a certain region of a country is dependent on the Population (POP), the normal Fraction of People Leaving (FPL) and the Departure Migration Multiplier (DMM).

The multiplier term could be there to account for periods of time when the normal fraction departing is tweaked as a result of, say, a temporary incentive. Where such constructs are employed in SD models they are inevitably dimensionless, that is to say they have no units. As well as a multiplier, any fraction, proportion, percentage or an index number would be dimensionless and be given units of '1'.

So we have: OMR=POP*FPL*DMM and persons/yrs= persons * 1/yrs * 1

Why is the FPL term in units of 1/yrs? This is because it is the number of persons leaving each year divided by the number there to start with, or (persons/yrs)/persons= 1/yrs. The same idea applies with an interest rate which is ($/yrs)/$ = 1/yrs (i.e. a percentage, which is dimensionless, but which can change over time).

Below are listed two possible equations to describe the Production Rate (PR). It is interesting to note that each is quite different but both are dimensionally balanced.

(3) Production Rate (PR) is a function of the Workforce (WF) and their Productivity (PROD). Productivity can crop up in a lot of business models and its dimensions can cause difficulty. It is a compound dimension expressed as (output) units/person/time unit, or (units/(persons*time)).

So we have: PR= WF*PROD and units/time= persons * (units/(persons*time))

(4) Production Rate (PR) is related to the Average Sales Rate (ASR), together with a Correction for a Stock Discrepancy (CSD) and a Correction for a Backlog Discrepancy (CBD). The correction terms will be accounted for separately in the model and they describe the product units produced per time unit that will eliminate any discrepancy between what is desirable and the state of affairs that exists.

So we have: PR= ASR + CSD + CBD and units/time= units/time + units/time + units/time

Which formulation for PR is the correct one? *Either* could be and there may indeed be other formulations which occur to the reader. The formulation employed is the one which is most appropriate given the purpose of the model and the circumstances prevalent in the actual system being modelled. A useful categorisation of commonly found formulations for rate and auxiliary equations is set out in the classic SD text by Richardson and Pugh (1981) and also in Sterman (2000). In addition the aspiring modeller should also study the many model listings provided by SD experts in texts and as supplementary material in journal articles.

To conclude a more complicated equation formulation example is described. It concerns the need to formulate an expression for the Extra Labour (EL) required to eliminate a greater than normal backlog of orders. Many operations experience this challenge, especially if there are seasonally induced gluts in orders. It is not feasible to employ a large workforce throughout time and it falls to the management to recruit more people when a very high backlog situation arises.

An initial formulation might be:

$$EL= (OB - NOB)/ PTAB$$

where OB= Order Backlog
NOB= Normal Order Backlog
PTAB= Planned Time to Adjust the Backlog

EL is obviously dimensioned as ‘persons’ and the expression is:

$$persons= units/time - units/time$$

The equation is not balanced dimensionally. It is necessary to introduce another variable (or constant) which will relate units/time to persons. A moment’s thought should make one realise that the concept of worker productivity (see example 3 above) is missing and so an additional parameter is required, say Normal Productivity of Labour (NPL). After some further thought it will be established that this parameter needs to be included in the denominator of the expression and as a multiplier. We now have:

$$EL= (OB - NOB)/ (PTAB*NPL)$$

Extracting just one of the terms in the numerator for the dimensional check yields:

$$persons= units/ (time * (units/(persons*time)))$$

and the two ‘time’ and ‘units’ elements cancel, leaving persons= persons and the equation is shown to be dimensionally correct. In the final step above it might be necessary to recall the mathematical dictum often chanted in school: “Invert the divisor and multiply”.

4 A COMPLETE SYSTEM DYNAMICS MODEL

The following gives a description of a problem at an IT Service Centre. We will conceptualise a model to address this issue and then develop and run simulation experiments (using Vensim) to explore a changed workforce recruitment policy which can improve dynamic performance. For more on workforce modelling see Cave & Willis (2020).

An IT service centre is staffed by 25 skilled specialists who respond to problems and difficulties which IT users make them aware of. When a user raises an issue a ticket is created containing a unique reference number for that problem. Once the matter is dealt with the ticket is completed and so is removed from the list of the IT ticket caseload.

From a historical perspective there is a normal turnover of IT staff. HR data shows around 3 staff have departed per month (12%) and the management have traditionally recruited new staff at the same average rate of departing staff. Typically, there will be 200 outstanding tickets in the caseload and these are raised at the rate of 500 per month and indeed completed at the same rate. So, an IT specialist will, on average, complete 20 tickets per month.

Over recent months there has been a surge in workload arising from a new software installation which users are having trouble familiarising themselves with. This has put pressure on the IT staff who are having to increase their workload (working longer hours etc) dealing with the surge in ticket submissions. In fact, it is having repercussions in terms of team morale and a lot more IT specialists are leaving the company than has hitherto been the case. The management are of course recruiting new skilled IT specialists but it doesn't seem to be adequately addressing the problem and the IT workload issue is not improving.

To conceptualise this issue we need to consider two resources: people and orders (tickets).

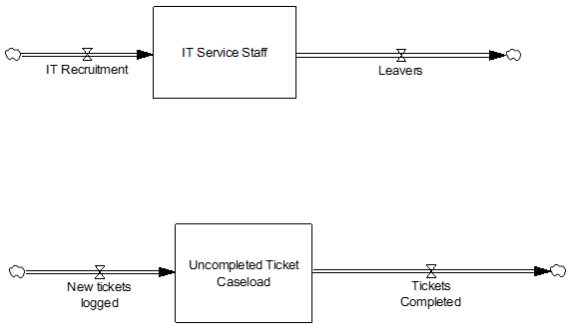


Figure 11 *Initial raw stock-flow conceptualisation of the workload problem*

We create two spinal stock and flow resource modules (as shown in figure 11). Next, we need to write expressions to capture the informational links which drive the flows. These will be either natural forces or managerially imposed forces (i.e. policies). In so doing we will need to introduce additional variables which lie in the information systems driving the flows; these are called auxiliary variables. This leads to figure 12. Also shown are the detailed equations and parameter values as produced by the Vensim model documentation feature.

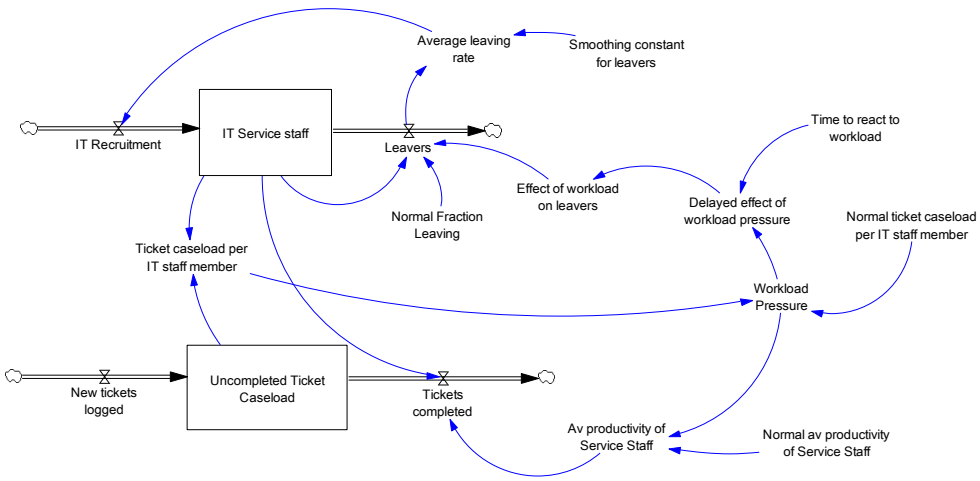
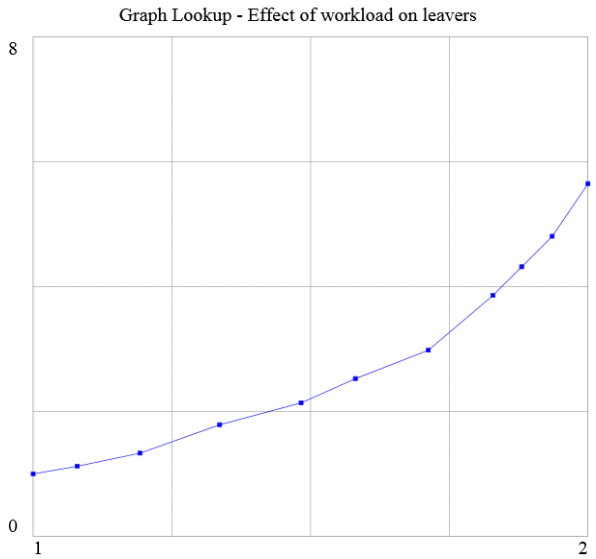


Figure 12 *Model with information links and parameters added*

(01) Av productivity of Service Staff=Normal av productivity of Service Staff*Workload Pressure
Units: tickets/(persons*mths)
(02)Average leaving rate=SMOOTH(Leavers, Smoothing constant for leavers)
Units: person/mths
(03)Delayed effect of workload pressure=SMOOTH3(Workload Pressure, Time to react to workload)
Units: Dmnl (dimensionless)
(04)Effect of workload on leavers= WITH LOOKUP (Delayed effect of workload pressure, ([(1,0)-(2,8)],(1,1),(1.07951,1.12281),(1.19266,1.33333),(1.33639,1.78947), (1.48318,2.14035),(1.58104,2.52632),(1.71254,2.98246),(1.82875,3.85965), (1.88073,4.31579),(1.93578,4.80702),(2,5.64912)))
Units: Dmnl (dimensionless) This makes use of a Vensim X-Y lookup as shown below.



(05)FINAL TIME = 24 Units: mths (The time horizon for the simulation.)
(06)INITIAL TIME = 0 Units: mths (The initial time for the simulation.)
(07)IT Recruitment=Average leaving rate Units: persons/mths
(08)IT Service staff= INTEG (-Leavers+IT Recruitment, 25) Units:persons
(Equation provided by Vensim – apart from initial value of 25)
(09)Leavers=IT Service staff*Normal Fraction Leaving*Effect of workload on leavers
Units: persons/mths
(10)New tickets logged=500+STEP(125,6)
Units: tickets/mths (Mimics the increase in workload: sudden 25% increase at t=6)
(11)Normal av productivity of Service Staff=20
Units: tickets/(persons*mths)
(12)Normal Fraction Leaving=0.12
Units: 1/mths
(13)Normal ticket caseload per IT staff member=8
Units: tickets/person
(14)SAVEPER = TIME STEP
Units: mths [0,24]
The frequency with which output is stored.
(15) Smoothing constant for leavers=3
Units: mths
(16)Ticket caseload per IT staff member=Uncompleted Ticket Caseload/IT Service staff

Units: tickets/person

(17)Tickets completed=IT Service staff*Av productivity of Service Staff

Units: tickets/mths

(18)TIME STEP = 0.03125

Units: mths [0,24]

The time step for the simulation.

(19)Time to react to workload=6

Units: mths

(20)Uncompleted Ticket Caseload= INTEG (New tickets logged-Tickets completed, 200)

Units: tickets (Again, provided by Vensim.)

(21)Workload Pressure=Ticket caseload per IT staff member/Normal ticket caseload per IT staff member

Units: Dmnl (dimensionless)

From this base case model we can show that the recruitment policy for new staff is too reactive. Figure 13 paints the poor performance.

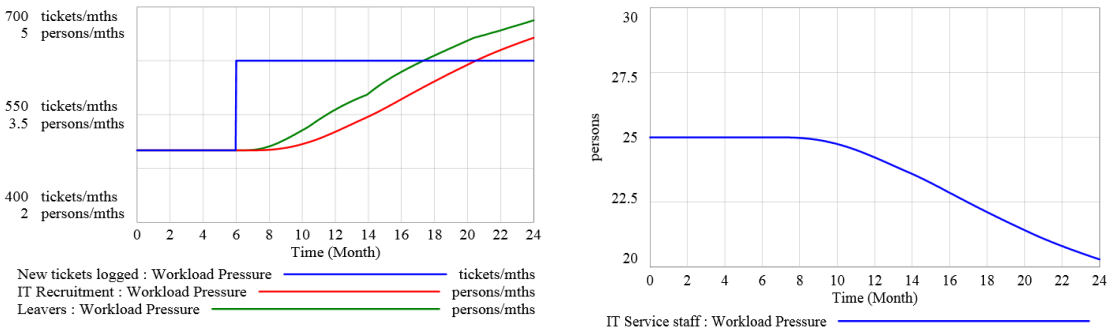


Figure 13 *Output graphs from the base case*

We need a policy based upon an improved 'early warning'. A possibility would be to monitor new tickets logged and then introduce the concept of *desired IT service staff* numbers based upon this. This revised policy is depicted in the model shown in figure 14.

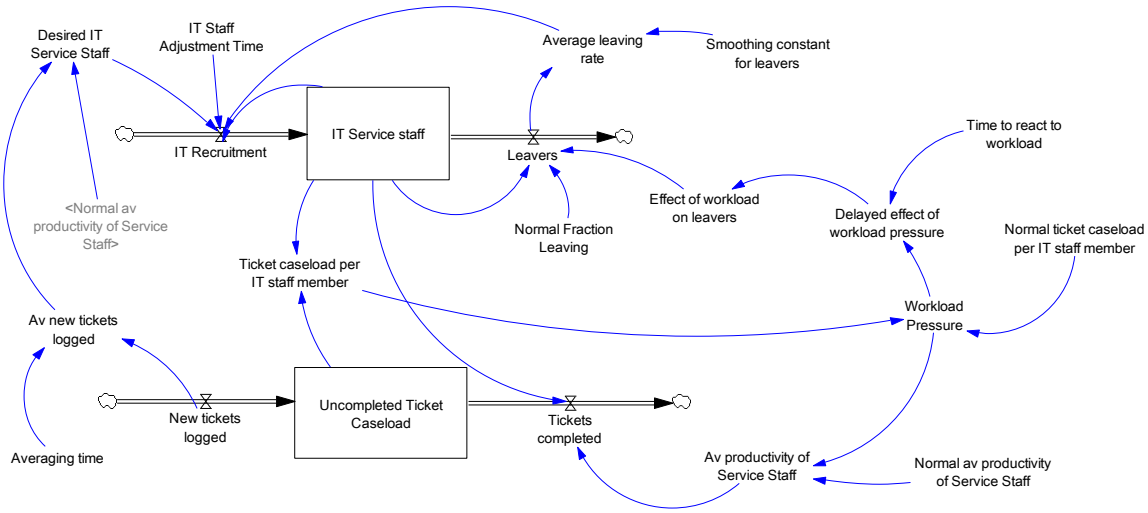


Figure 14 *Revised recruitment policy for IT staff*

The revised / additional equations are listed below:

(01)Av new tickets logged=SMOOTH(New tickets logged, Averaging time)
Units: tickets/mths
(02)Averaging time=1
Units: mths
(03)Desired IT Service Staff=Av new tickets logged/Normal av productivity of Service Staff
Units: persons
(04)IT Recruitment=(Desired IT Service Staff-IT Service staff)/IT Staff Adjustment Time+Average leaving rate
Units: persons/mths
(05)IT Staff Adjustment Time=1
Units: mths

The new recruitment policy has beneficial dynamic effects shown in figure 15.

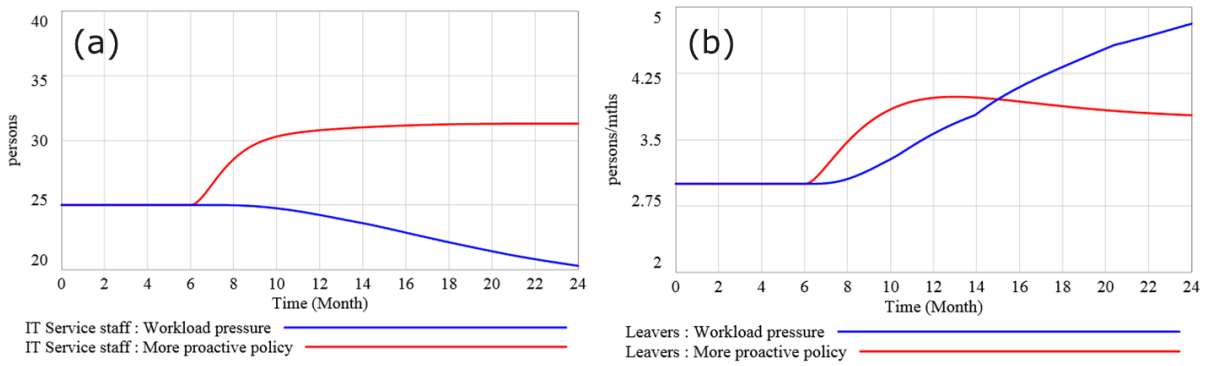


Figure 15 *Comparison plots of (a) IT services staff and (b) leavers under existing and revised policies*

FURTHER READING

It is impossible in this overview paper to fully do justice to what is now a significant methodology in the socio-economic, managerial, health, biological, environmental, energy and military sciences. However, three contemporary books will take the interested reader much further. These are purely the author's choice and they are listed in order of page count.

John Sterman's book (982pp) has arguably the most comprehensive coverage; see Sterman (2000) in the reference list. John Morecroft's (466pp) *Strategic Modelling and Business Dynamics: a feedback systems approach*, Wiley (2007) offers a very wide coverage of systems thinking and system dynamics and incorporates many practical model examples. Thirdly, Kambiz Maani and Bob Cavana have written a second edition of their offering (288pp): KE Maani and RY Cavana, (2007) *Systems Thinking, System Dynamics: managing change and complexity*, Pearson Education NZ (Prentice Hall), Auckland.

All these books come with a CD-ROM and/or a website which provides specimen models to be run and allows scenario experiments to be conducted. Exercises and instructor's manuals are also available.

An expanded version of the above material can be found in chapter 3 of the book by Brailsford, Churilov and Dangerfield (2014). This book covers both discrete-event simulation and system dynamics.

The book edited by Dangerfield (2020; 540pp) attempts to portray both methodological aspects and contemporary applications of SD now the field has just passed its 60th anniversary.

REFERENCES AND SELECT BIBLIOGRAPHY

Ballé M (1994) *Managing with Systems Thinking*, McGraw Hill, London.
Barlas Y (1996) Formal Aspects of Model Validity and Validation in System Dynamics. *System Dynamics Review*, 12:3, 183-210.

Brailsford S, Churilov L and Dangerfield B (Eds) (2014) *Discrete Event Simulation and System Dynamics for Management Decision Making*, Wiley, Chichester. See chapters by Robinson; Elder (DES & SIMUL8) and by Dangerfield; Jones (SD & VENSIM).

Cave S and Willis GJ (2020) System Dynamics and Workforce Planning. In: *System Dynamics*: *Theory and Applications*, Dangerfield BC (Ed.), Springer, New York. 431-457.

Coyle RG (2000) Qualitative and quantitative modelling in system dynamics: some research questions, *System Dynamics Review*, 16:3 225-244.

Coyle RG (2001) Rejoinder to Homer and Oliva, *System Dynamics Review*, 17:4 357-363.

Coyle G and Exelby D (2000) The Validation of Commercial System Dynamics Models, *System Dynamics Review*, 16:1, 27-41.

Dangerfield BC and Roberts CA (1996) An Overview of Strategy and Tactics in System Dynamics Optimisation, *Journal of the Operational Research Society*, 47:3 405-423.

Dangerfield BC and Roberts CA (1999) Optimisation as a statistical estimation tool: an example in estimating the AIDS treatment-free incubation period distribution, *System Dynamics Review*, 15:3, 273-291.

Dangerfield BC (2009) Optimization of System Dynamics Models. In: Meyers RA (ed.) *Encyclopaedia of Complexity and Systems Science*, Springer, New York 9034-9043. Reprinted in *Complex Systems in Finance and Econometrics*, Springer, New York, 2011, 802-811.

Dangerfield BC (Ed.) (2020) *System Dynamics: Theory and Applications*, Springer, New York.

Dangerfield BC & Duggan J (2020) Optimization of System Dynamics Models, In: *System Dynamics: Theory and Applications*, Dangerfield BC (Ed.), Springer, New York, 139-152.

Forrester JW (1958) Industrial Dynamics – a major breakthrough for decision makers, *Harvard Business Review*, 36:4, 37-66.

Forrester JW (1961) *Industrial Dynamics*, MIT Press, Mass (Now available from the System Dynamics Society, Albany NY.)

Forrester JW (2007) System Dynamics – a personal view of the first fifty years, *System Dynamics Review*, 23:(2-3) 345-358.

Forrester JW and Senge PM (1980) *Tests for Building Confidence in System Dynamics Models.* In: Legasto, AA Jr., Forrester, JW and Lyneis JM (Eds.): *System Dynamics*. Amsterdam: North-Holland Publishing Company, 209-228.

Goodman MR (1974) *Study Notes in System Dynamics*, Wright-Allen Press, Cambridge, MA (Now available from the System Dynamics Society, Albany NY.)

Graham AK and Ariza CA (2003) Dynamic, hard and strategic questions: using optimisation to answer a marketing resource allocation question, *System Dynamics Review*, 19:1, 27-46.

Hardin G (1968) The Tragedy of the Commons, *Science* 162:3859, 1243-1248. *DOI: 10.1126/science.162.3859.1243*

Homer J and Oliva R (2001) Maps and models in system dynamics: a response to Coyle, *System Dynamics Review*, 17:4 347-355.

Keloharju R and Wolstenholme EF (1988) The Basic Concepts of System Dynamics Optimisation, *Systems Practice*, 1: 65-86.

Keloharju R and Wolstenholme EF (1989) A Case Study in System Dynamics Optimisation, *Journal of the Operational Research Society*, 40:3, 221-230.

Maruyama M (1963) The Second Cybernetics: Deviation-Amplifying Mutual Causal Processes *American Scientist*, 51:2, 164-179.

Meadows DM and Robinson JM (1985) *The Electronic Oracle*, Wiley: Chichester. (Now available from the System Dynamics Society, Albany NY.)

Meadows DL and Sterman J (2012) Fishbanks: A Renewable Resource Management Simulation https://mitsloan.mit.edu/LearningEdge/simulations/fishbanks/Pages/fish-banks.aspx [Accessed 30 May, 2013]

Merrill JA, Deegan M, Wilson RV, Kaushal, R, Fredericks, K. (2013) A system dynamics evaluation model: implementation of health information exchange for public health reporting. *Journal of the American Medical Informatics Association*, 20:e131-e138 *DOI:10.1136/amiajnl-2012-001289*.

Richardson GP and Pugh AL (1981) *Introduction to System Dynamics Modelling with DYNAMO*, MIT Press, Cambridge, MA. (Now available from the System Dynamics Society, Albany NY.)

Richardson GP (2011) Reflections on the foundations of system dynamics, *System Dynamics Review*, 27:3 219-243.

Schwaninger M and Grösser S (2020) System Dynamics Modelling: Validation for Quality Assurance. In: *System Dynamics: Theory and Applications*, Dangerfield BC (Ed.), Springer, New York, 119-138.

Senge P (1990) *The Fifth Discipline*, Currency Doubleday, New York.

Senge P, A Kleiner, C Roberts, R Ross, B Smith (1994) *The Fifth Discipline Fieldbook*, Currency Doubleday, New York.

Sherwood D (2002) Seeing the Forest for the Trees: A Manager's Guide to Applying Systems Thinking, Nicholas Brealey, London.

Sterman JD (2000) Business Dynamics: Systems Thinking and Modeling for a Complex World, Irwin McGraw-Hill, New York.

Sterman JD (1984) Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models, Dynamica, 10:2, 51-66.

Tustin A (1953) The Mechanism of Economic Systems, Harvard University Press, Cambridge, MA.

Warren K (2020) Engineering of Strategy: A General Unified Theory of Performance and Strategic Management. In: System Dynamics: Theory and Applications, Dangerfield BC (Ed.), Springer, New York, 69-90.

Wolstenholme EF (2003) Towards the definition and use of a core set of archetypal structures in system dynamics, *System Dynamics Review*, 19:1, 7-26.

Wolstenholme EF & McKelvie D (2019) *The Dynamics of Care*, Springer, New York. (A book which sets out many of the health & social care models which the authors have developed during their careers as consultants.)

AUTHOR BIOGRAPHY

BRIAN DANGERFIELD has been teaching and researching using the system dynamics methodology for nearly fifty years. His career as an academic has involved spells in Business & Management departments at the universities of Liverpool, Salford and Bristol. His research output has included a range of system dynamics applications in areas such as business, health, economic modelling and epidemiology.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.004

TUTORIAL ON OPTIMISATION VIA SIMULATION: HOW TO CHOOSE THE BEST SET UP FOR A SYSTEM

Dr Christine Currie

University of Southampton
Highfield Road
Southampton
SO17 1BJ
christine.currie@soton.ac.uk

Dr. Thomas Monks

University of Exeter Medical School
Heavitree Road
Exeter
EX1 1TE
t.m.w.monks@exeter.ac.uk

ABSTRACT

In this tutorial we consider the problem of finding the best set up to use for a system, where the objective is measured using the output of a stochastic simulation model. What makes this a difficult problem is that the output is stochastic and consequently changes in each replication. Optimisation via simulation is a vast topic and we restrict ourselves to a small part of it – ranking and selection – in which a small number of discrete options are being compared. We describe two of the best-used methods, KN++ and OCBA. In these algorithms, just one solution is returned at the end of the optimisation and there is a single objective. We also discuss variations including best subset selection, multi-objective optimisation via simulation, and the minimisation of the expected opportunity cost. The tutorial is accompanied by a Github repository which includes Python code for the algorithms we describe here.

Keywords:

Optimisation via simulation, simulation, ranking and selection

1 INTRODUCTION

Optimisation via simulation (OvS) methods use the simulation as a proxy for the real system and the OvS algorithms will choose how to experiment on the simulation model in order to identify the optimal solution as efficiently as possible. Using mathematical notation, we wish to minimise an output $f(\mathbf{x})$, where $\mathbf{x}$ is a vector of decision variables and $f(\mathbf{x})$ is the expected value of the random output $Y(\mathbf{x})$,

$$f(\mathbf{x}) = \mathbb{E}[Y(\mathbf{x})].$$

We assume that the output $f(\mathbf{x})$ is a single number (or scalar) in most of what follows but in a multi-objective problem, it becomes a vector.

Hong and Nelson (Hong and Nelson 2009) provide a useful classification of simulation optimisation problems, dividing them into three main groups.

1. The feasible region for $\mathbf{x}$ has a small number of discrete solutions, *e.g. deciding between several different set-ups for a hospital to maximize throughput.*
2. The vector of decision variables $\mathbf{x}$ is continuous, *e.g. choosing the intervention or mix of interventions that leads to the least deaths when treating an infectious disease*
3. The vector of decision variables $\mathbf{x}$ is discrete and integer ordered, *e.g. optimizing the number of call centre staff on duty to minimize costs subject to constraints on response times.*

We would recommend *The Handbook of Simulation Optimisation* edited by Michael Fu (Fu 2015) as a good place to find out more about appropriate algorithms for these three different classes of problem along with tutorial papers from the Winter Simulation Conference.

In this tutorial we will focus on the first group of problems and the algorithms used to solve them, typically referred to as *ranking and selection* algorithms. We discuss the characteristics of ranking and selection problems in the next section and some variations on the basic single-objective, single-optimum setting. During the tutorial, we will make use of examples to illustrate the algorithms we are introducing here and we refer the reader to the Github repository for more details of these (https://github.com/TomMonks/ovs-tutorial).

Two main algorithms exist within the literature for solving these ranking and selection problems: indifference zone procedures such as KN++ and optimal computing budget allocation or OCBA methods. We describe each of these in turn in Sections 3 and 4. There are numerous extensions to these algorithms and also avenues for future research, which we touch on relatively briefly in Section 5 before concluding and summarising the key messages.

Before continuing it is worth defining a few key terms that we will use throughout the tutorial.

- **Probability of Correct Selection (PCS):** the probability that the option(s) output by the OvS algorithm is the true minimum. For a real problem, this must be estimated based on the simulation output but when testing algorithms on testbeds with known solutions, this is measured as the proportion of times that the algorithm finds the correct result.
- **Expected Opportunity Cost (EOC):** the estimated cost associated with choosing the wrong option. This can be a more practical objective to use as it distinguishes between cases where the cost difference between neighbouring options is small/large.

Both of the examples we use here and all of the algorithms that we introduce are included in a Github repository at https://github.com/TomMonks/ovs-tutorial and are free to access and download.

2 RANKING AND SELECTION

Ranking and selection algorithms are used when the number of options available is small, i.e. $\mathbf{x}$ can take only a few discrete values, and it is possible to sample at each of these values. The difficulty lies in the fact that the output of the simulation model is stochastic and that we have only a limited time or computational budget for experiments.

Assume that we are comparing m different options for the system set up, $i = 1, \ldots, m$. Each of these options has a true mean μ_i that is unknown to the experimenter and we wish to return the system that has the smallest μ_i. We obtain estimates of the μ_i by running the simulation model n_i times for each option i and finding the sample averages of the output variables $\mathbb{E}[Y(\mathbf{x}_i)]$. Ranking and selection algorithms aim to choose the n_i in such a way that the computation is as efficient as possible.

The original algorithms for solving these problems used two stages, e.g., see (Koenig and Law 1985) and (Chick and Inoue 2001), where in the first stage, each of the options is tested with the simulation model and in the second stage the number of subsequent observations of each system is optimised to ensure that either the PCS is guaranteed or, where the computational budget is fixed, the PCS is maximised. Two-stage approaches have the benefit that there is relatively little communication between the simulation model and the software running the ranking and selection and this can be beneficial when this communication is time-consuming. A further benefit of two-stage procedures is that they can be easier to program and use for non-experts in optimization, as discussed in (Monks and Currie 2018).

The majority of recent work in ranking and selection uses *sequential algorithms* in which the results of a simulation replication are fed back into the optimisation algorithm after each replication (or a small number of replications) to allow it to choose which option to test next. This allows the algorithms to react quickly to the output of the simulation model and hence improves their efficiency over the two-stage models. It is these algorithms that we focus on in this tutorial, considering first the KN++ algorithm introduced by Kim

and Nelson (Kim and Nelson 2006) and second the OCBA algorithm introduced by Chen and described in some detail in (Chen and Lee 2010).

There are many variations of the most basic single objective, single result problem and we do not have space to consider them all here. One which is particularly useful when helping with complex decisions is the *best subset selection* problem, in which the algorithm returns the top m options rather than just the top one. This allows a decision-maker to choose between several "good" solutions, allowing them to take into account factors that cannot be incorporated into the simulation model. In Section 4.1 we describe a sequential algorithm that can be used to solve this problem.

(Branke et al. 2007) compares different ranking and selection algorithms. They consider both OCBA and KN++, as we do here, but also include a *value of information procedure* or VIP. A VIP will allocate samples in such a way to maximize the expected value of information to be obtained from them. In the interests of space, we do not include these algorithms in this tutorial and would advise an interested reader to search for relevant references in (Branke et al. 2007).

3 INDIFFERENCE ZONE PROCEDURES

Indifference zone procedures provide a guarantee of finding the best system with a high probability $(1-\alpha)$ when the long-run average of the best system is at least δ better than that of the second-best system. Both α and δ are set by the decision-maker, with α typically being set to 0.05 or 0.10 to provide 95% or 90% confidence intervals respectively but it must be set such that $1-\alpha > 1/k$, where k is the number of systems. The indifference zone is defined by δ, the smallest difference in mean that would be significant to a decision maker.

We describe a sequential indifference zone procedure here, KN++, named after Kim and Nelson who describe their algorithm in (Kim and Nelson 2006). This extends the original *KN* algorithm, updating the variance estimator as more data are obtained, and is the standard indifference zone procedure, known for its efficiency but also its PCS guarantee. The procedure begins by running simulations for all systems and, as it proceeds, it will eliminate systems where the difference in their means and that of the best system exceeds some threshold, as described in the algorithm below. The procedure ends when only one system remains.

KN++ Procedure

1. Specify α, δ as discussed above; $n_0 > 2$, the number of replications of each system run during the initialisation step; η, the number of replications to make for each of the remaining systems at each step of the procedure.
2. Define I as the set of non-eliminated systems, $I \leftarrow \{1,\ldots,k\}$ and set the number of replications $n \leftarrow 0$. Set $\tau \leftarrow n_0$.
3. While $|I| > 1$
   (a) Run τ replications of each system in I. Set $n \leftarrow n+\tau$. Set $\tau = 1$.
   (b) Update the sample means $\bar{x}_i$ and sample variances σ_i^2 for each i in I. Set

$$\eta \leftarrow 0.5\left\{[2(1-(1-\alpha)^{1/(k-1)})]^{-2/(n-1)}-1\right\}$$

   and $h^2 \leftarrow 2\eta(n-1)$.
   (c) Find the difference in the means of all pairs of systems in I, $d_{ij} \leftarrow \bar{x}_i - \bar{x}_j$ for all $i, j \in I$ and $i > j$. Set $\varepsilon_{ij} = \max\{0, \delta/2n(h^2(\hat{\sigma}_i^2+\hat{\sigma}_j^2)/\delta^{*2}-n)\}$. If $d_{ij} > \varepsilon_{ij}$ then remove i from I, $I \leftarrow \backslash I\{i\}$ else if $d_{ij} < -\varepsilon_{ij}$ then remove j from I, $I \leftarrow I\backslash\{j\}$.
4. Return remaining system as the best.

4 OCBA

The basic idea of OCBA procedures is that additional replications are allocated to systems so as to best optimise the PCS. Effectively this means that the majority of the computational effort should be focused on competitive systems and little effort should be allocated to systems unlikely to be the best. This leaves us with a Catch 22 situation in that without carrying out all of the additional replications we cannot say exactly how the sampling will improve the PCS estimate and which systems are most important to simulate. However, it is possible to *estimate* the improvement and estimating the improvement in PCS from making additional samples with each system is the basis of determining how replications should be allocated. We refer readers to the excellent book by Chung Hun Cheng and Loo Hay Lee (Chen and Lee 2010) for the full technical details of OCBA and provide only the algorithm in what follows.

OCBA Procedure

1. Specify $n_0 \geq 2$, the number of replications to make of each system during initialisation; k, the number of systems/options being compared; η the number of replications to make at each stage; T, the total number of replications. For ease, set $T - kn_0$ to be a multiple of η.
2. Run n_0 replications for each design, set $l \leftarrow 0$ and $N_i^l \leftarrow n_0, i = 1, \ldots, k$.
3. Loop while $\sum_{i=1}^{k} N_i^l \leq T$.
   (a) Update the relevant statistics: sample mean, $\bar{X}_i = \frac{1}{N_i^l}\sum_{j=1}^{N_i^l} X_{ij}$; sample standard deviation, $s_i = \sqrt{\sum_{j=1}^{N_i^l}(X_{ij} - \bar{X}_i)^2/(N_i^l - 1)}$; $b = \arg\min_i \bar{X}_i$.
   (b) Allocate the new budget of η replications so that $\frac{N_i^{l+1}}{N_j^{l+1}} = \left(\frac{s_i(\bar{X}_b - \bar{X}_j)}{s_j(\bar{X}_b - \bar{X}_i)}\right)^2$ for all $i \neq j \neq b$ and $N_b^{l+1} = s_b\sqrt{\sum_{i=1, i\neq b}^{k}\left(\frac{N_i^{l+1}}{s_i}\right)^2}$
   (c) Perform an additional $\max(N_i^{l+1} - N_i^l, 0)$ replications for design i, $i = 1, \ldots, k; l \leftarrow l + 1$.

4.1 OCBA-M

The OCBA-M algorithm was introduced in (Chen et al. 2008) and has the objective of finding all of the top m designs for a simulated system. There are situations where a decision maker wishes to choose between a set of good designs, particularly where there are factors that cannot be included in a simulation model (e.g. political considerations). Top-m designs can also be beneficial as part of a global optimization routine where they are used to generate an elite set of solutions for optimisation in a subsequent stage.

OCBA-*m* Procedure

1. Specify $n_0 \geq 5$, the number of replications to make of each system during initialisation; k the total number of systems; m; η, the number of replications to make at each stage; T, the maximum number of replications where $T - n_0$ is a multiple of η; m the number of designs to include in the subset.
2. Initialise by running n_0 replications for each design, setting $l \leftarrow 0$ and $N_1^l, N_2^l, \ldots, N_k^l = n_0$.
3. Loop while $\sum_{i=1}^{k} N_i^l \leq T$.
   (a) Update statistics as follows. Calculate sample means for each system, $\bar{X}_i = \frac{1}{N_i^l}\sum_{j=1}^{N_i^l} X_{ij}$ and sample standard deviations $s_i = \sqrt{\sum_{j=1}^{N_i^l}(X_{ij} - \bar{X}_i)^2/(N_i^l - 1)}$, and compute $\hat{\sigma}_i = s_i/\sqrt{N_i^l}$ and $c = (\hat{\sigma}_{i_{m+1}}\bar{X}_{i_m} + \hat{\sigma}_{i_{m+1}}\bar{X}_{i_{m+1}})/(\hat{\sigma}_{i_m} + \hat{\sigma}_{i_{m+1}})$; update $\delta_i = \bar{X}_i - c$ for $i = 1, \ldots, k$.
   (b) Allocate computer budget as follows. The new computing budget increases by η and new replications are allocated so that $\frac{N_1^{l+1}}{(s_1/\delta_1)^2} = \frac{N_2^{l+1}}{(s_2/\delta_2)^2} = \ldots = \frac{N_k^{l+1}}{(s_k/\delta_k)^2}$. In practice, this can be achieved by setting $N_i^{l+1} \leftarrow N_i^l$, $i, = 1, \ldots, k$ and ordering the systems in non-decreasing order

of $\frac{N_i^{l+1}}{(s_i/\delta_i)^2}$. Until η observations have been allocated in this stage, add one to the system at the top of the list and reorder.

(c) Simulate an additional $\max(N_i^{l+1} - N_i^l, 0)$ for design i, $i = 1, \ldots, k; l \leftarrow l+1$.

5 WHAT NEXT?

Many problems that we consider in the OvS area are multi-objective and there are several ways of dealing with this. The review paper by Susan Hunter and co-authors (Hunter et al. 2019) provides an excellent overview of such multi-objective simulation optimisation problems and is designed as an advanced tutorial. They have a wider remit than just ranking and selection problems, considering problems with both integer-ordered decision variables and continuous decision variables. A concept used extensively in multi-objective optimisation is that of a *Pareto set*. We refer readers to (Hunter et al. 2019) for a formal definition and describe it here as a set of solutions for which one of the objective values is strictly better than any other that can be obtained feasibly, and no improvement can be made on any of the other objective functions by moving to a different feasible point. Identifying the Pareto set can then form the first part of a methodology in which the decision-maker either chooses from the full Pareto set, or obtains additional information from a decision-maker at interim points during the optimisation to guide it to the preferred solution.

Recent work in the area reflects the growing use of cloud computing in simulation for speeding up the generation of results. This area is still relatively new so it is likely that the classic algorithms have not yet been developed but several recent articles, e.g., (Pei et al. 2018) and (Kamiski and Szufel 2018), give an indication of how parallel computing can speed up these optimisation routines significantly.

A further area of research is designed to speed up optimisation by running initial tests on a low-fidelity model to obtain solutions that are likely to work well in the more complex simulation (and consequently the real system). In a recent article, (Xu et al. 2016), Xu and co-authors describe how the results of tests on low-fidelity models can be used to guide optimisation routines in the simulation models. The article refers to the use of digital twin technology and how optimisation can enhance this by not just mimicking the real system but suggesting better ways of running it into the future based on up-to-date input data. The process has some similarities to simheuristics (Juan et al. 2015), which are used to solve stochastic combinatorial optimisation problems. In a typical simheuristic algorithm, a metaheuristic is used to generate a set of good solutions for the deterministic version of the problem and these are then tested and refined using the simulation model.

6 SUMMARY

Our aim in this tutorial is to demonstrate the ease with which these sophisticated methods can be used in real simulations. The two key methods that we describe here and include in the Github repository are sequential methods, requiring constant communication between the simulation software and the optimisation algorithm. As discussed, this is sometimes not possible and we mentioned a possible alternative method (Monks and Currie 2018) that minimises these communications by going back to the two-stage models that were prevalent in the last century.

As discussed in the previous section, there are numerous variations on the original basic single-objective, single-solution problem, but understanding the algorithms for solving this original problem will provide an excellent base for understanding or indeed developing algorithms to solve the variations on the original problem.

In conclusion, if you need to guarantee the probability that you have selected the correct option for your system, use an indifference zone procedure such as KN++. If instead you have a fixed computational budget, OCBA is a good option for maximising the probability that you choose the correct option in a set number of simulation replications.

Finally, the Github repository (https://github.com/TomMonks/ovs-tutorial) is there to be used - please do so!

ACKNOWLEDGMENTS

Many thanks to the organisers of SW20 for giving us the opportunity to present this tutorial.

REFERENCES

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure". *Management Science* 53:1916–1932.

Chen, C.-H., D. He, M. Fu, and L. H. Lee. 2008. "Efficient Simulation Budget Allocation for Selecting an Optimal Subset". *INFORMS Journal on Computing* 20:579–595.

Chen, C.-H., and L. H. Lee. 2010, 07. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*.

Chick, S. E., and K. Inoue. 2001. "New Two-Stage and Sequential Procedures for Selecting the Best Simulated System". *Operations Research* 49:732–743.

Fu, M. C. 2015. *Handbook of Simulation Optimization*. 1st ed. New York, USA: Springer.

Hong, L. J., and B. L. Nelson. 2009. "A brief introduction to optimization via simulation". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 75–85. Piscataway, New Jersey: IEEE.

Hunter, S. R., E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia. 2019, January. "An Introduction to Multiobjective Simulation Optimization". *ACM Trans. Model. Comput. Simul.* 29 (1): 7:1–7:36.

Juan, A. A., J. Faulin, S. E. Grasman, M. Rabe, and G. Figueira. 2015. "A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems". *Operations Research Perspectives* 2:62–72.

Kamiski, B., and P. Szufel. 2018. "On parallel policies for ranking and selection problems". *Journal of Applied Statistics* 45 (9): 1690–1713.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Elsevier Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. Nelson. Waltham, MA: Elsevier.

Koenig, L., and A. Law. 1985. "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations". *Communications in Statistics - Simulation and Computation* 14:719–734.

Monks, T., and C. S. Currie. 2018. "Practical Considerations in Selecting the Best Set of Simulated Systems". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 2191–2200. Piscataway, New Jersey: IEEE.

Pei, L., S. Hunter, and B. Nelson. 2018. "A New Framework for Parallel Ranking & Selection Using an Adaptive Standard". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 2201–2212. Piscataway, New Jersey: IEEE.

Xu, J., E. Huang, L. Hsieh, L. H. Lee, Q.-S. Jia, and C.-H. Chen. 2016. "Simulation optimization in the era of Industrial 4.0 and the Industrial Internet". *Journal of Simulation* 10 (4): 310–320.

AUTHOR BIOGRAPHIES

CHRISTINE CURRIE is Associate Professor of Operational Research in Mathematical Sciences at the University of Southampton, UK, where she also obtained her Ph.D. She is Editor-in-Chief for the Journal of Simulation. Christine was co-chair of the Simulation Special Interest Group in the UK Operational Research Society until September 2013. Her research interests include simulation optimisation, mathematical modelling of epidemics, optimal pricing and applications of simulation in health care. See http://www.southampton.ac.uk/maths/about/staff/ccurrie.page for more details.

THOMAS MONKS is an Associate Professor of Health Data Science at the University of Exeter Medical School and The Institute of Data Science and AI, University of Exeter. He is also Turing Fellow at The Alan Turing Institute and Honorary researcher at the Clinical Operational Research Unit, UCL. Between 2014 and 2019 he led an applied healthcare Data Science team funded by the NIHR. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His speciality is simulation of healthcare systems of unscheduled and emergency care. He was co-chair of the UKs Simulation Workshop 2016 and 2018. His email is t.m.w.monks@exeter.ac.uk webpage is: https://www.turing.ac.uk/people/researchers/thomas-monks

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.005

A TUTORIAL ON INVOLVING STAKEHOLDERS IN FACILITATED SIMULATION STUDIES

Kathy Kotiadis

Kent Business School
University of Kent
Canterbury, Kent, CT2 7NZ UK

Antuela A.Tako

School of Business and Economics
Loughborough University
Leicestershire, LE11 3TU, UK

ABSTRACT

This tutorial introduces the PartiSim approach, aimed at supporting analysts and simulation modellers to carry out facilitated discrete event simulation studies. Facilitated simulation offers an alternative mode of engagement with stakeholders (clients) in simulation projects. It is particularly beneficial when modelling systems with complex behaviour, involving many stakeholders with plurality of opinions and objectives. PartiSim short for Participative Simulation, is a facilitated modelling approach developed to support simulation projects through a framework, stakeholder-oriented tools and manuals in facilitated workshops. A PartiSim study includes six stages, four of which involve facilitated workshops. PartiSim was developed more than 10 years ago. It can be applied to analyse operational problems in many contexts within the services and manufacturing domain. This tutorial presents the PartiSim framework and tools, some applications and example tools, a roadmap to adopting it and concludes with some tips for potential users.

1 INTRODUCTION

This tutorial describes the PartiSim approach to analysts and simulation modellers. PartiSim short for Participative Simulation, is a facilitated modelling approach developed to support analysts in involving stakeholders in the modelling process in a non-technical way. Stakeholders are engaged primarily in facilitated workshops to identify options and consider solutions through the use of simulation models. The approach was developed as part of a project funded by the UK's EPSRC back in 2007. PartiSim consists of a framework (Tako and Kotiadis 2015), tools and manuals (Kotiadis et al 2014, Kotiadis and Tako 2018) that support the analyst in carrying out modelling activities involving stakeholders throughout the project. Its framework, tools and manuals were developed and tested in two UK healthcare settings in the UK. Subsequently a toolbkit was developed including a user guide, tools and manuals in 2010 (Kotiadis and Tako 2010), which was updated in 2018. These are available for modellers to download for free from the PartiSim website (www.partisim.org).

The authors have trained modellers on using PartiSim, mainly in the UK through the UK OR society training programme and to the best of our knowledge it has been embedded on at least two occasions in the curriculum of an undergraduate business and a postgraduate engineering module at two UK institutions. Further applications have followed, three of which we are aware of and two are from different teams of analysts, who report in the academic literature on its use. For example, Proudlove et al (2017) report using a similar approach to PartiSim to undertake facilitated modelling in three health care projects. Philips (2017) used PartiSim to explore uncertainty and production smoothing in a complex pharmaceutical manufacturing environment. It was furthermore applied in a healthcare ambulance setting as part of a masters dissertation project (Puntambekar, 2016) under the supervision of one of the co-authors (Tako). The success of these studies varies, however, they all identify the benefit of engaging the stakeholders in conversations to co-develop options and solutions for their own problems.

More specifically using the PartiSim approach, tools and guidance available (Kotiadis and Tako 2010) the modelling team can benefit, not only because the activities set out can help the modelling team to make sense of the complexities involved in their settings, but also because it allows the modelling team to engage concurrently with all the stakeholders leading to common views and consensus being built in a transparent way at one meeting (workshop). It furthermore allows for the stakeholders to be part of the process and the solutions identified, while at the same time non-technical language is used to extract their views. The dedicated tools supporting each workshop allows for a more structured and leaner modelling process throughout the study, compared to studies where the modeller is developing the model on his own and checks or validates the model with individual stakeholders on a one to one basis. The suggestions and tips available in the tools and manuals for the facilitator to use enable better communication with the stakeholder group rather than making up the questions on the spot. Undertaking the simulation study in a participative way can help save time in building the model on the computer, mainly because the workshops enable a common understanding between the modeller and stakeholder team on what should be included in the model, as well as commitment and quick access to the data needed to develop the model.

Based on our experience of developing and using the PartiSim framework, this tutorial aims to guide the analyst in using the PartiSim framework and tools in their participative simulation studies. Section 2 provides an overview of the PartiSim framework, including the activities and tools used to support each stage of the simulation study. Section 3 illustrates applications of PartiSim in real life studies, based on our experience of using it and discusses the outcomes of these studies. Section 4 introduces three example tools used in PartiSim workshops to give modellers an insight of how they are used in practice. Section 5 provides a roadmap of the journey that the modelling team should take at an individual and team level in adopting the approach. Section 6 concludes this tutorial with some practical tips for using the PartiSim approach and its tools for potential adopters.

2 OVERVIEW OF THE PARTISIM FRAMEWORK & TOOLS

The PartiSim approach is designed to support the modellers' interaction with a group of stakeholders throughout the DES study lifecycle. A framework and tools support the modeller in undertaking the different modelling activities during a simulation study. The framework, outlined in Table 1, consists of six key stages and/or five sub-stages (column 1); each includes a number of prescribed activities (column 2), tools (column 3) and corresponding stakeholder-oriented deliverables (outputs) (column 4), which enable participative DES modelling to take place.

The main PartiSim stages include: 1. Initiate simulation study; 2. Define Problem; 3. Define conceptual model; 4. Model Coding; 5. Experiment with model; 6. Implement Findings (Tako and Kotiadis 2015, Kotiadis and Tako 2010). The sub-stages support the main stages, either to prepare for the workshop-based stages or to tidy up outputs developed in workshops and confirm these with the stakeholders. Model coding, a middle stage in PartiSim, is not undertaken in a facilitated mode and that is acceptable practice in facilitated DES (Robinson et al 2014).

The aims of each stage (and sub-stage) are achieved by undertaking the prescribed dedicated activities (Table 1, column 2), which are distinguished in two types: modelling and workshop activities. The modelling activities are aimed at supporting the modelling process while workshop activities support the facilitation of the group of stakeholders. The activities for the sub-stages are mainly undertaken by the modelling team, who report back to the stakeholders the outputs agreed in the workshops or seek further reflections and clarifications. Some activities such as those undertaken in stage 1 and mostly in the sub-stages are generic in nature and related mainly to organising the simulation project or liaising with the stakeholder team. They could be used in any type of analysis carried out in a facilitated mode. Other activities are adapted or borrowed from Soft Systems Methodology (Checkland 1999). For example, the activity "Define system & boundaries" (stage 2), involves decomposing the system into the activities that take place in that system. Traditional DES modelling activities are adapted to be carried out in a facilitated environment, giving stakeholders the space to express their preferences and discuss

alternatives. For example in the "Debate desirable and feasible solution space" activity (stage 5) the results of relevant scenarios are presented and debated with the stakeholders.

Each stage is supported by tools and the associated manuals which support the modelling team and stakeholders to reach to the prescribed dedicated outputs for each stage (column 3, Table 1). Scripts are also available for some of the stages, aimed mainly at the facilitator. These are different from the tools or manuals in that they include advice to support the facilitation process for activities that do not require any specific tools to be used. These are paper based and freely available on the PartiSim website (www.partisim.org).

Most of the activities support the development of the intermediate deliverables or outputs (Table 1, column 4). They are called intermediate because they can be revised or converted into a different output in the next stage. Some, for example "A bounded system within which the problem to be addressed exists" (sub-stage 2.a), are developed in a sub-stage with the view to using and leading the discussion during the workshop in stage 3. While others such as the conceptual model (stages 2 and 3), are developed during the workshop, but refined during a sub-stage (3.a) and converted into a different output (a simulation model) in stage 4.

Table 1: The PartiSim Framework, including stages, activities, tools and outputs

Stage & purpose	**Activities**[1]	**Tools**	**Outputs**
1. Initiate Study <u>Purpose:</u> Identify stakeholder team Identify key problem situation(s)	The modelling team undertake: - informal meetings and/or - on-site observations and/or - one-to-one interviews - with project champion and key stakeholder(s), to address preliminary information needs	- Feasibility of simulation modelling and its use Script - Situation of Interest Tool with manual - Recording Observations Tool with manual - Bank of questions Script - Stakeholder details Tool with manual - List of reading materials Tool with manual	List of stakeholder team roles. Preliminary understanding of the problem situation Study proposal, incl. initial study aims and timescales
1.a Pre-workshop (Sub-stage) <u>Purpose:</u> Preparations for workshop 1	- Identify modelling team and stakeholder team roles. - Modelling team prepare preliminary materials to be used in workshop 1 - Decide workshop venue and time slots. - Stakeholders are invited to workshops		
2: Define the Problem (workshop 1) <u>Purpose:</u> Agree on the problem situation and the wider system, within which it exists.	*Agree problem statement* *Define the system* *Draw a system model*	- Define the system Tool with manual - Draw the System Model Tool with manual	Overall study objectives/aims System map

2.a Post workshop1/Pre-workshop 2 stage <u>Purpose</u>: Disseminate workshop 1 outputs and prepare for workshop 2	Modelling team re-draw tools & disseminate workshop outputs to stakeholders Prepare preliminary materials for use in workshop 2		
3. Define conceptual model (workshop 2) <u>Purpose:</u> Define specific elements of the conceptual model	Participating stakeholders take part in a facilitated workshop process to: - *Brainstorm study objectives* - *Draw the Performance Measurement Model (PMM)* - *Define simulation study objectives* - *Draw communicative model* - *Discuss data collection*	- Performance Measurement Model (PMM) with manual - Study objectives Tool with manual - Communicative Model Tool with manual2018	Model inputs, outputs and contents Simulation objectives Process flow diagram A list of data requirements
3.a Post workshop 2 (sub-stage) <u>Purpose:</u> Disseminate workshop 2 outputs and refine conceptual model	Modelling team: - Prepare report detailing Refined workshop outputs and Data requirements - Liaise with the stakeholder team over correctness of workshop 2 outputs.		
4. Model coding <u>Purpose:</u> Conceptual model is converted into a computer model	- Data collection (modeller and stakeholders) - Build simulation model on the computer (modeller)		Model results Model validation and verification Preliminary future scenarios
4.a Pre-workshop 3 sub-stage <u>Purpose:</u> Preparations for Workshop 3	- Prepare preliminary materials for use in workshop 3 (stage 5): • Liaise with the project champion over correctness of model & its results (modeller and project champion) • Review preliminary scenarios with project champion • Prepare preliminary materials for use in the next workshop		
5. Experimentation stage (workshop 3) <u>Purpose:</u> Define alternative	Stakeholders are invited to: - *Validate the simulation model & its results* - *Rate performance measures (linked to model results)*	- Model validation tool - Rating the Performance Measures tool (or	Model validation and verification Alternative future scenarios

scenarios to experiment with model	- *Debate desirable and feasible scenarios*	VISA) with manual - Debating the Alternative Scenarios tool with manual	
5.a Post-workshop 3/ Pre-workshop 4 sub-stage <u>Purpose:</u> Refine alternative scenarios & prepare for workshop 4	Modelling team: - Tweak or correct simulation model - Implement additional scenarios suggested (based on stakeholder feedback from workshop 3.) - Liaise with the stakeholder team over correctness of model results - Prepare preliminary materials for use in workshop 4		New alternative future scenarios Revised simulation model Revised model results
6. Implementation stage (workshop 4) <u>Purpose:</u> Define an implementation plan	Stakeholders are invited to: - *Review learning & changes implemented* - *Risk analysis and feasibility of change* - *Agree action trail*	- Script for Identifying changes in the system - Feasibility and Risks Scale tool with manual - Barriers to Change tool with manual - Action and Communication Plan tool with manual	Agreeable and feasible scenario(s) to be taken forward Action plan with deliverables (including due date and person responsible)

[1] Activities in italics are workshop activities

3 APPLICATIONS OF PARTISIM

In this section we refer to some real life applications in which the PartiSim framework and tools have been used in practice. All three applications happen to be in health care in light of the authors' industry contacts and opportunities for collaboration. These are the Obesity (Tako et al 2014), Colorectal and Ambulance Service study. As noted in the introduction, there are more adaptations of PartiSim by other teams, however we concentrate here on the studies we have had direct experience with. A brief summary of each study follows.

The obesity study involved a newly set up service that provides services for London and Northern Ireland, offering three types of treatments: lifestyle treatment (i.e. advice on diet, exercise and behavioural change), pharmacotherapy (administration and management of weight loss medication) and bariatric surgery (also known as obesity surgery). The later involved three main types of surgery: gastric band, sleeve gastrectomy and gastric bypass. The service providers wanted to understand how to configure their resources (i.e. surgeons and physicians) in order to consistently meet the 18 week target in the foreseeable future, without adding unnecessary capacity, by employing new resources such as surgeons and physicians. At the time of the study (2009), the service was experiencing increasing numbers of referrals and an increased pressure to meet the demand for consultation and treatment. The pressure was mostly experienced in the parts of the system providing pharmacotherapy treatment and surgery. The service

referrals were increasing each year at an exponential rate which made planning difficult. For more details see Tako et al (2014).

The Colorectal study involved an outpatients surgical care service at a UK NHS Hospital which at the time (2009) had been experiencing increased demand for its clinics due to a then recently launched bowel screening programme. In addition, stakeholders believed that some patient categories, particularly those categorised as less urgent may have excessive waits during their journey along the colorectal cancer care pathway. The surgical service was offering out-of-hours outpatient clinics and colonoscopy tests in order to meet the increased demand and reduce the proportion of patients breeching waiting time targets. The stakeholders were interested to gain a better understanding of the demand for services and the existing levels of resource available i.e. staffed time for clinic appointments, colonoscopy examinations and surgery. The study explored the impact of introducing improvements to the colorectal pathway through a combination of re-organising and/or increasing the levels of some resources (e.g. clinic slots) on the performance of the clinic in terms of the size of the waiting lists and the proportion of patients breaching Department of Health targets (2 week, 18 week etc). The Obesity and Colorectal study were undertaken at the time of developing the PartiSim approach on a pro-bono basis, with the view to testing the tools and process.

The Ambulance Service (AS) study involved an NHS ambulance service Trust that provides pre-hospital emergency and urgent care services and patient transport to a specific local area population (Puntambekar, 2016). As with all UK's NHS services, this particular AS faced high demand levels for its services especially in the winter months, which in turn increases the pressure on the service to deliver safe care to patients within the required response time targets. At the time of the study (2016), the specific service was interested in improving the efficiency of its call cycle by reducing its overall call cycle times and the number of patients conveyed to emergency departments when not needed. Policies such as providing advice over the phone (hear and treat), treating patients at the scene (see and treat) and taking patients to alternative non-hospital destinations, such as urgent care centres, were being introduced and the service was keen to understand the impact of these changes on the AS performance. Clinical advisors had been hired by the AS Trust to provide Hear and Tread services over the phone to patients. In order to deliver valued analytical support to the AS a facilitated modelling approach was undertaken involving stakeholders from the AS throughout the study. The project was undertaken as a masters consultancy project on a pro-bono basis and one of the authors supervised the project and facilitated the workshops (Puntambekar, 2016).

All three studies followed the same PartiSim process and tools. The first two were used as case studies to test the tools developed, whereas the last was utilised by a novice modeller (masters student) to provide consultancy services as part of the summer project. The models developed represent mainly queuing systems of patients (or patient calls), which were amenable to modelling using a discrete-event simulation approach. Due to space limitations the models developed are not provided in this paper, however these were presented to and discussed with the relevant stakeholder teams. A summary of the key characteristics of these studies can be found in table 2 below.

Table 2: A summary of the key characteristics of the Obesity, Colorectal and Ambulance Service studies.

	Obesity study	Colorectal study	Ambulance service study
Stakeholder participation	Multidisciplinary	Multidisciplinary although surgeons accounting for majority	Mainly paramedics and clinical team mentors (CTMs) attended workshops. Strategic Innovation Programmes Manager (project champion) was also involved, but unfortunately did not

			attend the workshops.
Simulation Study objectives	To explore: • reducing the waiting list for a number of clinics in the pathway • reducing the number of beds required in post op care • the achievement of the 18 week target for referrals	• To understand the patient pathway • To explore reducing patient throughput time	• To identify ways to improve the efficiency of the ambulance service call cycle by increasing the percentage of: • Hear and Treat calls • Sea & Treat cases • cases conveyed to Alternative care providers
workshop involvement	4 workshops (average duration 2 hours) Most meetings took place in a hospital meeting room	4 workshops (average duration 2 hours) Most meetings took place in external conference room	4 workshops (average duration 2.5hrs) Most meetings took place in a seminar room at Loughborough University.
Action resulting from study	More operating slots and decision to build new obesity surgery operating theatre	Decision to introduce a new process in the care pathway	Agreement that the service should increase involvement of the clinical assessment team in the call cycle to provide advice over the phone.

In all three studies stakeholders engaged well with the process and the tools used as part of the workshops, interacting either with the facilitators or each other. The stakeholder team participating in the workshops of the three studies was different. In the obesity study the group comprised of many different specialties. The divergence is less in the colorectal study and far less in the AS study where only front end staff that went out in ambulance calls attended the workshops. Their managers were reluctant to attend the workshops as they were worried this would affect the free expression of views among their staff. Despite the efforts of the modelling team to include also members outside the organization, such as clinical staff from the associated emergency departments in interconnected hospitals, this was not considered suitable from the AS management for confidentiality purposes, hence not pursued. Contrary to the first two applications the project champion in the AS study was not able to attend any of the workshops, which meant that the support experienced in the previous two studies during the workshops was not present. Despite not having attended any of the workshops, the AS study project champion supported the study and the modeller fully throughout.

Considering, the conversations that took place in the workshops, participants were fully involved and contributed enthusiastically in all the tasks when invited. It is observed that there were more heated

discussions among the participants of the obesity study than the colorectal and the AS study. This is likely to be related to the consistency of the group of participants, which included managers (non-clinicians) and nurses with more differences in their experience of the system and therefore their thinking. However, all arguments were resolved within the workshop and in all three studies the stakeholders gave equal praise to the modelling team about the overall experience and the knowledge gained as a result of these workshops.

All three studies reached to a consensus about the action to be taken as a result of discussions taking place within the workshops, however the level of implementation differs between the different studies. In the first two studies, the project champion met with other stakeholders outside of the workshops in an effort to push forward action. In both cases this took place between the third and the fourth workshop. The modelling team was not aware of these meetings until the fourth workshop. This turn of events was surprising to the modelling team, we however believe that the project champions, which in both cases were powerful and influential, were motivated by the knowledge gained and discussions conducted during the experimentation workshop. On the contrary, the project champion, filled in by the strategic innovation programs manager in the AS study, was not able to attend any of the workshops. He was however equally supportive of the study outside the workshops, who met with the modeller to discuss data input needs, validate the model and propose scenarios. After having read the stakeholder report post workshop 3, he/she was very enthusiastic about the findings and as a result arranged for the modelling team to present the results of the study to the Board of Directors of the AS trust. Despite the results being received enthusiastically by the service and its management team, the year after (2017) a re-organisation of the call cycle and the way time targets are counted throughout the service was centrally introduced. This change meant that it took away the attention of the service from the simulation study, hence we are not clear about the outcomes of the study, however the modelling team is aware that the AS continues to make use of the clinical assessment team to provide hear & treat care to patients. Hence we can conclude that the outcomes of all three studies were positive and that the process undertaken and discussions that took place at the workshops played an important role in generating ideas and reaching consensus, which may have not been possible if we were to speak to stakeholders individually. We next provide some examples of tools used in PartiSim workshops.

4 EXAMPLE TOOLS

This section provides some examples of tools used in PartiSim stakeholder-oriented workshops to give the reader a feel of the process followed and the facilitation at the workshops. We choose three tools, draw a system model (part of Workshop 1), rating the performance measures (part of workshop 3) and Analyzing risks and feasibility of change (part of workshop 4) to give the reader an insight of how they work in practice.

4.1 Draw a system model

The system model consists of a graphical representation of the key activities occurring in the system of interest. It is completed as part of the third and last activity in workshop 1 (Define the conceptual model), after the problem statement activity and define the system with the stakeholders takes place. The Draw the System Model Tool (Figure 1) and manual can be utilised, which consists of paper-based tools that stakeholders complete during the workshop with the facilitator's support.

The process of developing a System Model consists of collecting the verbs that describe the activities that take place in the care system, based on the logical dependencies involved (Checkland and Scholes, 1999). We group the key activities that take place in healthcare systems, into three generic categories: clinical/operational, managerial and research. The clinical/operational part can be a closer representation to the computer model, depending on the problem situation studied (Kotiadis and Robinson, 2008). Whereas, the research and managerial parts are considered useful in order to enrich the understanding of the operational (clinical in health care settings) needs leading to a better model. The facilitator can find

guidance and tips in the accompanying manual for this tool, such as questions to be directed to the participants while using the tools. The process of designing the Care System Model (CSM) with the stakeholders helps to gain further insights about the problem situation by both stakeholders and the modelling team.

An example of the tool completed at the workshop for the obesity study is presented in Figure 1. This exercise served as means of bringing out some additional problems and inefficiencies involved in their obesity system that had not emerged during the problem statement activity. Concerns were raised regarding inefficiencies present in the care system such as patients wrongly being referred to some clinics resulting in long waiting lists. Stakeholders were then asked to identify interrelations between the three groups of activities (managerial, clinical and research). For example, the managerial activity "Design and set up patient group forum" is connected to the clinical activity "Provide group forum for patients" in Figure 1.

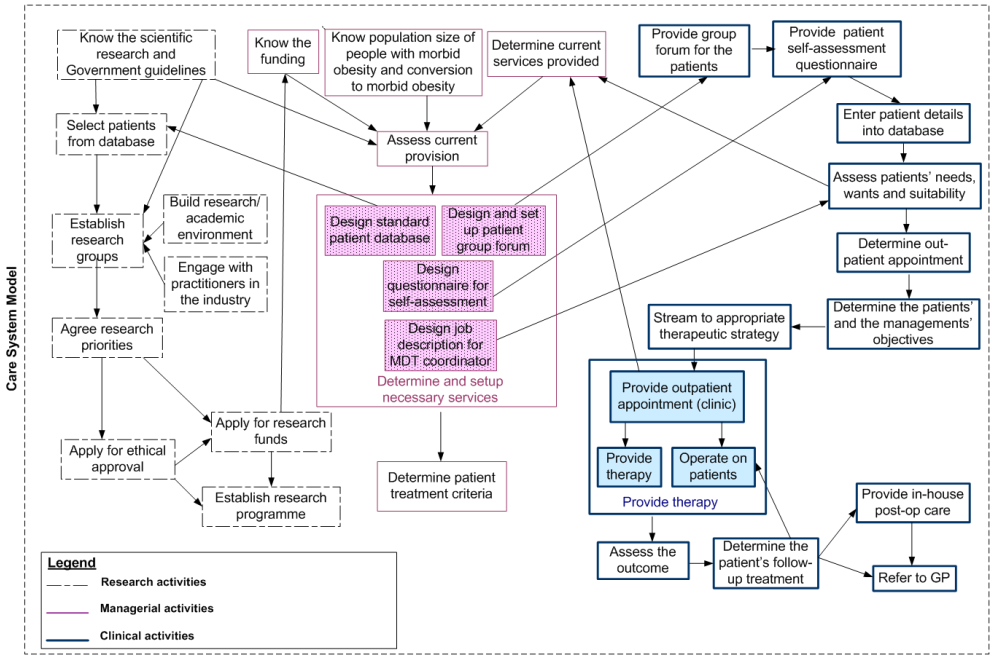


Figure 1: A System Model representing the research, managerial and clinical activities in the obesity care system

4.2 Rating the performance measures

Rating the performance measures is the second activity in workshop 3 (Experimentation stage, Table 1). Performance measures are the key model outputs. The aim of this activity is to get the stakeholders to focus on the most important measures (model outputs), which are then subsequently used to ultimately narrow the solution space of the scenarios. They have been initially identified in workshop 2 as part of defining the conceptual model. After the simulation model is developed, it is brought to Workshop 3 for the stakeholders to validate, including the model outputs. In this activity, the participants are guided through a process to identify and negotiate the importance attached to each performance measure.

The activity is guided by the Rating the Performance Measures Tool and its manual (Column 2, Table 1). This tool is based on multi-criteria decision analysis (MCDA) (Belton and Stewart 2002). It is available as paper-based tool and as a software tool, such as VISA software (http://www.visadecisions.com), for which one needs to have a license. This tool consists of a value tree representing model results (performance measures) and the weight in terms of importance attached to each one by the stakeholders. An example of a value tree developed for the obesity study (Table 2), using

the VISA software is presented in Figure 2. In this case, the value tree was set up prior to the workshop using the performance measures that were identified in the previous workshop (workshop 2) and during model coding. It should be noted that the modelling team had only recently started to learn and use VISA and to avoid any unexpected technical hitches and subsequent delays, prepared printouts of VISA outputs in advance. Nevertheless, the VISA software has the potential to be used live, if the modelling team is familiar with using it. The benefit of using VISA live in the workshop lies in that the results of different scenarios, can be connected with the agreed value tree in order to evaluate each scenario and to identify the most desirable and feasible scenario/(s). This is because the VISA software is compatible with the simulation software we used (www.simul8.com) to develop the DES model. It is also possible to rate the performance measures using the paper-based tool, without the VISA software and/or anonymously, as explained in the PartiSim User Guide and Toolkit (Kotiadis and Tako 2010). In the subsequent two studies (Colorectal and Ambulance service) the paper-based tool was used instead, in the former for the purpose of trialing the tool and the latter because the modeller did not have access to the VISA software.

At the workshop in the obesity study, the facilitator started the activity by asking the stakeholders to express their opinions about the importance of each performance measure, by weighing each one on a scale from zero to one hundred (Figure 2). During the validation part of the workshop, it had already become clear that the waiting lists were of high importance to all stakeholders, especially for the pharmacology and surgery clinics. The stakeholders on the whole agreed with the weights assigned prior to the workshop (Figure 2) so no changes were needed. Subsequently, the stakeholders moved on to the next workshop activity to debate desirable and feasible scenarios based on the performance of the scenarios of interest for the most important outputs.

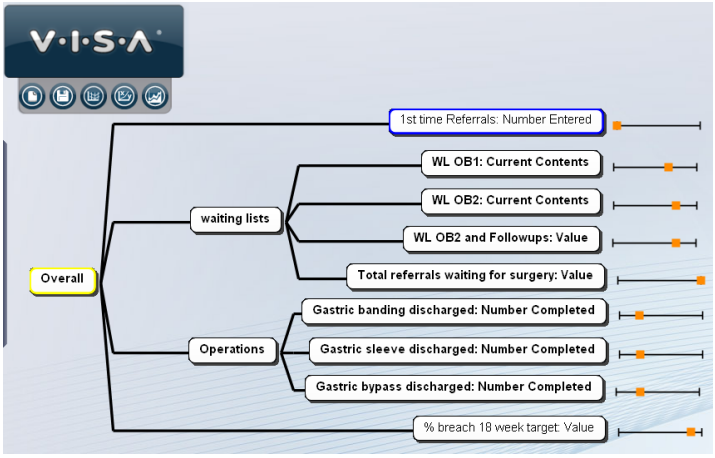


Figure 2: Value tree rating performance measures of the obesity system using VISA software

4.3 Analysing risks and feasibility of change

Analysing risks and feasibility of change is the second activity in Workshop 4 (Implementation Stage), after a discussion where the learning and changes that may have been introduced so far is reviewed. This workshop activity focusses on the scenario identified as most desirable, based on it achieving the highest performance for the most important performance measures (model outputs). The Feasibility and Risks Scale Tool is used and the aim is to narrow the solution space to ideally one scenario that could be implemented, by identifying the factors that may hinder implementing the changes linked to the chosen scenario, with the view to weighing up the feasibility of the scenarios chosen. It is recognised that factors such as psychological perceptions may hinder the stakeholders from taking action (Ajzen 1991).

At the workshop in the obesity study, out of the six scenarios explored the third scenario was the best performing for most performance measures. This was also the most preferred scenario by all stakeholders.

The facilitator started this activity by asking the stakeholders to consider how this scenario could be put in place and hence the inhibiting factors were discussed.). The Feasibility and Risks Scale Tool (Figure 3) and its manual are used to identify the reasons for which this scenario was feasible and the reasons for which it was not feasible. All stakeholders were encouraged to contribute to the discussion. The facilitator put forward two columns, one for reasons supporting the feasibility of the scenario and the other for reasons against it and recorded on a flipchart. The points made were listed and the scale was constructed by drawing a sloping line, dipping in this case on the not feasible side of the scale.

As a result of this process, Scenario 3 was deemed to be not feasible in the short term because of the timescale of adding new resources in the real system. In the real life system, a delay of a few months in introducing the additional resources would not guarantee its results. As the admissions and waiting lists in the real system would be increasing it would take longer to reach equilibrium in the system, where key targets are not breached. An example of the discussion that took place at the workshop is shown below, where physical space was identified as an issue for implementation of the scenario:

Stakeholder A: I don't think this is working. I think this system internally, for us, having a third surgeon here, the third surgeon, the issue is not really physically, in terms of surgery, it's a case of space.

Stakeholder B: Beds and space.

Project Champion: We've assumed the space will just magically appear.

<Laughter>

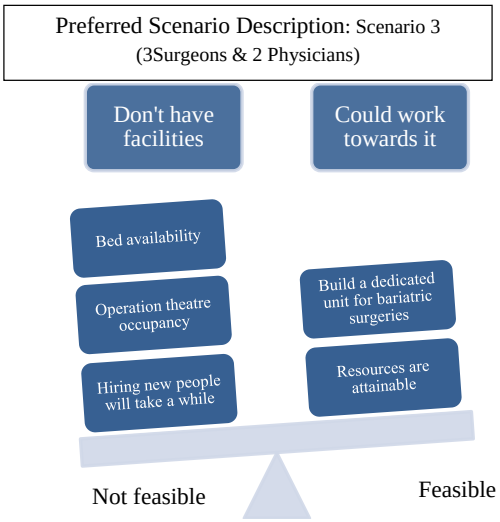


Figure 3: Example of using the feasibility and risk scale Tool to analyse a scenario

As a result of this analysis, it was accepted that scenario 3 was not feasible mainly due to timing issues. A number of other scenarios was discussed till a scenario considered feasible by the group was identified, before moving on to the next workshop activity.

We next provide a roadmap of the journey that potential users interested in adopting PartiSim in their project should be undertaking.

5 ROADMAP TO ADOPTING PARTISIM

In this section we explore the process of adopting PartiSim and of undertaking a facilitated modelling mode in a DES modelling project. This guidance alongside the PartiSim materials freely available will

support a modeller or a modelling team to change their practice from an expert to facilitated mode of DES practice. The adoption of PartiSim can be considered at two levels: the individual and the team level. This means that effort is required from individuals within a team, as well as the whole team, in order to become competent in undertaking PartiSim as part of an intervention. We next consider each level separately.

The individual level training can be undertaken by members of the modelling team such as those taking on the role of a workshop facilitator or simulation modeller. It is advised that all those in the modelling team embark on this individual development prior to coming together as a team. This could be considered as an ongoing 4-stage loop (left loop, Figure 4) with each iteration making the individual reflect on their knowledge and experience and thus taking on a continuous improvement plan at a personal level. The questions asked should include: "What did I do well?" and "What should I have done differently to engage clients?"

In our experience of PartiSim we have found that in each intervention we have gained experience and enlightenment leading to better practice in subsequent case studies. Moving from expert mode to facilitated mode is an ongoing journey of personal development. Hence the loop starts and ends at the same point, with reflection (left loop, Figure 4). At the personal level the individual should engage with the framework and tools prior to each intervention in order to familiarise him/herself with the content taking into account all the updates to practice. Indeed the development of the PartiSim website by the authors is aimed at providing a knowledge base of up-to-date practice and all teams engaging in PartiSim are encouraged to contribute to its ongoing refinement and development.

We acknowledge that workshop facilitation is an art that requires ongoing refinement and individuals looking to take facilitation roles are encouraged to update their competencies though reading or practice on an ongoing basis (see bottom activity of left loop, Figure 4). The art of facilitation extends beyond simulation (Robinson 2014, Tako and Kotiadis 2015) and OR (Franco and Montibeller 2010, Taket 2002, Ackermann 1996) to other fields (Kaner 2007) and is constantly evolving. The DES community has a lot to learn from the research into facilitation led by the Problem Structuring Community (also known as soft OR community largely based in the UK and Europe) in OR. Other areas that should also be considered is that of behavioural OR, a newcomer to the field, that concerns itself with how groups interact with models and the modelling process, providing research and understanding that could feed into the facilitated and participative DES practice (Franco and Hamalainen 2016). Following on from updating and developing competencies the individual is encouraged to take part in an actual intervention. At that stage one enters the next 5-stage loop, the PartiSim team development (right loop, Figure 4), with "Apply PartiSim to real case study". This is discussed in more detail in the next paragraph. The point made here is that one cannot be fully proficient in PartiSim unless they engage in real practice. The first time an individual undertakes the loop, he/she should be encouraged to consider their journey as a learning experience where improvements and adjustments will be necessary in future applications.

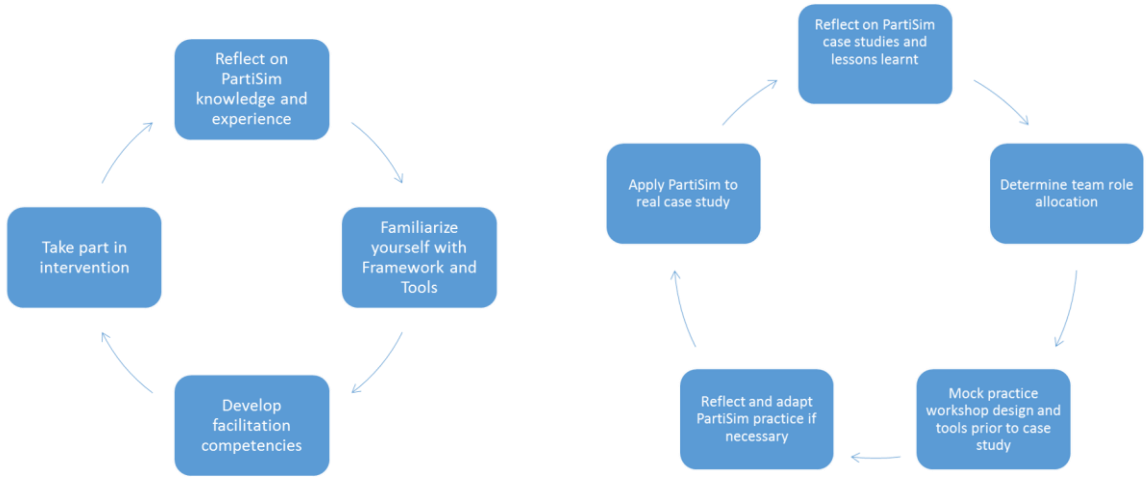


Figure 4: Personal (left) and team (right) development for adopting PartiSim in DES interventions

Now we consider the process that could be followed by a team adopting PartiSim (right loop, Figure 4). Similarly to an individual's development journey towards PartiSim it is advised that a team is formed at the beginning of any intervention. In the very first loop the modelling team should hold a PartiSim awareness event where existing literature is discussed and any concerns and issues are raised with a view that all the team have a good initial grasp of the process, guidance and tools before commencing practice. At this point it could be that some individuals within a team commence their personal journey (left loop, Figure 4) although it would be better if that has taken place to some extent before the team meet.

The PartiSim framework identifies the roles that will enable the delivery of the simulation study, from both the stakeholder and modelling team, to include roles such as the facilitator, modeller and recorder, but also key stakeholders, project initiator and project champion (Kotiadis et al 2014). The project champion comes from the stakeholder team; he/she has good communication and interpersonal skills to create awareness, confidence and consensus, but has also authority and influence within the organization to build up commitment to the project. At the end of the study, they can in turn support and ensure the delivery of implementation plans agreed at the end of the study. Ideally we would suggest that the modelling team embarking on a change in practice should have at least one stakeholder (ideally the project champion) involved in this early reflection stage in order to get feedback on the process. Having familiarised themselves with PartiSim, the modelling team should discuss the roles that they are prepared to trial in the first loop. Modellers that are confident communicators should consider developing their skills in facilitation but equally if the skillset is not currently present within a team additional members, possibly outside of DES modelling, could be sought. Obviously, at this point, it is expected that a prospective intervention has been already identified and the team would be preparing for the first workshop. In our experience we found that holding mock practice workshops without the actual stakeholders (the modelling team and/or other externals to the intervention acting as stakeholders) helped improve the flow of the actual workshop. For example, at this stage an experienced facilitator should engineer opportunities for others in the team to trial facilitation in small time chunks (e.g. 30 minutes) as part of the team's training and development.

Having embarked on mock workshops the modelling team should hold a debriefing to reflect on the workshop process, flow and duration with the view of adapting practice to their strengths for the real application. The allocation of roles and development of competencies within the modelling team, should also be reconsidered. Following this, the modelling team should be ready to engage in a real application. At the end of the intervention a meeting should be held by the modelling team to reflect on workshops, roles and competency development with a view to improving practice in the subsequent loop/application. Given that modelling team membership may deviate from one intervention to another it is advised that

modelling teams consider the loop for PartiSim team development (right hand loop, Figure 4) for each application. PartiSim is just as much about the collaborative approach within the modelling team as it is between the modelling team and the stakeholder team during the intervention.

6 PRACTICAL TIPS FOR USING PARTISIM

We conclude the tutorial with some additional practical tips for using the PartiSim approach and its tools for potential adopters of the approach to consider, as listed below:

- Identify from the outset of the study whether the stakeholder team are willing and/or need to be involved in the study. If dealing with a complex problem, where people in the system hold different opinions and contradicting views about the problem, with little communication amongst teams, a participative study would be suitable.
- It is beneficial that the membership of the stakeholder team is consistent throughout the study to ensure that there is continuity in the outputs and learning from one workshop to the other. For this reason an agreement from the beginning of the study should be made with participants to commit to attending all four workshops and dates agreed in advance if at all possible. A good way to incentivize good participation is to create a good rapport with the stakeholder group and to offer opportunities for informal chats at breakouts, i.e. coffee/lunch breaks.
- Being flexible and willing to accommodate stakeholder requirements. In all three studies discussed in section 3, we have found that working with clinicians and healthcare staff, with high risk responsibilities and busy schedules we have had to make a conscious effort to keep workshops duration as short as possible and accommodate workshops around stakeholders' commitments. Some examples include being flexible on the start time (e.g. 7 am) and location (e.g. hospital meeting room) of workshops to suit stakeholders' busy schedules.
- Besides keeping workshops as short as possible, ideally approximately two hours, we also recommend leaving time between workshops, between 2-4 weeks to give time to the modelling team to summarize workshop outputs, prepare for the next workshop, collect data or information required for the model, etc. This time is also beneficial for the stakeholder team to let ideas sink in and come up with fresh ideas in subsequent workshops.
- From the modeller's perspective, being able to apply the PartiSim approach effectively, one needs to be prepared and open to deploying a multi-paradigm approach, meaning moving between the soft and hard paradigms between the different activities (Tako and Kotiadis, 2015). For novice modelers or those more familiar with the hard paradigm, this can mean being consumed by the model and its results rather than focusing on the client interaction and the process (a framework, its stages and outputs). More details about how each paradigm is deployed at each PartiSim stage is provided in Tako and Kotiadis (2015). It is beneficial to be familiar with Soft Systems Methodology (SSM) (Checkland 1999) and more generally the problem structuring field.

DES modellers and analysts are invited to carry out a PartiSim study in their simulation projects and reflect on the facilitation skills needed to develop. We believe that using the overall framework and tools is especially useful for novice modellers and those looking to develop their facilitation skills by undertaking the journey described in the roadmap (section 5). The PartiSim materials, user guide, tools and manuals are available for interested modellers to access for free from our website (www.partisim.org).

REFERENCES

Ackermann F (1996). Participant's perceptions on the role of facilitators using group decision support systems. Group Decision and Negotiation 5 (1): 93–112.

Ajzen, I. (1991). "The theory of planned behaviour." Organisational Behaviour and Human Decision Processes 50: 179-211.

Belton, V. and T. J. Stewart (2002). Multiple criteria decision analysis: an integrated approach. Dordrecht, Kluwer.

Checkland, P. (1999). *Systems thinking systems practice*. Chichester, Wiley.

Checkland P and Scholes J (1999). Soft Systems Methodology in Action. Wiley. Chichester. UK.

Franco, L. and Hamalainen, R., 2015. Behavioural operational research: returning to the roots of the OR profession. European Journal of Operational Research, 249(3), pp. 791-795.

Franco, L. and G. Montibeller (2010). "Facilitated modelling in operational research." *European Journal of Operational Research* 205: 489-500.

Kaner S (2007). *Facilitator's Guide to Particpatory Decision Making*. San Francisco, CA, Jossey-Bass.

Kotiadis K and Tako AA (2018) Post-model coding facilitation in DES: A case study in Healthcare. European Journal of Operational Research.

Kotiadis K and Tako AA (2010) PartiSim User Guide to Facilitation Version 1 (December 2010), DOI: 10.13140/RG.2.1.3659.1201

Kotiadis K, Tako A, Vasilakis C, (2014) 'Participative and Facilitative Conceptual Modelling'. Journal of the Operational Research Society, 65 (2), 197–213.

Phillips C. (2017) 'Participative simulation and collaborative model design used to manage uncertainty and align production with demand', Working paper, Bangor University

Proudlove, N, Bisogno, S, Onggo, BSS, Calabrese, A & Levialdi-Ghiron, N 2017, 'Towards fully-facilitated discrete event simulation modelling: Addressing the model coding stage' European Journal of Operational Research, vol 263, no. 2, pp. 583-595.

Puntambekar N. (2016) Exploring the efficiency of a UK ambulance service through Discrete-Event Simulation, Master Dissertation, Loughborough University.

Robinson, S., Worthington, C., Burgess, N. and Radnor, Z.J. (2014). Facilitated modelling with discrete-event simulation: Reality or myth? *European Journal of Operational Research*, 234 (1): 231–240.

Tako AA, Kotiadis K, (2015) 'PartiSim: A Framework for participative simulation modelling'. European Journal of Operational Research, 244(2), p555-564.

Tako AA, Kotiadis K, Vasilakis C, Miras A, le Roux C W, (2014) 'Modeling patient waiting times for an sobesity service: a computer simulation study'. BMJ Quality and Safety, 23 (5) p373-381.

Taket, A. (2002). "Facilitation: some contributions to theorising the practice of operational research." *Journal of the Operational Research Society* 53(2): 126-136.

AUTHOR BIOGRAPHIES

ANTUELA A. TAKO is a Reader in Operational Research at the School of Business and Economics, Loughborough University. She holds a PhD in Simulation and an MSc in Management Science and Operational Research from the University of Warwick. Her research interests include the comparison of simulation approaches (discrete-event simulation and system dynamics), facilitated and participative simulation modelling, conceptual modelling and health care modelling. She is Associate Editor of the Journal of the Operational Research Society. Her email address is a.a.takou@lboro.ac.uk and her webpage is: http://www.lboro.ac.uk/departments/sbe/staff/anthi-antuela-tako/.

KATHY KOTIADIS is Kathy Kotiadis is a Reader (Associate Professor) in Management Science/Operational Research at the Kent Business School, University of Kent. She graduated with a BSc (Hons) in Management Science and went on to do a PhD in Operational Research at the University of Kent. Her main research interests include discrete event simulation modelling applied to health care and the development of the simulation methodology through Problem Structuring Methods. In 2009 she was awarded the KD Tocher Medal by the OR society for the best simulation paper published in the Journal of Simulation over the period 2007-8. Her email address is K.Kotiadis@kent.ac.uk.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.006

TUTORIAL ON SIMULATION MODEL VERIFICATION AND VALIDATION

Prof. Stewart Robinson

School of Business and Economics
Loughborough University
Loughborough, LE11 3TU
s.l.robinson@lboro.ac.uk

ABSTRACT

The activities of verification and validation aim to give confidence that a simulation model can be used to aid decision-making. This tutorial explores the definitions of verification and validation, and discusses how they fit into the process of performing a simulation study. Difficulties in validating a model are then discussed. This leads to the conclusion that it is impossible to completely assure the validity of a model, but it is possible to build confidence in a model through verification and validation. Some practical methods for performing verification and validation are described.

Keywords: Simulation Model, Verification, Validation, Confidence

1 INTRODUCTION

Verification and validation aim to determine the accuracy with which a simulation model predicts the performance of the real system it is representing. In this tutorial the concepts of verification and validation are explored as well as some methods for model testing. The paper is split into three parts. First, the terms verification and validation are defined, and various forms of verification and validation are described and set in the context of the process of performing a simulation study. There is then a discussion on the difficulties that are encountered when trying to perform verification and validation. Third, some useful verification and validation methods are described.

2 WHAT IS VERIFICATION AND VALIDATION?

Verification is the process of ensuring that the model design (conceptual model) has been transformed into a computer model with sufficient accuracy (Davis, 1992). Validation, on the other hand, is the process of ensuring that the model is sufficiently accurate for the purpose at hand (Carson, 1986). Verification has quite a narrow definition and in many respects it can be seen as a subset of the wider issue of validation.

There are two key concepts in validation: the ideas of *sufficient accuracy* and models that are built for a specific *purpose*. No model is ever 100 percent accurate (for all purposes); although it might be possible to generate a perfect model if the system is very simple and the purpose is very narrow. Anyhow, in general, a model is not meant to be completely accurate, but a simplified means for understanding and exploring reality (Pidd, 2009). In verification and validation the aim is to ensure that the model is sufficiently accurate. Further, this accuracy is with reference to the purpose for which the model is to be used. As a consequence, the purpose, or objectives, of a model must be known before it can be validated. This purpose may have been determined at the start of the simulation study, being expressed through the objectives, or it may be an alternative use for an existing model. Under this definition for validation it is possible to think in terms of absolute validity; a model is either sufficiently accurate for its purpose or it is not. In other words, validity is a binary decision with a conclusion of 'yes' or 'no'. Proving this is a different matter, as is discussed in section 3.

Verification and validation can be further understood by mapping the verification and validation requirements onto the process of performing a simulation study (Robinson, 2014). Figure 1 shows that for each activity in a simulation study, at least one verification or validation activity is performed in parallel.

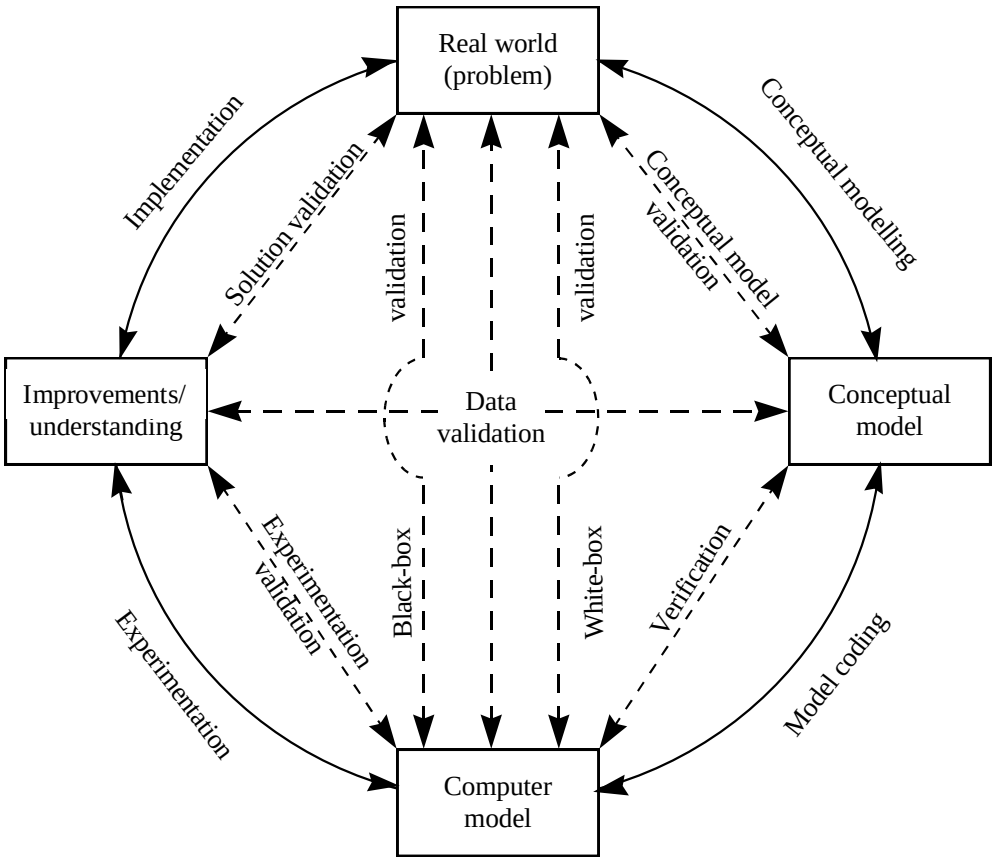


Figure 1 *Simulation Model Verification and Validation in a Simulation Study (adapted from Landry et al (1983))*

Various forms of validation are identified, which can be defined as follows:

- *Conceptual Model Validation*: determining that the content, assumptions and simplifications of the proposed model are sufficiently accurate for the purpose at hand. The question being asked is: does the conceptual model contain all the necessary details to meet the objectives of the simulation study?
- *Data Validation*: determining that the contextual data and the data required for model realisation and validation are sufficiently accurate for the purpose at hand. As shown in figure 1, this applies to all stages in a simulation study, since data are required at every point.
- *White-Box Validation*: determining that the constituent parts of the computer model represent the corresponding real world elements with sufficient accuracy for the purpose at hand. This is a detailed, or micro, check of the model, in which the question is asked: does each part of the model represent the real world with sufficient accuracy to meet the objectives of the simulation study?
- *Black-Box Validation*: determining that the overall model represents the real world with sufficient accuracy for the purpose at hand. This is an overall, or macro, check of the model's operation, in which the question is asked: does the overall model provide a sufficiently accurate representation of the real world system to meet the objectives of the simulation study?
- *Experimentation Validation*: determining that the experimental procedures adopted are providing results that are sufficiently accurate for the purpose at hand. Key issues are the requirements for removing initialisation bias, run-length, replications and sensitivity analysis to assure the accuracy

of the results. Further to this, suitable methods should be adopted for searching the solution space to ensure that learning is maximised and appropriate improvements identified.

- *Solution Validation*: determining that the results obtained from the model of the proposed solution are sufficiently accurate for the purpose at hand. This is similar to black-box validation in that it entails a comparison with the real world. It is different in that it only compares the final model of the proposed solution to the implemented solution. Consequently, solution validation can only take place post implementation and so, unlike the other forms of validation, it is not intrinsic to the simulation study itself. In this sense, it has no value in giving assurance to the client, but it does provide some feedback to the modeller.

Verification is also identified on figure 1 as a test of the fidelity with which the conceptual model is converted into the computer model (as per its definition). On the surface verification and white-box validation may look very similar in that they both involve checks of the detail in the model. The difference lies in what the model is checked against. In verification the reference point is the conceptual model and as such the modeller can carry out verification on his/her own without reference to the real world. Meanwhile, in white-box validation the reference point is the real world and so the modeller must work with domain experts.

What should be apparent is that verification and validation is not just performed once a complete model has been developed, but that *verification and validation is a continuous process that is performed throughout the life-cycle of a simulation study*. In the same way that modelling is an iterative process, so too is verification and validation. From the early stages of developing a conceptual model this model should be validated. As the project progresses the conceptual model is likely to be revised as the understanding of the problem and the modelling requirements change. As a consequence, the conceptual model needs to be revalidated. While the conceptual model is being transformed into a computer model, the constituent parts of the model (particularly those recently coded) should be continuously verified. Similarly, the details of the model should be checked against the real world throughout model coding (white-box validation). Black-box validation requires a completed model, since it makes little sense to compare the overall model against the real world until it is complete. This does not imply, however, that black-box validation is only performed once. The identification of model errors and continued changes to the conceptual model necessitates model revisions and therefore further black-box validation. In a similar way, the experimental procedures need to be validated for every revision of the model, including the experimental scenarios. It cannot be assumed that the requirements for experimentation are the same for every model version.

Although white-box validation and black-box validation are often lumped together under one heading, operational validity (Sargent, 2013), it is because they are performed as separate activities during a simulation study that a distinction is drawn between them here. White-box validation is intrinsic to model coding, while black-box validation can only be performed once the model code is complete.

3 THE DIFFICULTIES OF VERIFICATION AND VALIDATION

Before discussing specific methods of verification and validation it is important to recognise that there are a number of problems that arise in trying to validate a model.

3.1 There is No Such Thing as General Validity

A model is only validated with respect to its purpose. It cannot be assumed that a model that is valid for one purpose is also valid for another. For instance, a model of a production facility may have been validated for use in testing alternative production schedules, however, this does not mean that it is necessarily valid for determining that facility's throughput. A model could only be described as generally valid if it could be demonstrated that it was suitably accurate for every purpose to which it might ever be put. Not only is it unlikely that every potential purpose for a model could be determined, but also such a model would probably be very extensive, requiring vast amounts of code, data and run-time. This goes against the principle of keeping models as simple as possible for the task at hand. Indeed, reality is the only 'model' which is generally valid.

3.2 There may be No Real World to Compare Against

Much validation requires a comparison of the model to the real system (conceptual model validation, white-box validation and black-box validation). However, many models are developed of proposed systems, for instance, new production or service facilities. As a consequence, there is no real world to use for comparison. Even if the model is of an existing system, its purpose is to investigate alternative operating practices, for which again no real world exists. The model may be shown to be valid when it is representing the existing operation, but this does not guarantee that it is valid once it represents some change to the system.

3.3 Often the Real World Data are Inaccurate

Validation often involves a comparison of some facet of the model, for instance throughput, against real world data. The model is run under the same conditions as the real world to see if it performs in a similar manner. There are two difficulties that arise with this procedure. First, the real world data may not be accurate. Indeed, the purpose of data validation is to determine the accuracy of the data that are being used. If the data are not accurate, however, this creates problems in determining whether a model's results are correct.

Second, even if 'accurate' real world data do exist, it must be remembered that these are only a sample, which in itself creates inaccuracy. For instance, data may have been collected on the throughput of a production facility over a ten-week period. If, however, data had been collected for a further ten weeks this would no doubt have changed the sample distribution and summary statistics (e.g. the mean) of the data. To exacerbate the problem, the simulation itself is providing only a sample; results of say ten weeks of operation. This means that the real world-to-model comparison is a comparison of two samples. Although statistical procedures can be used to determine whether these two samples are similar, these only provide a probabilistic and not a definitive answer.

3.4 Which Real World?

Different people have different interpretations of the real world, described as Weltanschauung or world views by Checkland (1981). An employee in a bank may see the bank as a means for earning money, while a customer may see it as a means for safely storing money, or as a means for borrowing money. Depending on who we speak to, we obtain different interpretations of the purpose and operation of the bank. Every day we can read multiple accounts of the same event in our newspapers, each with subtle (or not so subtle!) differences. The event was the same, but the reporters' interpretations vary. This problem becomes more extreme as we move from modelling what are primarily physical systems to modelling human centric and social systems.

This presents a problem when validating models. If people have different world views, which interpretation(s) should be used for developing and validating a model? A model that is valid to one person may not be valid to another.

3.5 There is Not Enough Time to Verify and Validate Everything

There is simply not enough time to verify and validate every aspect of a model (Balci, 1997). Those that develop software have experienced users breaking what was thought to be perfectly sound code. This is a problem that affects both verification and validation. The modeller's job is to ensure that as much of the model is verified and validated as possible, both in terms of the model details (conceptual model validity, verification, white-box validation and data validation), the overall validity (black-box validation) and the experimental procedures (experimentation validation).

3.6 It is Impossible to Prove that a Model is Valid: Confidence not Validity

The conclusion of this is that although, in theory, a model is either valid or it is not, proving this in practice is a very different matter. Indeed, it is not possible to prove that a model is valid. Instead, it is only possible to think in terms of the confidence that can be placed in a model. The process of verification and validation is not one of trying to demonstrate that the model is correct, but is in fact a process of trying to prove that the model is incorrect. The more tests that are performed in which it

cannot be proved that the model is incorrect, the more the clients' (and the modeller's) confidence in the model grows. The purpose of verification and validation is to increase the confidence in the model and its results to the point where the clients are willing to use it as an aid to decision-making. It is also important for the modeller to have confidence that the simulation should be used for decision-making. However many tests are performed, it is always possible that the next test might show the model to be invalid, and so validity can never be completely assured.

4 METHODS OF VERIFICATION AND VALIDATION

There are many methods of verification and validation available to simulation modellers. Here a summary of some useful approaches is provided. For detailed reviews of verification and validation techniques see Balci (1994), Yilmaz and Balci (1997) and Sargent (2013). It should be noted that good quality documentation provides significant help to any verification and validation effort.

4.1 Conceptual Model Validation

There are no formal methods for validating a conceptual model. A project specification document is a means for helping to determine what confidence should be placed in the model. The specification should be circulated among those who have a detailed knowledge of the system and feedback sought on whether the model is appropriate. Where issues occur, these should be dealt with either by adjusting the conceptual model, or by clarifying any misunderstandings. By gaining wide acceptance for the conceptual model the confidence of the modeller and the clients is increased.

It is useful for the modeller and the clients to jointly assess the assumptions for the level of confidence that can be placed in them and their likely impact on the results of the model:

- *Confidence* is the level of certainty that an assumption about the real system is correct
- *Impact* is the estimated effect on the model results if the assumption is incorrect

For every assumption both the level of confidence and the impact could be assessed as being high, medium or low. The impact of any simplifications should also be assessed, although we would generally not expect this to be high, otherwise the simplification should probably not have been selected in the first place. Albeit purely based on judgement, such an assessment both ensures that the potential effect of all the assumptions and simplifications is considered and helps identify any areas of particular concern. For instance, those assumptions about which there is low confidence, and that it is believed have a high impact, need to be addressed. One approach is to try and remove them by altering the model or investigating the real system further. Alternatively, and when it is not possible to remove them, sensitivity analysis can be performed later in the project to quantify their impact.

4.2 Data Validation

Data are obviously a potential source of inaccuracy in a simulation model and can in their own right move a model from being sufficiently accurate to being invalid. Every effort should be made to ensure that the data are as accurate as possible. The modeller should investigate the sources of data to determine their reliability. The data should be analysed for inconsistencies and any cause for concern investigated. Beyond this, much has to be put down to trust especially when the modeller is simply presented with data.

4.3 Verification and White-Box Validation

Although verification and white-box validation are conceptually different, they are treated together here because they are both performed continuously throughout model coding. Also, they are both micro checks of the model's content. Verification ensures that the model is true to the conceptual model, while white-box validation ensures that the content of the model is true to the real world (in this way it is an indirect form of conceptual model validation). Verification can be performed by the modeller alone, comparing the computer model to the conceptual model description. Meanwhile, white-box validation requires the involvement of those knowledgeable about the real world system. Whereas verification can be performed almost continuously during model coding, white-box

validation is performed less frequently since it requires the involvement of more than just the modeller.

Various aspects of the model should be checked during model coding:

- Timings e.g. cycle times, repair times and travel times
- Control of elements e.g. breakdown frequency and shift patterns
- Control of flows e.g. routing
- Control logic e.g. scheduling and stock replenishment
- Distribution sampling e.g. the samples obtained from an empirical distribution

Three methods of verification and white-box validation are now discussed.

Checking the Code

The modeller needs to read through the code to ensure that the right data and logic have been entered. This is especially true for areas of complex logic. A useful idea is to get someone else to read the code, or to explain the code to someone else as a second check. If no modelling experts are available, then most simulation software vendors offer a help-desk service with which specific areas of code could be discussed. Alternatively, by expressing the code in a non-technical format a non-expert could check the data and the logic. This is especially useful for obtaining the opinion of those who have a detailed knowledge of the system being modelled.

Visual Checks

The visual display of the model proves to be a powerful aid for verification and validation. By running the model and watching how each element behaves both the logic of the model and the behaviour against the real world can be checked. Various ideas aid this approach:

- Stepping through the model event by event
- Stopping the model, predicting what will happen next, running the model on and checking what happens
- Interactively setting up conditions to force certain events to take place
- Creating extreme conditions, such as a very high arrival rate, to determine whether the model behaves as expected
- Isolating areas of the model so it runs faster, reducing the time to perform thorough verification and validation of that part of the model
- Explaining the model as it runs to those knowledgeable about the real system in order to gain their opinion
- Tracing the progress of an item through the model

It is useful simply to watch a model running for a period of time. In so doing a lot can be learnt about the behaviour of the simulation. It is also useful to demonstrate the model, formally and informally, to those who have a detailed knowledge of the system. Not only does this enable them to identify any shortcomings in the model, but by involving them this should increase the credibility of the work (assuming that not too many errors are found!).

Inspecting Output Reports

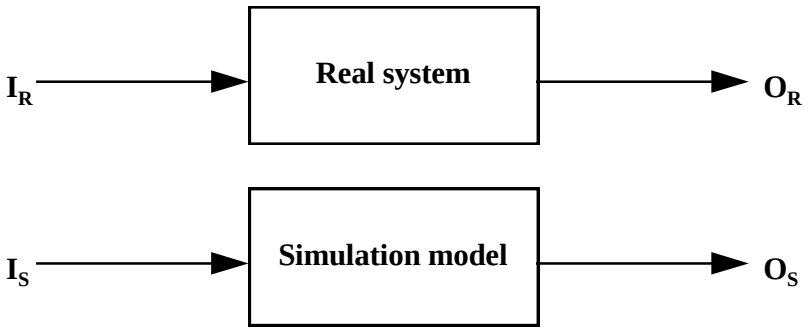
By inspecting the reports from a simulation run, the actual and expected results can be compared. Of interest in verification and white-box validation is the performance of the individual elements, for example, service point utilisations. Graphical reports of samples from input distributions, for instance, machine repair times, are an aid in checking that they are being modelled correctly. More formal (statistical) methods for comparing distributions can be employed to provide a more rigorous check.

A report which may be of some use is a 'trace' of a simulation run. This is a blow-by-blow history, normally written to a file, of every event that takes place during a simulation run. Inspecting this report can help to diagnose and rectify any problems.

4.4 Black-Box Validation

In black-box validation the overall behaviour of the model is considered. There are two broad approaches to performing this form of validation. The first is to compare the simulation model to the real world. The other is to make a comparison with another model. The second approach is particularly useful when there are no real world data to compare against.

Comparison with the Real System
If confidence is to be placed in a model then, when it is run under the same conditions (inputs) as the real world system, the outputs should be sufficiently similar. As shown in figure 2, when I_S is set to be the same as I_R then O_S should be similar (with sufficient accuracy) to O_R. This concept is expressed as the null hypothesis (H_0) in figure 2. If the null hypothesis is rejected then, at the chosen level of significance, the model is believed to be invalid. The approximation sign shows that the model need only be sufficiently accurate.



I_R - inputs to real system
O_R - outputs from real system
I_S - inputs to simulation model
O_S - outputs from simulation model

H_0: If I_S=I_R then $O_S \approx O_R$

Figure 2 *Black-Box Validation: Comparison with the Real System*

As already stated, the significant difficulty with this form of validation is that there may not be any accurate real world data with which to perform such a comparison. If this is the case then the comparison can be made against the expectations and intuition of those who have a detailed knowledge of the real system. Comparison against approximate real world data such as these may not give absolute confidence in the model, but it should help to increase confidence.

Historic (or expected) data collected from the real system, such as throughput and customer service levels, can be compared to the results of the simulation when it is run under the same conditions. It is important to check not only the average levels of these data, but also to compare their spread. This can be performed by judging how closely the averages from the model and the real world match, and by visually comparing the distributions of the data. Various statistical tests also lend themselves to such comparisons (Kleijnen, 1995). Assuming that the same quantity of output data is generated from the simulation model as is available from the real system, then a confidence interval for the difference in the means can be calculated as follows:

$$\overline{X}_S - \overline{X}_R \pm t_{2n-2,\alpha/2}\sqrt{\frac{S_S^2 + S_R^2}{n}}$$

where:

$\overline{X}_S$ = mean of simulated output data

$\overline{X}_R$ = mean of real system output data

S_S = standard deviation of simulated output data
S_R = standard deviation of real system output data
n = number of observations (this must be the same for the simulated and real system data)
$t_{2n-2,\alpha/2}$ = value from Students t-distribution with $2n$-2 degree of freedom and a significance level of α /2

If the sample size (n) is different then a more complex calculation is required (Montgomery and Runger, 2002). Of course, it is probably simpler to delete observations from the larger sample in order to make the sample sizes equal, although this loses some valuable information.

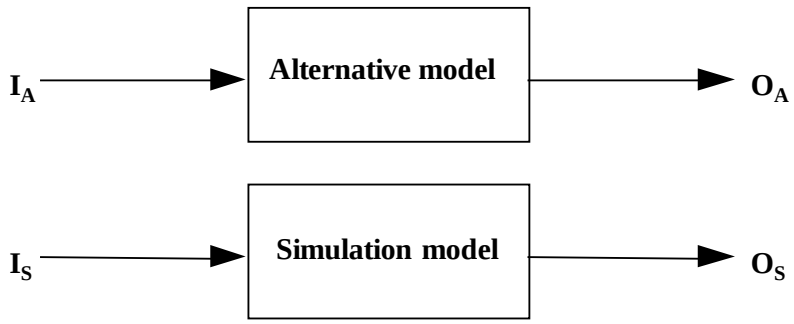
Apart from using a confidence interval to compare the output from the model with the real world, P-P plots and chi-square tests could be used to compare the distributions of the output data. Another powerful approach is to run a simulation from a trace of historic data, enabling a more direct comparison of the model with the real world (Kleijnen, 1995; Kleijnen et al, 1998; 2001).

An alternative approach is to compare the relationships between the inputs and outputs in the model and the real world. For instance, if it is known that when an input (e.g. a storage area) is increased by 20 percent in the real world there is a corresponding 10 percent increase in one of the outputs (e.g. throughput), a similar relationship should be obtained from the model.

In a Schruben-Turing Test (Schruben, 1980) the model reports are made to look exactly the same as the reports provided by the real system. One or more reports from the model and from the real world are given to someone who is knowledgeable about the system. He/she is then asked to try and distinguish between the two. If he/she is unable to detect any difference, this increases the confidence in the model. If differences are detected, then the reason for these should be investigated and corrected in the model if they are deemed significant. Even if real world reports are not available, it is still worth asking a domain expert to review the model reports.

Comparison with Other Models

As an alternative to comparison against the real world, the simulation can be compared to other, normally simpler models (figure 3). This group of methods is particularly useful when no real system data are available. However, this does not preclude their use when these data are available. Indeed, comparisons with other models in addition to comparisons with the real world can only serve to generate more evidence for helping to assess the confidence that can be placed in a model.



I_A - inputs to alternative model
O_A - outputs from alternative model
I_S - inputs to simulation model
O_S - outputs from simulation model

H_0: If I_S=I_A then $O_S \approx O_A$

Figure 3 *Black-Box Validation: Comparison with an Alternative Model*

One approach is to compare the simulation model against a mathematical model. It is unlikely that a mathematical model is able to predict the outcome of the simulation exactly, otherwise the simulation would probably not have been built. However, for the purposes of comparison a

mathematical model may be able to give a crude approximation to the outputs of the real system. Examples of mathematical models that might be used are paper calculations, spreadsheet analysis and queuing theory (Winston, 2003). This approach is sometimes referred to as static analysis because it does not (cannot) take into account the full dynamics of the simulation model.

In order to aid comparison it is sometimes useful to simplify the simulation model to the extent that a mathematical model can predict exactly, or at least more exactly, the outcome of the model. One specific, and extreme, case of this is the use of deterministic models. This is a simulation model from which all the random events are removed. In many cases it is possible to determine mathematically the exact outcome of such a model. (This approach is also beneficial for verification and white-box validation as logic errors can sometimes be more easily spotted when stochastic behaviour has been removed from the model.)

Comparisons can also be made against other simulation models of the same or similar systems. For instance, a more detailed model of the system may have been developed for some other purpose. This presupposes, of course, that the other model is itself valid.

Onggo and Karatas (2016) provide an example of validation through comparison with a simpler mathematical model. In this case the simulation is an agent-based model of maritime search operations.

4.5 Experimentation Validation

Assuring the accuracy of simulation experiments requires attention to the issues of initial transient effects, run-length, the number of replications and sensitivity analysis. Also, the search of the solution space should be sufficient to obtain an adequate understanding and identify appropriate improvements. Methods for dealing with these issues are described in some detail in Robinson (2014).

4.6 Solution Validation

The aim of all modelling and verification and validation efforts is to try and assure the validity of the final solution (or improvement). Once implemented, it should be possible to validate the implemented solution against the model's results. This is similar in concept to the comparisons with the real world performed in black-box validation, except that the comparison is between the final model of the proposed solution and the implemented solution. Therefore, the techniques of black-box validation discussed above can be applied.

Solution validity should also involve checking whether the implemented solution is indeed the most suitable. In practice, however, it is unlikely that this is possible, since it is not usually practical to implement alternative solutions to determine their effect; this, no doubt, is the reason for using a simulation in the first place. Neither is solution validation possible if the simulation is only used to develop a better understanding of the real world and not directly to identify improvements. A form of reverse validation, however, may be possible. An improved understanding may lead to the implementation of new ideas. These ideas could then be included in the simulation model and a comparison made to the real world, thereby checking the accuracy of the model.

From discussions with simulation practitioners it is apparent that solution validation is rarely carried out even though it is the only true test of the outcome of a simulation study. A key problem is that the implementation may take many years to complete, by which time the momentum for the simulation work, and possibly the simulation modeller, have long disappeared. Another issue is whether the solution is properly implemented and so whether a meaningful comparison can be made.

5 CONCLUSION

It is not possible to prove that a model is valid. Therefore, model verification and validation is concerned with creating enough confidence in a model for it to be used in decision-making. This is done by trying to prove that the model is incorrect. The more tests that are performed in which it cannot be proved that the model is incorrect, the more confidence in the model is increased. For verification and validation the general rule is: the more testing the better.

Of course, the modeller and the clients may have different thresholds for confidence. Some clients may derive their confidence simply from the model's display, others may require more in-depth verification and validation before they are willing to believe the results. The modeller is responsible for guiding the clients and ensuring that sufficient verification and validation is performed.

Finally, the modeller should remember that the acceptance of a simulation study and its results does not rest solely on the validity of the model. Verification and validation assures (content) quality in the sense that the model conforms to the clients' technical requirements for a model and a set of results that are sufficiently accurate. What it does not determine is the extent to which the simulation study meets the clients' expectations concerning the process of project delivery (process quality).

ACKNOWLEDGMENTS

This paper is based on Robinson, S. (1999). Simulation Verification, Validation and Confidence: A Tutorial. *Transactions of the Society for Computer Simulation International*, 16 (2), pp. 63-69. Copyright © by Simulation Councils, Inc. Reproduced by permission. It is also adapted from Robinson, S. (2014). *Simulation: The Practice of Model Development and Use, 2nd edition*. Palgrave, London, chapter 12.

REFERENCES

Balci O (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study, *Annals of Operations Research* **53**: 121-173.

Balci O (1997). Principles of simulation model validation, verification, and testing, *Transactions of the Society for Computer Simulation* **14(1)**: 3-12.

Carson J S (1986). Convincing users of model's validity is challenging aspect of modeler's job, *Industrial Engineering* **18(6)**: 74-85.

Checkland P (1981). *Systems Thinking, Systems Practice*. Wiley: Chichester, UK.

Davis P K (1992). Generalizing concepts of verification, validation and accreditation (VV&A) for military simulation. R-4249-ACQ, October 1992, RAND: Santa Monica, CA.

Kleijnen J P C (1995). Verification and validation of simulation models, *European Journal of Operational Research* **82(1)**: 145-162.

Kleijnen J P C, Bettonvil B and Groenendaal W V (1998). Validation of trace-driven simulation models: a novel regression test, *Management Science* **44(6):** 812-819.

Kleijnen J P C, Cheng R C H and Bettonvil B (2001). Validation of trace-driven simulation models: bootstrap methods, *Management Science* **47(11)**: 1533-1538.

Landry M, Malouin J L and Oral M (1983). Model validation in operations research. *European Journal of Operational Research* **14(3)**: 207-220.

Montgomery D C and Runger G C (2002). *Applied Statistics and Probability for Engineers, 3rd ed.* Wiley: New York.

Onggo B S and Karatas M (2016). Test-driven simulation modelling: A case study using agent-based maritime search-operation simulation. *European Journal of Operational Research* **254(2)**: 517-531.

Pidd M (2009). *Tools for Thinking: Modelling in Management Science, 3rd ed*. Wiley, Chichester, UK.

Robinson S (2014). *Simulation: The Practice of Model Development and Use, 2nd edition*. Palgrave: London.

Sargent R G (2013). Verification and validation of simulation models, *Journal of Simulation* **7(1)**: 12–24.

Schruben L W (1980). Establishing the credibility of simulations, *Simulation* **34(3)**: 101-105.

Winston W L (2003). *Operations Research: Applications and Algorithms, 4th ed.* Duxbury Press: Belmont, CA.

Yilmaz L and Balci O (1997). Object-oriented simulation model verification and validation. *Proceedings of the 1997 Summer Computer Simulation Conference* (Obaidat, M.S. and Illgen, J., eds.). Society for Computer Simulation International: San Diego, CA, pp 835-840.

AUTHOR BIOGRAPHY

STEWART ROBINSON is Professor of Management Science and Dean at the School of Business and Economics, Loughborough University, UK. His research interests are in the practice and use of simulation models where he focuses on conceptual modelling, model validation, output analysis and alternative simulation methods (discrete-event, system dynamics and agent based). His home page is www.stewartrobinson.co.uk.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.007

TUTORIAL: TEXT ANALYTICS FOR SIMULATION WITH PYTHON

Dr. Roger McHaney

Kansas State University
Management Information Systems
1301 Lover's Lane
Manhattan, KS USA 66506
mchaney@ksu.edu

ABSTRACT

Text-based data analytics offers many opportunities related to simulation practice. This tutorial describes these with three primary examples: (1) using sentiment analysis to drive a simulation; (2) using social network analytics to structure a simulation; and, (3) developing model timelines using Twitter analysis. The tutorial includes all aspects of data analytics including data collection, cleaning, analysis and use---all specifically adapted to simulation practitioners. The tutorial uses a hands-on approach with Python in Jupyter notebook and Spyder environments and data sources as JSON, CSV, Twitter/social media, and web scraping/crawling. Other techniques such as setting up a cloud-based data collection platform are also described. The three examples use VADERsentiment, Pandas, word clouds, text analytics, topic modeling, social network analytics and other tools. Example code is provided during the tutorial. Those attending the tutorial will learn ways to enhance the human element in their models based on real-world data sources.

Keywords: Data Analytics, Text Analytics, Simulation Applications, Python, Gephi

1 INTRODUCTION

Text-based data analytics offer a wide range of opportunities for understanding system operation and enhancing analysis techniques. This is true in many fields, including simulation. For instance, simulation practitioners may need to develop mechanisms that represent the frequency, categories, types, and nuances of data that enter a system. This is particularly applicable to situations where human factors come into play. Text analytics offer insights into data regarding its polarity, emotional content, topical aspect and overall intent. Data such as these may be used to drive human factor modeling and allow better representation of the unpredictable nature people bring into systems.

Many traditional data collection and analysis techniques can be supplemented or replaced with automated processes influenced by the widespread proliferation of text-based data sources such as web sites, social media, Twitter, emails, closed captioning, business documents, contracts, resumes, proposals, PDF files and many others. The current tutorial leverages this phenomenon and focuses on data collection, cleaning, analysis, and use adapted to simulation practitioners and researchers. Many of the techniques and practices described in this tutorial have broader application but all three primary examples are tailored to model development and use.

1.1 Simulating Human Behavior

Simulating human behavior is a challenge and perhaps even limitation in many models. In fact, Baines et al. (2004) suggested that human aspects were rarely included in discrete event simulations and this represented a major shortcoming in many studies. More than a decade later, Greasley and Owen (2018) found this challenge remains and requires more work.

Several approaches related to big data can help mitigate challenges related to the human element in models. For instance when building a model of cybersecurity-related policy, Twitter can be

analyzed for sentiment to reveal polarity (e.g. negative or positive feelings), future intent, or other important topics related to the human choices and potential behaviors (Gupta, Sharma, & Chennamaneni, 2016). Findings such as these can be incorporated into a model to realistically represent the ways people may react to policy changes. Likewise, in models where feelings, motivations, and actions are based on human factors, text analytics has the potential to provide realistic data for driving the model. For example, a model of consumer behavior related to purchasing airline tickets could be improved by incorporating consumer sentiments derived from real world datasets (Khan & Urolagin, 2018).

1.2 Text-based Analytics in the Modeling Lifecycle

Text-based analytics techniques fit well into several phases of the simulation modeling lifecycle. Prior research into DES modeling has used a 4-stage model to facilitate understanding in other areas (Tako & Kotiadis, 2015). We will use the same framework in this tutorial to communicate where text analytics might be applied in the modeling process. Table 1 provides examples.

Table 1 *Text Analytics Applications Mapped to DES Stage (Tako & Kotiadis, 2015, p. 557)*

DES Stage	Text Analytics Applications
Problem structuring/ Conceptualisation	Problem formulation/discovery, data acquisition, cleaning, and processing; Visualizations of input data for better understanding of system and conceptual model
Simulation model coding	Transforming data into model constructs
Obtaining solutions	Visualizations based on output data from model
Implementation	Transference of knowledge gained; Storytelling

In general, text analytics techniques can be used to discover problems, collect data for model constructs, and provide understanding in early model stages. Later, input data and acquired knowledge can be used in model construction. Model outputs can be analyzed using text analytics tools, and finally the outcomes of the model can be communicated using storytelling techniques common to data analytics.

1.3 Tutorial Approach

The current tutorial starts with building a data analytics environment for use by simulation analysts. Real world data can be obtained for use in modelling from a variety of sources. In some instances, it may have to be derived from unstructured data sources that require careful collection, cleaning, and interpretation. This environment used in the tutorial relies on Anaconda, which is an open source implementation of Python and R programming languages for scientific computing. We use Jupyter Notebooks and/or Spyder within Anaconda's distribution to run Python scripts that assist with text analytics in this tutorial.

Within this environment, we will cover several powerful tools for text analytics. We start with data acquisition using CSV and JSON data files, Tweets from Twitter, web scraping and crawling, and other sources. Collected data will be cleaned using a variety of tools that remove stop words, short words, irrelevant tokens, and other non-meaningful symbols. Further data analysis will put the collected and cleaned data into a form ready for further analysis. This includes lists and Pandas data frames. Specific analysis takes place next. Techniques described in the tutorial will include descriptive measures, readability, sentiment analysis, text analytics (e.g. supervised and unsupervised machine learning), topic modeling, and others. Environments used in visualization and storytelling (such as

Microsoft Power BI and Tableau) will be described at this stage. Finally, the data will transformed for use in simulation models to illustrate usefulness to modelers.

1.4 Starting Correctly

One of the most important steps in determining what tools to use and how to approach data collection, involves understanding an overall project goal. Robinson points out two important aspects of accomplishing this: first, developing a comprehensive problem description and second, formulating project goals (Robinson, 2008; Robinson, Arbez, Birta, Tolk, & Wagner, 2015). This will enable a simulation team to better understand what data to collect supported by project goals and purpose and will drive the storytelling function.

2 TOOLS FOR TEXT ANALYTICS

Text analytics requires collecting, cleaning, and working with unstructured data that comes from many sources in a variety of formats. Frequently, this data is intended for uses that do not consider collection and analysis. Therefore, a flexible tool set is required to work effectively in this area.
The current tutorial focuses on the Python programming language which has been specifically adapted to data analytics tasks with libraries, objects and scripts. We use the Anaconda distribution of Python and focus on the Jupyter Notebooks IDE (integrated development environment). An alternate approach would be to use 'R' and its scripts and toolsets. Another potential tool is LIWC (McHaney, Tako, & Robinson, 2018).

Python is an object-oriented, programming language. It originally was considered a scripting language associated with web and app development but has evolved into a high-level, widely used general purpose programming language. It has an easily understood syntax that lends it to being highly readable and, in some ways, self-documenting. Python supports user developed packages that can be installed and implemented in specialized ways. Most features used for data analytics fall into this category. For instance, Numpy, Pandas, Matplotlib, and Scikit-Learn are popular packages used in data science applications. Adding to the appeal, many Python packages, its standard library, and interpreter all are free.

Anaconda is a popular deployment of Python for data science applications. This is an open source distribution that includes both Python and R programming languages. Its primary purpose is to simplify package installation, management and deployment. Anaconda makes it possible to easily use more than 1500 data science libraries and packages. It runs on computers with Windows, Linux, or Mac operating systems. Anaconda's Python distributions include an interpreter and several Python editors (e.g. Jupyter Notebook and Spyder) as well as numerous Python tools and packages. The Anaconda distribution for data science can be downloaded from this website (Figure 1): https://www.anaconda.com/distribution/.

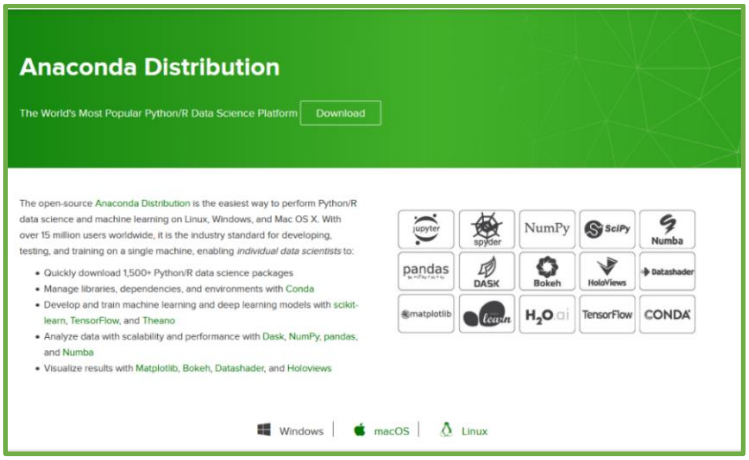


Figure 1 *Anaconda Distribution for Data science Download Website*

At the time this tutorial was developed, the current version of Python was 3.7. The first step in installation is to press on the appropriate installer for your computer. Figure 2 indicates saving the download and then doubleclicking on the file to initiate the local installation process. Press *run* as shown in Figure 3 then the Anconda installer begins.

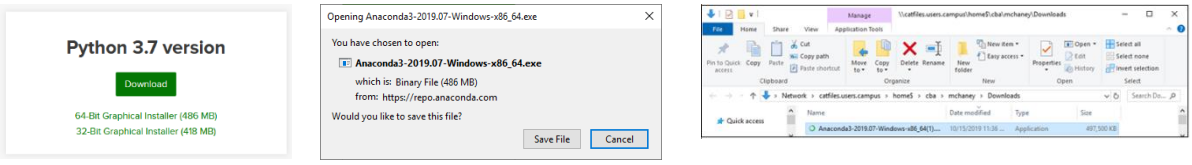


Figure 2 *Download File and Double Click to Install*

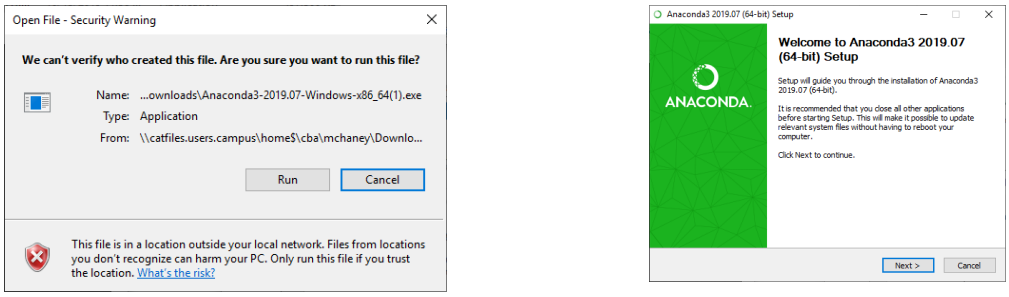


Figure 3 *Verify Installation before Anaconda Installer Begins*

As shown in Figure 4, agree to the open source license and install as *Just Me* to avoid problems later when running packages on your computer for this tutorial. Finally, choose an installation location that provides read/write permission like C:\Users\yourname\anaconda3 in Windows environments.

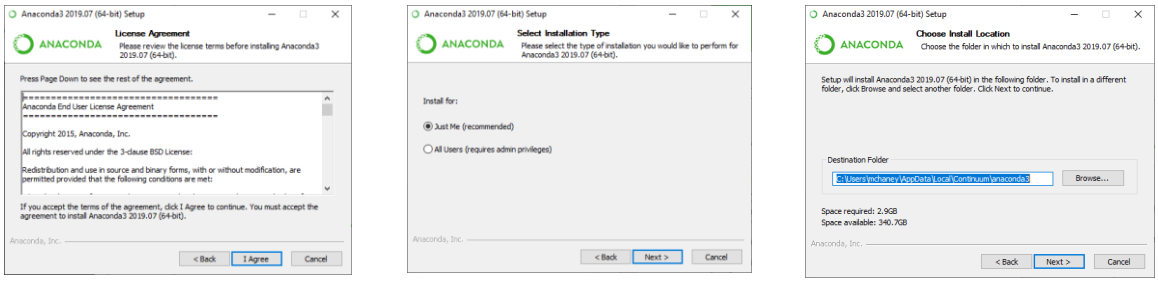


Figure 4 *Select Options to Ensure Usable Installation*

Finish the installation by adding Anaconda to the PATH environment variable by checking the boxes under advanced options (Figure 5). After that, proceed with the defaults on the following step.

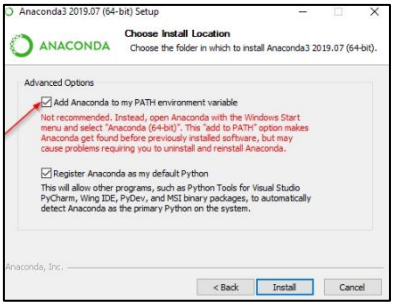


Figure 5 *Add Anaconda to PATH environment variable*

Once the environment is installed, Python can be accessed and used through Jupyter Notebook. It runs in several ways. It can be started from the command prompt by typing*: jupyter notebook*, or from the Windows Start Icon as shown in Figure 6, or from within Anaconda as shown in Figure 7. Once Jupyter Notebook starts, it will appear in a browser window as shown in Figure 8.

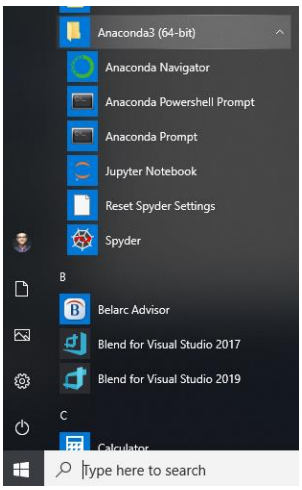


Figure 6 *Starting Jupyter Notebook from Windows start icon*

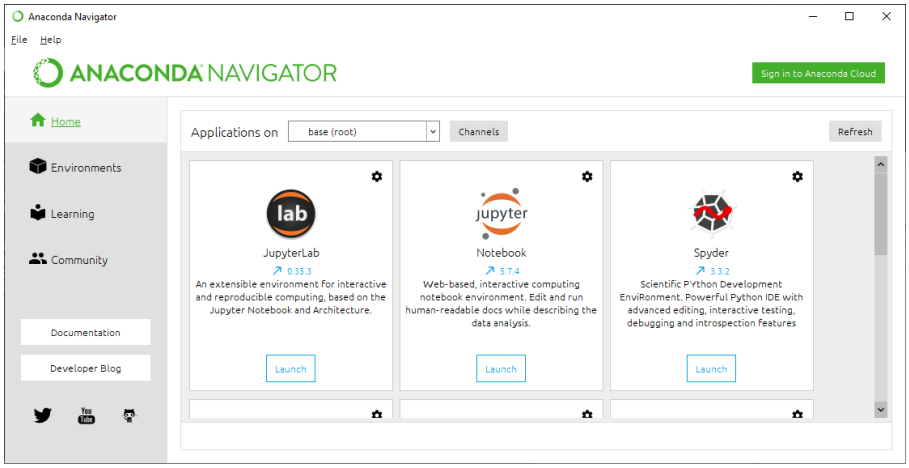


Figure 7 *Starting Jupyter Notebook from Anaconda Navigator*

Figure 8 *Jupyter Notebook File Browser Window*

Example files for this tutorial can be edited and run by clicking on names from the Jupyter Notebook file listing window.

3 DATA ACQUISITION

Unstructured text data can be acquired from several sources. It can be input directly into Python code. It can be read from files including CSV, Excel, and JSON. Or it can be scraped from web sites or obtained using an API such as the one provided by Twitter. The following sections provide examples of each.

3.1 Data Loaded Directly into Python

Perhaps the least efficient way to acquire data for further analysis is by pasting it directly into Python code. This method can be useful when smaller data sets are used or when cutting and pasting is the only method available. The following lines of code provide an example of how this might appear:

```
Text1 = "Line 14 produced 12 errata pieces today."

Text2 = "It was a hard day at work, we had 3 people die in the emergency room."

Text4 = “Nothing happened at all. It was very boring.”
```

3.2 Data Loaded from a File

A more commonly used method is to read existing information from a file. The following lines of code demonstrate this for data contained in a csv file:

```
# Create an empty list
DataList = []

# Open a CSV file
openfile = open('data/facultysalary.csv', 'r', encoding = "ANSI")

# Read the file into an object
r = csv.reader(openfile)

# Process each line and append to a list
for i in r:

       # Append the data to the list
        DataList.append(i)

# Close the file
openfile.close()
```

Other file types can be used as well. For example, JSON is a very popular file format, particularly when web services or API interfaces are used. The following code provides example JSON formatted data (that data would be found in an external file rather than in Python code):

```
data = [

{
"display": "Student JavaScript Tutorial",
"url": "http://www.w3schools.com/js/default.asp"
},

{
"display": "Student HTML Tutorial",
"url": "http://www.w3schools.com/html/default.asp"
},

{
"display": "Student CSS Tutorial",
"url": "http://www.w3schools.com/css/default.asp"
} ]
```

3.3 Data Loaded from Twitter or other APIs

Many social media services and other information-rich sources offer APIs (application program interfaces) which include routines, protocols (e.g. rules), and tools for accessing data and other objects. For example, Twitter provides an API for gathering Tweets. To collect Tweets, it is necessary to get a Twitter Developer account. The following steps detail methods for doing this:

1) Obtain a regular Twitter Account (if you don't have one already);
2) Go to the Twitter Developer Website here:

   `https://developer.twitter.com/en/apply-for-access.html`

3) Click on the Apply Button as shown in Figure 9;

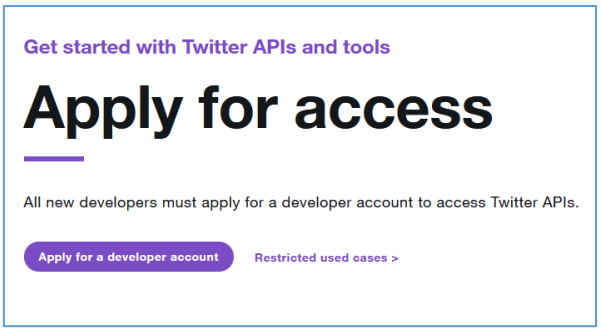


Figure 9 *Apply for Twitter Developer Account*

4) Be sure to indicate your account is for academic use. More information about the application process can be found here: https://docs.inboundnow.com/guide/create-twitter-application/.

Once the Twitter Developer account is approved (which may take several days), you will have access to API keys which permit you to acquire data. To access your API keys, login to your account at: developer.twitter.com and then go to *App* as shown in Figure 10.

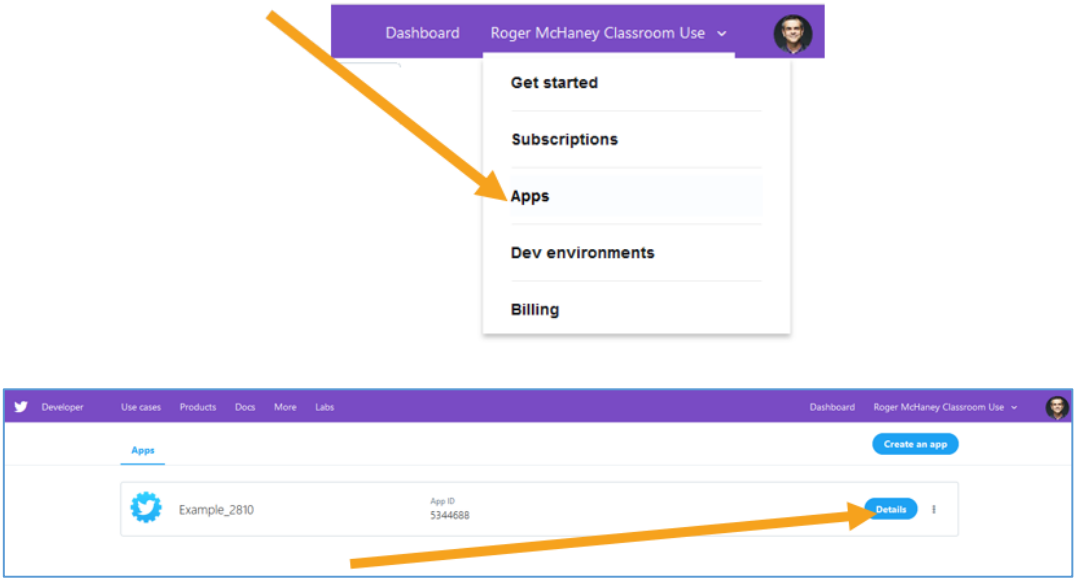


Figure 10 *Getting to your Twitter API keys. First Apps, then to Details*

Then, the Keys and Tokens can be accessed and used in Python scripts or other programmatic ways. See Figure 11

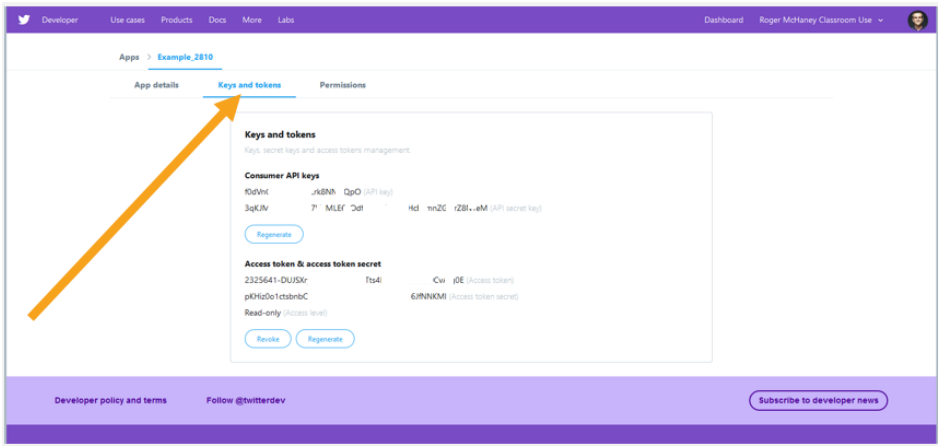


Figure 11 *Get Keys and Tokens for Use in Scripts and Software*

Figure 12 provides a screen shot of a Python script that uses Twitter API keys to aid in data collection. The script can be customized to collect data based on hashtags, keywords, users or other items of interest.

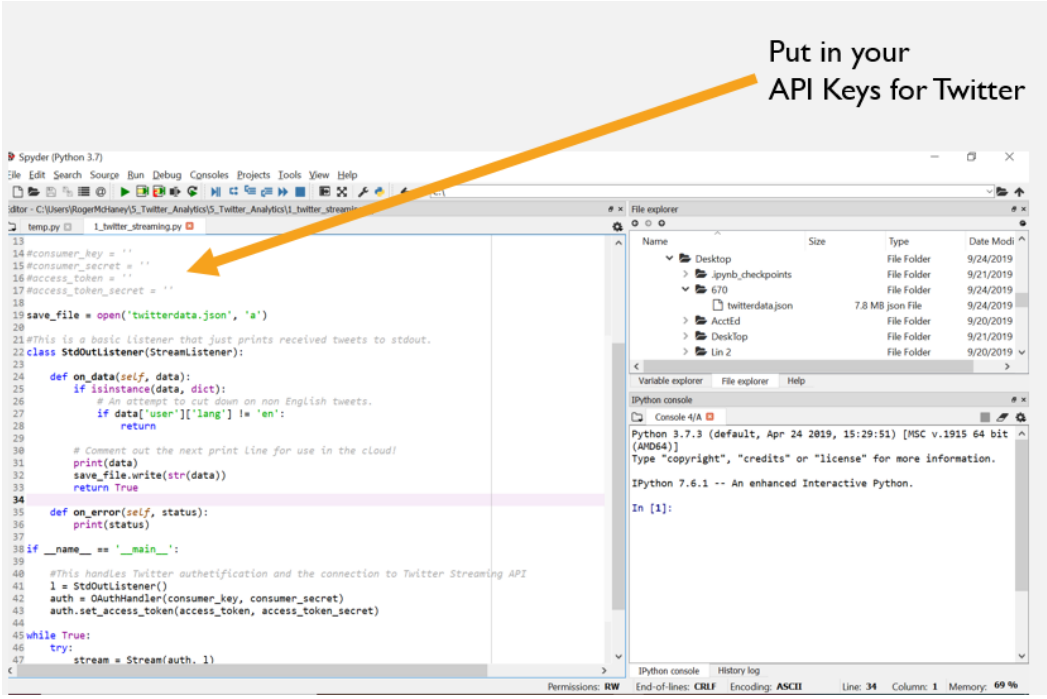


Figure 12 *Use API keys Python Script to Aid with Data Collection*

For applications that might require a long period of time for data collection, running the script in a cloud environment may be necessary. Google, Microsoft, and Amazon all provide cloud facilities that can host data analytics applications. For this tutorial, we use a free cloud-based environment called PythonAnywhere (https://www.pythonanywhere.com). Figure 13 shows where a limited version of this cloud-based tool can be accessed with a free account.

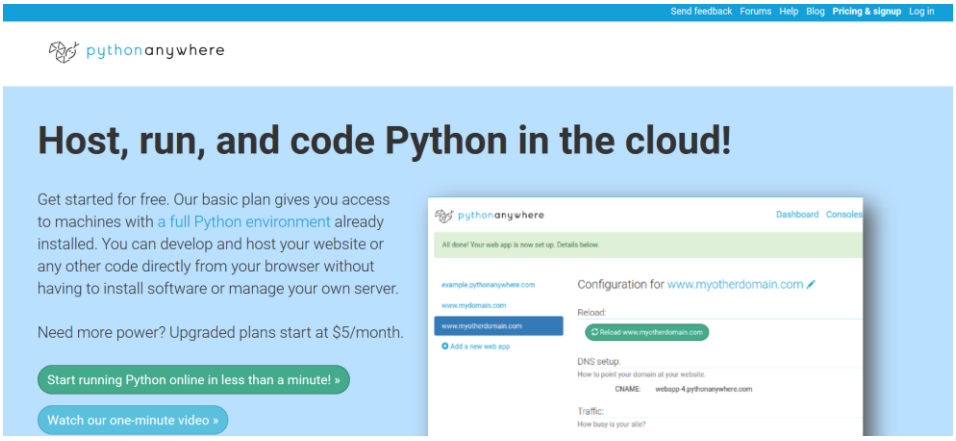


Figure 13 *Use PythonAnywhere to Run Data Collection for Long Periods of Time*

3.4 Data Loaded from Web Scraping and Crawling

Another source of data used in unstructured data analytics is web pages. Data that is displayed on a website can be either scraped or crawled depending on the way the HTML is structured. Web scraping generally refers to obtaining data from a single page whereas crawling involves setting up a program that scraps multiple pages. Web scraping can be accomplished using a query tool called XPATH. This tool aids in drilling into the page's HTML structure and programmatically retrieving desired information. Figure 14 provides an example of XPATH. XPATH code can be tested using a sandbox such as the one found at: https://www.freeformatter.com/xpath-tester.html

//li[a]	Selects the `li` elements which enclose an `a` element.

Figure 14. *XPATH Statement*

3.5 Data Loaded from Files

Often, data is obtained from an organization's data warehouse or other sources and may be found in JSON, CSV, or Excel files. Each can be readily imported using a set of Python commands. Figure 15 provides an example.

```
# Define list
sms = []

# Get data from csv file
openfile = open('data/smsdata_classification.csv', 'r')

r = csv.reader(openfile)

# Put into list
for i in r:
    sms.append(i)
    sms_data.append(i[0])
    sms_labels.append(i[1])

openfile.close()
```

Figure 15 *Sample Code to Load CSV File*

4 DATA CLEANING

A primary activity of data analytics is to clean captured data. In most cases, the data will contain a variety of special characters, white space and other formatting that must be removed. Python provides several tools for accomplishing these tasks so 'clean' data can be stored in lists or data frames for later analysis. Pandas is a popular Python library with tools for cleaning data, fixing missing values, organizing data, and changing data types. Other tools can be used as well. For instance, Python has functions for making letters upper or lower case. Likewise, Python includes a library of commands, called RegEx, or Regular Expressions, that are sequences of characters used to form search patterns or perform character replacement/substitutions in strings. Figure 16 provides a brief example of built-in functions and a RegEx string to clean data.

```
# Lowercase

tokens = tokens.lower()

# Remove useless numbers and alphanumerical words

tokens = re.sub("[^a-zA-Z0-9]", " ", tokens)
```

Figure 16 *Sample Code to Help Clean Data*

5 DATA ANALYSIS

After collecting data and ensuring it has been cleaned and made ready for analysis, several different things can be done with it. Descriptive statistics can be derived from unstructured data. These include frequency distributions for words, topics, sentences, and other tokens. Likewise, word clouds can be developed to communicate outputs based on frequency.

Many times, the data includes words that do not contribute to the meaning of the analysis. These are words are known as stop words and can be removed from the data. Other words have different forms due to plurals, tense and other features. They can be combined into a common form for analysis. This is done through stemming or lemmatizing. Code examples provided in the tutorial provide examples of performing these operations in ways useful to simulation practitioners.

Following cleaning, the normalized data set can be used for sentiment analysis, which provides information about positive or negative feelings, future intentions, and topics related to sentiments. Other text analytics can be performed to classify unstructured data into categories using machine learning techniques, text analysis, topic modeling, and social network analysis. These types of analyses can be useful in development of simulations and models. Specific code examples will be provided in the tutorial.

6 EXAMPLES RELEVENT TO SIMULATION

Unstructured data can be useful in simulation models, particularly when dealing with human elements and decision making. For instance, real world review data can be used to derive a distribution of happy versus unhappy customers. The distribution can be incorporated into a model to determine different patterns of behavior. Likewise, intention analysis, can be used to drive model entity behavior according to real world distributions.

Other useful information relevant to simulations includes timing data. In other words, the rate with which comments are made in Twitter or other social media platforms is useful. Modeling consumer behavior can be based on product reviews or other information from a variety of text-based sources. Likewise, text analytics and topic modeling can be used as sources for data built into simulation models. For example, text analytics provides classification models. Various attribute values attached to simulation entities can be used for categorization and ultimately direct entities down different paths

within a model. Topic modeling can be used to better understand dynamics within a corpus of unstructured data. For instance, using topic modeling to analyze adverse medical event data can help modelers structure their simulation in ways that match reality more closely, and provide categories and distributions for the occurrence of particular events (Zhu et al., 2019). The current tutorial will provide code examples for each of these categories to further illustrate usefulness to simulation development. Three examples are provided in the following subsections.

6.1 Sentiment Analysis to Drive Model Human Factors

Sentiment analysis uses computation techniques to identify and classify opinions expressed in text messages or social media. The outcome determines the attitude of the writer toward the subject of the analyzed message. Usually the outcome will be positive, negative or neutral. The current tutorial will demonstrate how to scrape text from a set of product reviews then use that material to build a distribution for use in a model. As brief example, Figure 17 shows use of vaderSentiment to build scores based on customer comments. The resulting scores are loaded into EasyFit to develop a distribution (Figure 18) for use in a model where a decision point is located. As can be seen, the customers either loved or hated the product.

vaderSentiment

- http://www.nltk.org/howto/sentiment.html

In [1]:
```
# Import Needed Libraries from Natural Language Tool Kit
import nltk
from nltk.corpus import stopwords

# Sentiment analysis items required (e.g Vader)
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

In [2]:
```
comments = ["I hate this product and its taste.",                  # positive sentence example
            "I absolutely love wheatabix.",                         # positive sentence example
            "Wheatabix is my top choice and I like it alot",        # punctuation emphasis handled correctly (sentiment intens
            "Buy me more Wheatabix! Healthy and delicious.",        # booster words
            "Nasty stuff that Wheatabix. DAMN nasty.",              # emphasis for ALLCAPS handled
            "Wheatabix is my fav! Hate it!!!!",                     # combination of signals - VADER appropriately adjusts int
            "Wheatabix sucks and I hate it. FRIGGIN Nasty stuff!!!",  # booster words & punctuation
            "Wheatabix is good.",                                     # positive sentence
            "Wheatabix is kind of good.",                             # qualified positive sentence is handled correctly
            "Wheatabix was good, but the consistency is unappealing and the flavor is not great.", # mixed negation sentence
            "At least it isn't a horrible cereal.",                   # negated negative sentence with contraction
            "Make sure you :) or :D today when you eat wheatabix!",   # emoticons handled
            "Wheatabix SUX!",                                         # negative slang with capitalization emphasis
            "Wheatabix only kinda sux! But I'll get by on it, lol"    # mixed sentiment example with slang and constrastive c
             ]

analyzer = SentimentIntensityAnalyzer()
# Get Sentiment Scores for Each Data Item
for comment in comments:
    vs = analyzer.polarity_scores(comment)
    print(vs["compound"])
```

```
-0.5719
0.6697
0.5106
0.7888
-0.8918
-0.4344
-0.9136
0.4404
0.3832
-0.5409
0.431
0.8633
-0.5461
0.2228
```

Figure 17 *Python VaderSentiment Application*

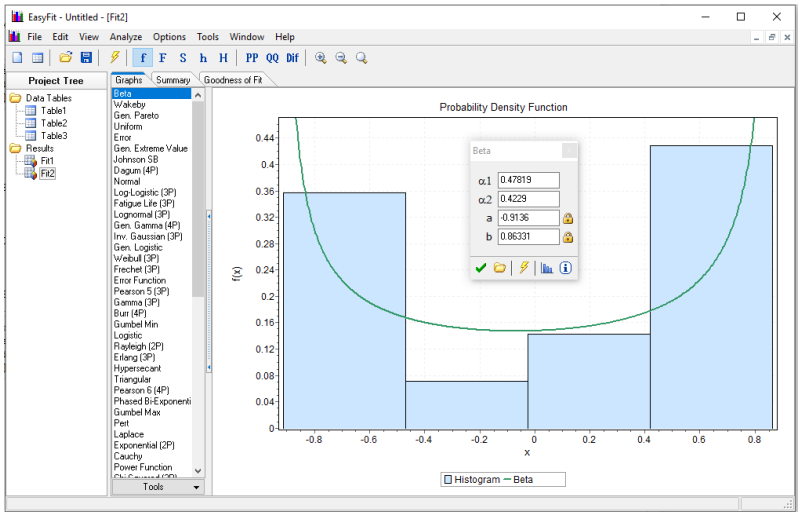


Figure 18 *Data from Sentiment Analysis Fit to Beta Distribution*

6.2 SNA Approach to Structure Model

Social network analysis investigates social structure and formation with network development based on graph theory. The result of a network analysis is a set of nodes (e.g. actors or entities) and edges (e.g. relationship or links). SNA is useful in model structure development and can help inform developers of an agent-based network or communication within a network of entities. Figure 19 provides a glimpse of a network of station communications developed using the open source SNA tool, Gephi. Communication patterns and frequencies can be derived and used to structure a simulation model.

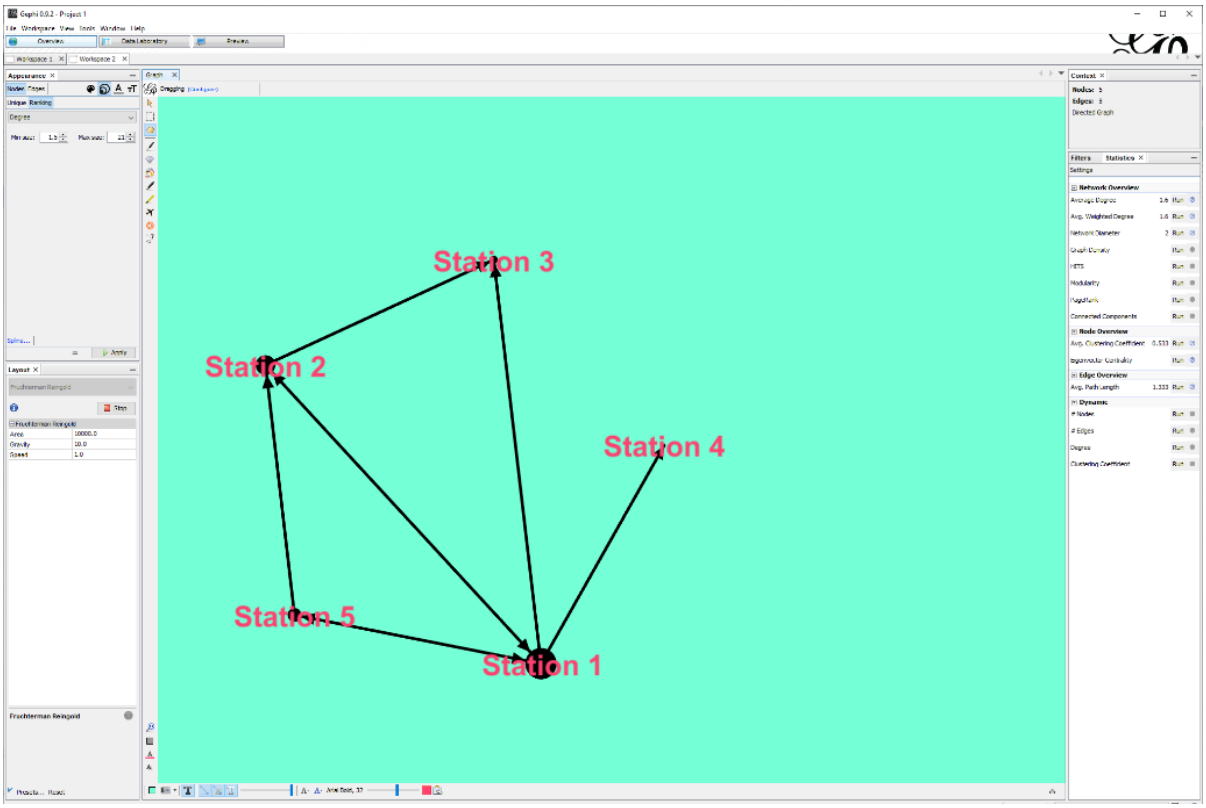


Figure 19 *Using Gephi to Develop Model Structure*

6.3 Twitter Data to Derive Model Timing Sequence

Another example of deriving value from text mining can be illustrated with extracting timing information from Twitter's JSON-formatted, Tweet data. In this example, Python was used with a cloud service to collect Twitter data using the hashtag, simulation. The tweets were mined for their creation times and this was used to develop a distribution for interarrival time. Figure 20 shows the Twitter collection code and Figure 21 shows the collected times. Finally, Figure 22 provides the output for fitting the distribution using EasyFit software.

```
In [ ]: # Action 3 Required: pip if needed
        !pip install algorithmia

In [1]: # Use Twitter Developer API Info to Collect Tweets!
        import Algorithmia

        input = {
          "query": "simulation",
          "numTweets": "50",
          "auth": {
              "app_key": "f0dVn[redacted]",
              "app_secret": "3q[redacted]Z8NYeM",
              "oauth_token": "2[redacted]CvAfg0E",
              "oauth_token_secret": "pKHiz0o1ctsbnb0[redacted]e6JfNNKMl"
          }
        }
        client = Algorithmia.client('simBq26eRxHp151xckLDLL3dO6Q1')
        algo = client.algo('twitter/RetrieveTweetsWithKeyword/0.1.5')
        print(algo.pipe(input).result)
        ended_entities': {'media': [{'additional_media_info': {'monetizable': False}, 'display_url': 'pic.twitter.com/WXtLAg173D
        ', 'expanded_url': 'https://twitter.com/Mr_DrinksOnMe/status/1194458362504732672/video/1', 'id': 1194325008241299457, 'id
        _str': '1194325008241299457', 'indices': [62, 85], 'media_url': 'http://pbs.twimg.com/ext_tw_video_thumb/1194325008241299
        457/pu/img/9bXIOUQsDF0LLGFP.jpg', 'media_url_https': 'https://pbs.twimg.com/ext_tw_video_thumb/1194325008241299457/pu/img
        /9bXIOUQsDF0LLGFP.jpg', 'sizes': {'large': {'h': 720, 'resize': 'fit', 'w': 408}, 'medium': {'h': 720, 'resize': 'fit', '
        w': 408}, 'small': {'h': 680, 'resize': 'fit', 'w': 385}, 'thumb': {'h': 150, 'resize': 'crop', 'w': 150}}, 'type': 'vide
        o', 'url': 'https://t.co/WXtLAg173D', 'video_info': {'aspect_ratio': [17, 30], 'duration_millis': 7834, 'variants': [{'co
        ntent_type': 'application/x-mpegURL', 'url': 'https://video.twimg.com/ext_tw_video/1194325008241299457/pu/pl/Te-0CKLQqAO_
        cxXS.m3u8?tag=10'}, {'bitrate': 632000, 'content_type': 'video/mp4', 'url': 'https://video.twimg.com/ext_tw_video/1194325
        008241299457/pu/vid/320x564/FR6mlk4n01YZOonm.mp4?tag=10'}, {'bitrate': 832000, 'content_type': 'video/mp4', 'url': 'http
        s://video.twimg.com/ext_tw_video/1194325008241299457/pu/vid/408x720/iZHtP1wE2WknKbev.mp4?tag=10'}]}}}]}, 'favorite_count':
        162457, 'favorited': False, 'geo': None, 'id': 1194458362504732672, 'id_str': '1194458362504732672', 'in_reply_to_screen_
        name': None, 'in_reply_to_status_id': None, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id': None, 'in_reply_to_
        user_id_str': None, 'is_quote_status': False, 'lang': 'en', 'metadata': {'iso_language_code': 'en', 'result_type': 'recen
        t'}, 'place': None, 'possibly_sensitive': False, 'retweet_count': 68839, 'retweeted': False, 'source': '<a href="https://
        about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>', 'text': 'Learning to fly a helicopter. Lucky it was
        just a simulation. https://t.co/WXtLAg173D', 'truncated': False, 'user': {'contributors_enabled': False, 'created_at': 'S
        un May 09 13:43:13 +0000 2010', 'default_profile': False, 'default_profile_image': False, 'description': 'This is our pla
        ce, we make the rules. Web Designer at @MalavKarkar. Instagram: mrdrinksonme_. For business or endorsements: ☐ hello@mr
        drinksonme.com', 'entities': {'description': {'urls': []}, 'url': {'urls': [{'display_url': 'malavkarkar.com', 'expanded_
```

Figure 20 *Python Code to Collect Twitter Data*

```
In [5]: # View the Tweet Times

        for i in data[:100]:
            print(i['created_at'])

        Thu Nov 14 16:27:43 +0000 2019
        Thu Nov 14 16:27:43 +0000 2019
        Thu Nov 14 16:27:43 +0000 2019
        Thu Nov 14 16:27:42 +0000 2019
        Thu Nov 14 16:27:41 +0000 2019
        Thu Nov 14 16:27:40 +0000 2019
        Thu Nov 14 16:27:37 +0000 2019
        Thu Nov 14 16:27:37 +0000 2019
        Thu Nov 14 16:27:35 +0000 2019
        Thu Nov 14 16:27:35 +0000 2019
        Thu Nov 14 16:27:34 +0000 2019
        Thu Nov 14 16:27:31 +0000 2019
        Thu Nov 14 16:27:31 +0000 2019
        Thu Nov 14 16:27:28 +0000 2019
        Thu Nov 14 16:27:28 +0000 2019
        Thu Nov 14 16:27:27 +0000 2019
        Thu Nov 14 16:27:27 +0000 2019
        Thu Nov 14 16:27:27 +0000 2019
        Thu Nov 14 16:27:27 +0000 2019
        Thu Nov 14 16:27:27 +0000 2019
        Thu Nov 14 16:27:27 +0000 2019
```

Figure 21 *Collected Time Data from Twitter*

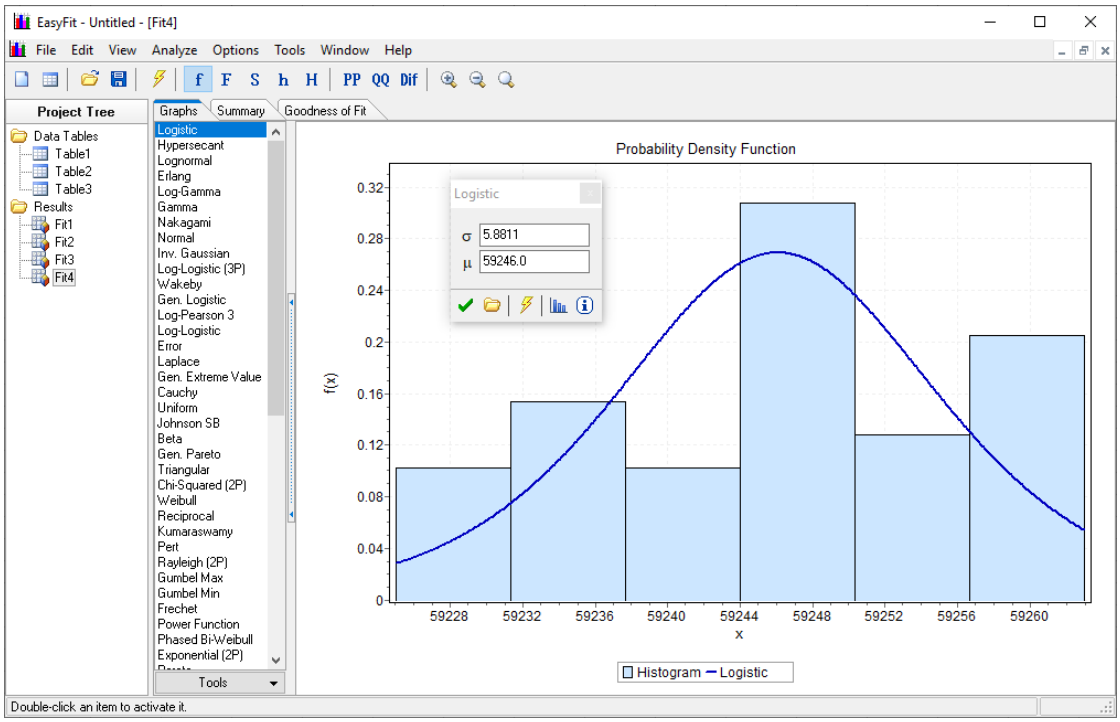


Figure 22 *Fitting Distribution for Interarrival Time using EasyFit Software*

7 VISUALIZATION AND STORY TELLING

Big data visualization and storytelling tools such as Power BI or Tableau are useful in both simulation model input data analysis and to provide output data reporting and analysis. Both tools contain a wide range of visualizations that help ensure data is more easily understood. Simulations are excellent tools for generating big data visuals to illustrate how outputs can add another dimension to a model's usefulness. This tutorial will provide examples.

8 CONCLUSIONS

The purpose of this tutorial is to provide a working set of tools specific for unstructured data analysis for use by simulation practitioners. The examples use the Python programming language with Jupyter Notebook from Anaconda. Specific examples derived from descriptive analytics, text analytics, topic modeling and sentiment analysis were demonstrated with data from sources that included web scraping, Twitter API, and CSV files.

ACKNOWLEDGMENTS

The author would like to acknowledge Dr. Bongsug Chae from Kansas State University for developing some of techniques described in this tutorial.

REFERENCES

Baines, T., Mason, S., Siebers, P.-O., & Ladbrook, J. (2004). Humans: the missing link in manufacturing simulation? *Simulation Modelling Practice and Theory*, **12(7–8)**, 515–526.

Greasley, A., & Owen, C. (2018). Modelling people's behaviour using discrete-event simulation: a review. *International Journal of Operations & Production Management*, **38(5)**, 1228–1244.

Gupta, B., Sharma, S., & Chennamaneni, A. (2016). Twitter Sentiment Analysis: An Examination of Cybersecurity Attitudes and Behavior. In *Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium Innovations in Data Analytics.* **17** (pp. 1–10). Dublin, Ireland: AIS.

Khan, R., & Urolagin, S. (2018). Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data. *International Journal of Advanced Computer Science and Applications*, **9(6)**, 380–388.

McHaney, R., Tako, A., & Robinson, S. (2018). Using LIWC to choose simulation approaches: A feasibility study. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2018.04.002

Robinson, S. (2008). Conceptual modelling for simulation Part I: definition and requirements. *The Journal of the Operational Research Society*, **59(3)**, 278–290.

Robinson, S., Arbez, G., Birta, L. G., Tolk, A., & Wagner, G. (2015). Conceptual modeling: definition, purpose and benefits. In *Proceedings of the 2015 winter simulation conference* (pp. 2812–2826).

Tako, A. A., & Kotiadis, K. (2015). PartiSim: A multi-methodology framework to support facilitated simulation modelling in healthcare. *European Journal of Operational Research*, **244(2)**, 555–564.

Zhu, L., Reychav, I., McHaney, R., Broda, A., Tal, Y., & Manor, O. (2019). Combined SNA and LDA methods to understand adverse medical events. *International Journal of Risk & Safety in Medicine*, (Preprint), 1–25.

AUTHOR BIOGRAPHY

ROGER MCHANEY is a Professor of Management Information Systems and University Distinguished Teaching Scholar at Kansas State University, College of Business. He holds the Daniel D. Burke Chair for Exceptional Faculty. McHaney is the author of books on simulation, education, and health technology. Prior to becoming an academic, he developed simulations primarily in manufacturing and materials handling. His current research includes understanding how simulation and modelling techniques can best be applied in business organizations, and data-driven simulations. His email address is mchaney@ksu.edu.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, Eds.
DOI: https://doi.org/10.36819/SW21.008

AGENT-BASED MODELS: A TUTORIAL

Dr Duncan A. Robertson

School of Business and Economics
Loughborough University, Loughborough, LE11 3TU
d.a.robertson@lboro.ac.uk
http://www.duncanrobertson.com

ABSTRACT

We introduce agent-based modelling in this tutorial paper. We introduce the concepts of agent, emergent behaviour, and show these concepts in three different agent-based models.

Keywords: Tutorial paper, agent-based modelling, techniques

1 INTRODUCTION

Agent-based modelling ('ABM'), or multi-agent simulation, is a technique where individual 'agents' are modelled, where the behaviour of these agents combine to make up the overall system. This micro-level modelling can be contrasted with modelling stocks and flows in system dynamics ('SD') models or modelling systems as a series of events in discrete event simulation ('DES').

While agent-based modelling is somewhat intuitive, in that interactions between individual agents is the fundamental building block of the modelling technique, agent-based modelling still requires a degree of high-level computer coding rather than using drag-and-drop building blocks to create a model.

2 AGENT-BASED MODELS: FUNDAMENTALS

Agent-based models are a form of simulation, where individual components of the system are modelled. However, these components – or agents – can have autonomy. They can take into account the environment, or the interactions with other agents, in order to allow their behavior to *adapt*. Unlike SD models, where the movement of objects through a system is not explicitly modelled, ABM uses these agents as the building blocks of the simulation: it is a bottom-up rather than top-down methodology. And unlike DES, agents within an agent-based model can have autonomy rather than being directed through a system by a controller.

This tutorial will introduce three agent-based models to show the fundamental building blocks of an agent-based model.

2.1 The Agent: Autonomy

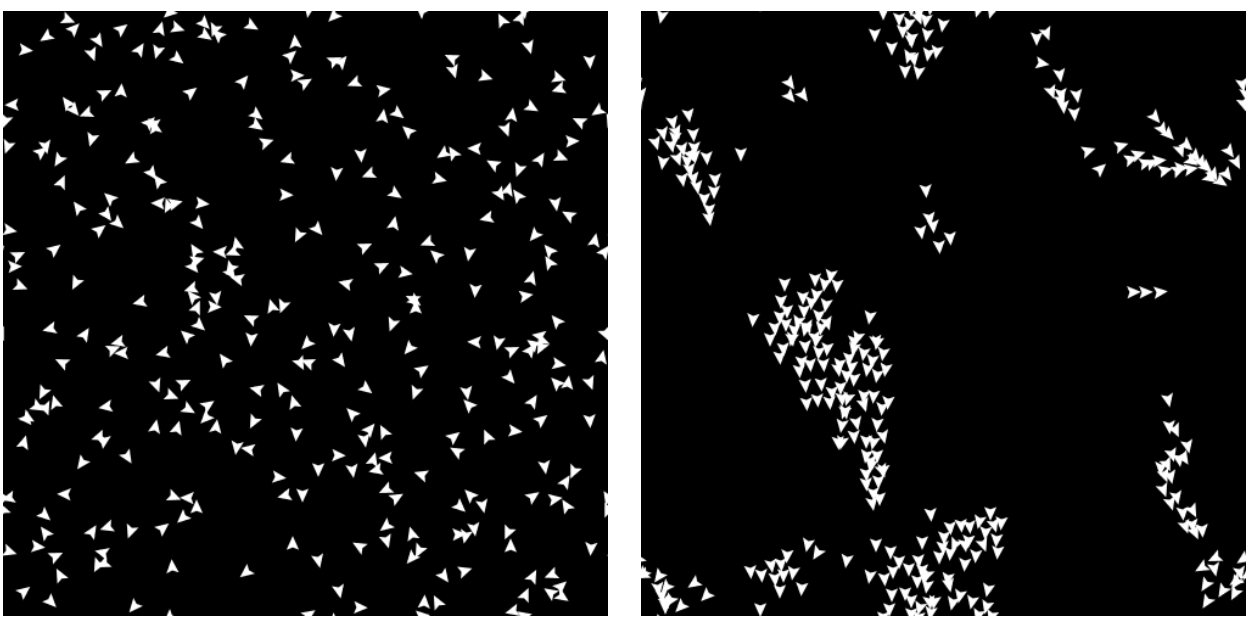
Agents are the fundamental components of agent-based models. These are (or can be) autonomous, in that their behaviour is controlled by the agent themselves, rather than being directed: they can 'think' for themselves. As such, the use of agent-based models has been used by researchers in several disciplines to model social behaviour as well as the behaviour of physical systems.

'Thinking' in an agent-based modelling context is a term that encompasses the ability of agents to respond to stimuli; it does not necessarily mean that the agent has cognitive ability. Take for example the interaction of three masses that experience gravitational or electrostatic attraction to each other. In each case, the bodies – or agents – follow forces proportional to the inverse of the square of the distance between the objects. In an agent-based model, the direction of the force vector could be computed *by each agent* and the resulting direction of movement calculated. In this case, however, the agents do not have cognition, they merely respond to the force vector to which they are exposed: the masses do not know the location of the other masses, and therefore cannot be expected to take this into account when reacting to their position. Agent-based modelling is concerned with modelling the social, physical, psychological, behavioural 'forces' that affect the state or actions of an individual.

Agents within agent-based models are autonomous, in that their behaviour is modelled as being controlled by the agent themselves without any over-arching controller of the system: in this way, the macro-level model behaviour in built up from the individual actions of the agents that collectively comprise the system.

2.2 Emergence

An early use of agent-based models was to model the spatial interaction of birds. One can observe the phenomenon of flocking in bird populations, but the modelling of this using conventional simulation techniques is difficult. Reynolds (1987) attempted to model the flocking behaviour not by modelling the entire flock, but by simulating each individual bird. These birds do not have access to information about the position, speed, and direction of each bird in the population, but only about their *local neighbourhood*, parameterized by a radius and angle of observation. By imposing simple rules on the behaviour of individual birds, namely separation (steer to avoid other birds); alignment (steer towards to the average heading of birds within the local neighbourhood); and cohesion (move towards the position of birds within the local neighbourhood), 'lifelike' flocking behaviour is simulated.



(a) Start of Simulation (b) During the Simulation

Figure 1 *Flocking Model showing Emergent Flocking Behaviour*

In Figure 1, birds are initially positioned at random on the space (note that the space wraps around so that agents can spill over from the left to the right; from the top to the bottom, or vice versa – this is a torus, a

common topology for agent-based models in that it avoids hard boundaries and the necessity of modelling what happens at these boundaries). These bird agents then follow simple rules that take into account their local observations, which then leads to the system-wide emergent property of flocking.

2.3 Bounded Information

In the flocking model above, agents need not (and generally do not) have full information about the state of the system: they tend to be myopic, in that their knowledge of the system is local. This leads to boundedly rational behaviour (i.e. their behaviour may have been different if they had full information). As such, this can lead to interesting outcomes.

In Schelling's (1971) model of segregation – generally accepted as one of the first agent-based models, individuals in a city determine whether they are happy or unhappy based on the number of neighbours who are the same colour as themselves. Agents – who have a property of colour - are first arranged randomly on a grid. A model parameter determines the threshold at which agents are happy: each agent calculates the proportion of their neighbours who are a different colour to them: if this is above the threshold value, the agent moves to a vacant space, as shown in Figure 2

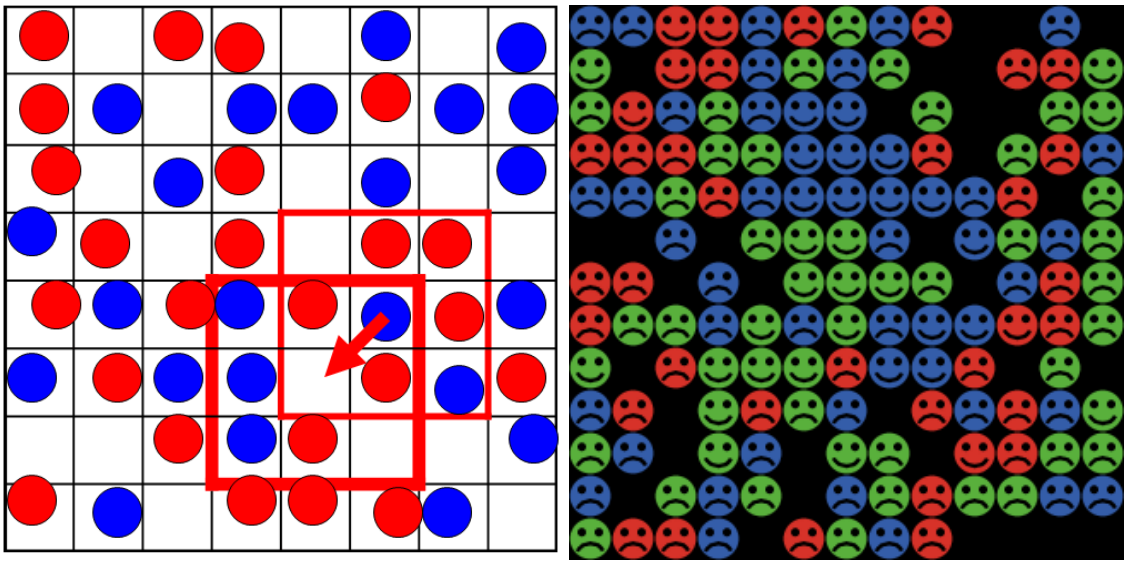


Figure 2 *Schelling's Segregation Model (images from duncanrobertson.com)*

Schelling's model is important in that it demonstrates that macro level behaviour can come from micro level interactions. In particular, the agents *self organize* into colour segregated clusters without the need of a central planner. In particular, if the threshold value is below 50%, i.e. agents are happy to be in a local minority, segregation still forms. This behaviour cannot readily be modelled using traditional techniques.

2.4 Applications in Management

There are several applications of agent-based modelling in business. Within the strategy field, individual firms' strategies can be modelled, and the effect on the business landscape and other firms' strategies can be modelled. Robertson (forthcoming) and Robertson and Caldart (2010) set out a model of interacting firms by using an agent-based model to show how strategic movement of one firm deforms the fitness landscape which in turn alters the strategic movement of other firms. Robertson (2019) provides an overview of Agent-Based Models in Strategic Management

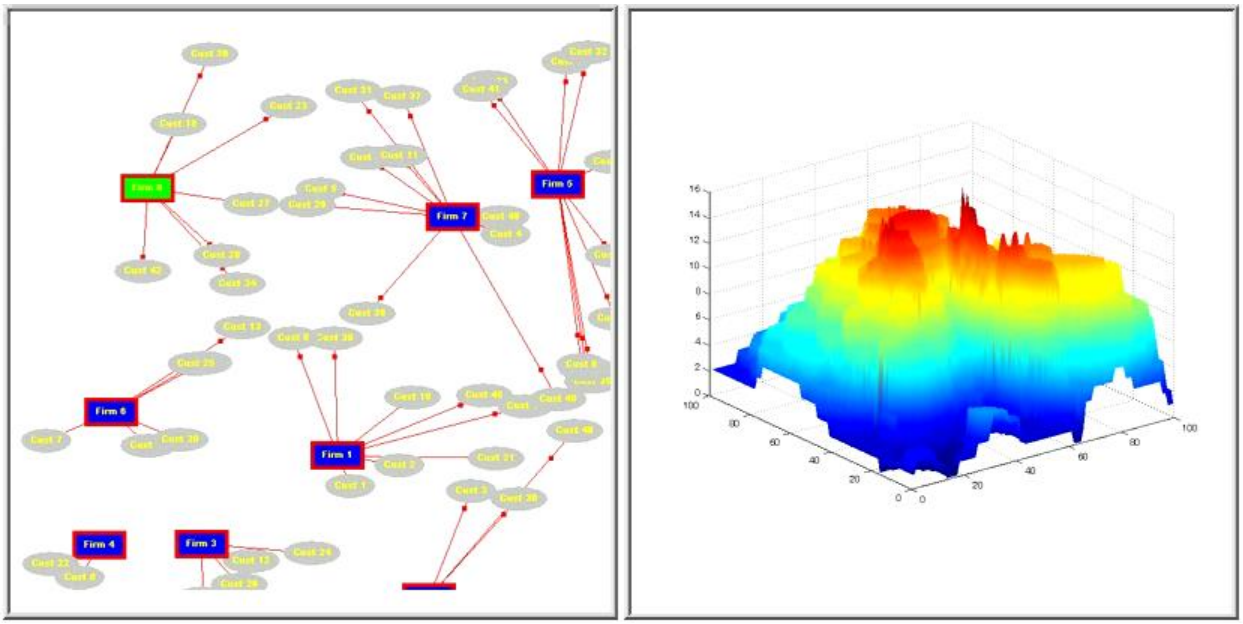


Figure 3 *Robertson's Dynamic Competition Model (image from Robertson (forthcoming))*

3 CONCLUSION

The concepts of emergence and self-organization are best demonstrated by the use of agent-based models. The models introduced in this tutorial other models will be introduced in the SW20 workshop, which is intended to be an interactive session where agent-based models can be developed interactively with workshop participants.

REFERENCES

Gilbert, N. and Troitzsch, K. (2005) *Simulation for the Social Scientist*, Buckingham: Open University Press

Reynolds, C. (1987) 'Flocks, Herds and Schools: A Distributed Behavioral Model', *SIGGRAPH '87: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. Association for Computing Machinery*, 25–34

Robertson, D. A. and Caldart, A. A. (2010) *The Dynamics of Strategy*, Oxford: Oxford University Press

Robertson, D. A. (2019) *Agent-Based Strategizing*, Cambridge: Cambridge University Press

Schelling, T. C. (1971) 'Dynamic Models of Segregation', *Journal of Mathematical Sociology*, **1**, 143-186

AUTHOR BIOGRAPHY

DUNCAN ROBERTSON is a member of the Management Science and Operations group at Loughborough University and a Fellow of St Catherine's College in the University of Oxford. He received his undergraduate degree in physics at Imperial College London and his DPhil at the Said Business School, University of Oxford. Further information and details about agent-based modelling can be found at http://www.duncanrobertson.com/

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.009

HYBRID SIMULATION WITH SIMUL8 - A HEALTH ECONOMICS EXAMPLE

Fiona Lindsay

SIMUL8 Corporation
29 Cochrane Street, Glasgow, G1 1HL
tbc@SIMUL8.com

Christoph Werner

SIMUL8 Corporation
29 Cochrane Street, Glasgow, G1 1HL
tbc@SIMUL8.com

ABSTRACT

Economic evaluations take place when a new drug or device is being taken to market and is also used by healthcare payers and providers when considering whether to adopt any innovation. Clinicians and administrators need to understand not just whether the new intervention is cost effective when applied to theoretical cohorts of patients, but also what it would mean when implemented in their own organization or health economy. Simulation allows modelers to simulate both the transition between disease states and their likely costs and QALYs as well as to understand how the individual patient will now use services and resources. Using a case study example, this paper demonstrates how health decision-makers can use SIMUL8 simulation software to create a hybrid model for this purpose.

Keywords: Hybrid simulation, SIMUL8, health economic modelling

1 INTRODUCTION

Discrete event simulation is the most widely used simulation technique in industry, but there are also some scenarios where agent or continuous simulation are the right tool for the job. In some use cases, there are also real benefits to combining techniques to create a hybrid simulation. SIMUL8 is a multi-method simulation tool that incorporates, continuous and agent functionality to give users the power to model any scenario. This whitepaper will talk through a use case of health economic modelling to show the benefits of using a hybrid approach and how to achieve this using SIMUL8.

2 THE USE CASE FOR USING HYBRID SIMULATION IN HEALTH ECONOMICS

Research publications are recognizing that economic modelling which does not take into account waiting times and delays and their impact on the effect of treatment and costs can be erroneous. Discrete event simulation combined with agent simulation is being increasingly recommended as an economic evaluation technique.

Clinicians and administrators need to understand not just whether the new intervention is cost effective when applied to theoretical cohorts of patients, but also what it would mean when implemented in their own organization or health economy.

Simulation allows modelers to simulate both the transition between disease states and their likely costs and QALYs as well as to understand how the individual patient will now use services and resources. Decision-makers will need to be assured on all these points prior to approving a business case for implementation of a new way of working or the adoption of a drug or device. Ron Shannon from GHE lists the following reasons to use discrete event simulation, it:

- represents clinical reality
- presents the course of disease naturally with few restrictions
- is flexible: no mutually exclusive branches or states required
- follows the natural concept of time, the simulation clock keeps track of the passage of time (no fixed cycles)
- offers flexibility for handling perspectives and sensitivity analyses

- permits transparency
- allows queuing (e.g., if a health resource is not available at a given time)
- enables modeling of limited resources, bottlenecks, if applicable to the problem
- defines patients as explicit elements with specific attributes (e.g., sex, age, event history) that can be modified over time
- provides the option of updating variables continuously or at specific time periods
- and, in economic evaluations, discrete event simulation has the flexibility to accommodate a richer structure without making it unmanageable in size

While many health economists are using discrete event simulation, others prefer Monte Carlo and Markov models. SIMUL8 simulation software combines state transition modeling with discrete event simulation to allow users to have the full flexibility to model key healthcare questions. This paper will show how to use a hybrid simulation technique to model challenges such as health economics.

3 SIMUL8 CASE STUDY - DISEASE MODEL

The most important part of your simulation will be modeling the disease/condition as the different stages of the disease will determine what treatments, costs or resource are required by the patient. The simplest way to do this in SIMUL8 is to use State Charts.

3.1 State Charts

State Charts allow you describe what is happening to the patient in a supplemental way to a typical discrete process. This significant benefit of using the method is that it removes the need for complicated logic. They describe a series of states (conditions) through which patients will flow. Typically the flow through a series of states is not related to the physical position of the patient. Instead we can link the physical position of the patient by triggering events in a discrete process which depend on which state (condition) the patient is in.

- *States* are used to describe the various condition which the disease can progress through. These could be Low, Moderate or Severe stages of a disease.
- *Decisions* are used when a patient transitions from the state but a decision as to what state they will transition too happens. Such as from the state Low 50% of patients will transition to Moderate and 50% to High.
- *Transition arrows* allow you to control how a patient transitions from one state to another. There are various options to how patient can transition such as chance per day or a distribution of time spent in state.

3.2 Building the Disease Model

To build your disease model simply drag and drop states and/or decision points from the building blocks tab at the left hand side of the . Once all the required states are on screen these can be connected up with transition arrows. To connect states with transition arrows hold down the Shift key and click and drag from state to state to create an arrow. Transition arrows can go both to and from states.

For this guide we will be using an example model to talk through the concepts, the example model is based on a project looking at the cost effectiveness of different treatments for Chronic Venus Leg Ulcers. The disease pathway is outlined in *Figure 1*.

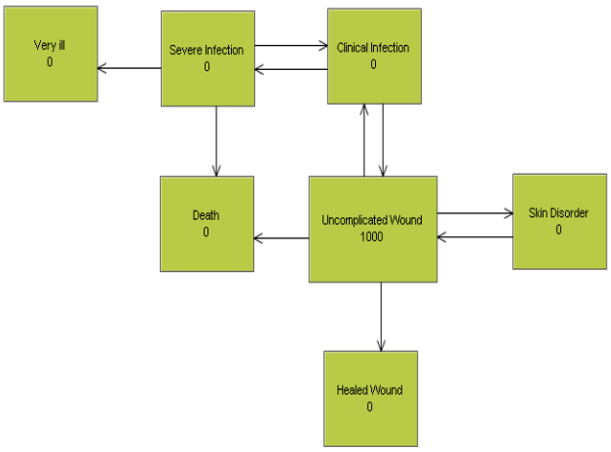


Figure 1 *Leg Ulcer Disease Pathway*

3.3 State Properties

Once you have dropped your required states on screen we can then populate these with properties. State properties are accessed in the same way as any other SIMUL8 object, by selecting the object and using the ribbon menus or by double clicking on the object.

Under the state preferences tab, the first box is where you can name the state. By doing so SIMUL8 will automatically create a Label also with this name as States require Labels to work. Alternatively you can select a pre-defined Label and SIMUL8 will automatically set the name of this state to the Label.

Also within this tab we can define start up values for the states, in the example model in *Figure 1* this is used to define the patient cohort that will be run through the model. All patients will start in the state Uncomplicated Ulcer, to do this enter a value in the box named 'Initial Content'.

3.4 State Transitions

The next step is to define how patients will transition from state to state, to enter these properties double click on a transition arrow. Select a transition type from the drop down menu, a short description of each transition type is provided below:

- *Chance each per time unit*: Enter the chance each patient has of transitioning per time unit.
- *Percent of all per time unit*: Enter the percentage of patients that will transition per time unit
- *Elapsed time in state*: Enter the time it will take for a patient to transition.
- *Rate (number per time unit)*: Enter the number of patients that will transition per time unit.
- *Time of day*
- *One every N time units*: Enter the time at which one patient will transition.
- *All every N time units*: Enter a time unit in which all patients will transition.
- *Always*: Patients entering this state will always transition.

When selecting a transition type it is important to remember that the transitions are based on the simulation time units. By default SIMUL8 sets the simulation time units to minutes. To change these settings go to the Data and Rules tab and select clock properties and select from the required unit from the top. For the ulcer example model select the setting 'Days'. In the ulcer example model there are three transition types used, the first is Chance each per time unit. The table below outlines the values and states that use these transitions.

Table 1 *Values and states using transitions*

From State	To State	Chance Per Day
Uncomplicated	Death	0.000131
Uncomplicated	Clinical Infection	0.000548
Uncomplicated	Skin Disorder	0.0000224
Clinical Infection	Severe Infection	0.0000167
Severe Infection	Death	0.000141

The transitions Clinical Infection and Skin Disorder back to Uncomplicated and Severe Infection to Very Ill are based on a distribution, meaning that patients are likely to spend a certain amount of time in that state before transitioning. To create distributions select the data and rules tab and distributions, use the data in *Table 2* to create the distributions required for the ulcer example.

Table 2 *Defining distributions*

Distribution Name	Time in State (Days)	Distribution Type
Clinical Infection	9	Average
Severe Infection	9	Average
Skin Disorder	8	Average

Once the distributions have been created, apply them to correct transition arrow. To do this, first select the transition type Elapsed time in state, use the (…) to open the formula editor and select the correct distribution. Distributions can be found by selection the 'Object' option at the left hand side. *Table 3* shows which distributions should be applied to each transition.

Table 3 *Distributions to apply to each transition*

From State	To State	Chance Per Day
Clinical Infection	Uncomplicated	Clinical Infection (9,Average)
Skin Disorder	Uncomplicated	Skin Disorder (8, Average)
Severe Infection	Very Ill	Severe Infection (9, Average)

The final transition is that from Uncomplicated to Healed Wound. This is based on a percentage chance of healing which changes each week the patient is receiving treatment. To set up this transition the heal time data needs to be entered into an internal spreadsheet. To do this select the Data and Rules tab, then the information store and create a new spreadsheet. Next, copy and paste the heal time data into the spreadsheet. The heal time data can be found in *Appendix A* of this paper.

The next step is to set the Healed Wound transition, double click the arrow and select the transition type percent of all per time unit. Like with distributions, use the (…) button to reference the spreadsheet. All spreadsheets are found in 'Information'. After selecting the spreadsheet the column and row from which to read the data needs to specified. The column will always be 2 and the row will be based on the week. There is a MATH function WEEK to control this which will automatically increment the row depending on what week the simulation is running by using the reserved SIMUL8 variable 'Simulation Time'. The images below outline how this should be populated.

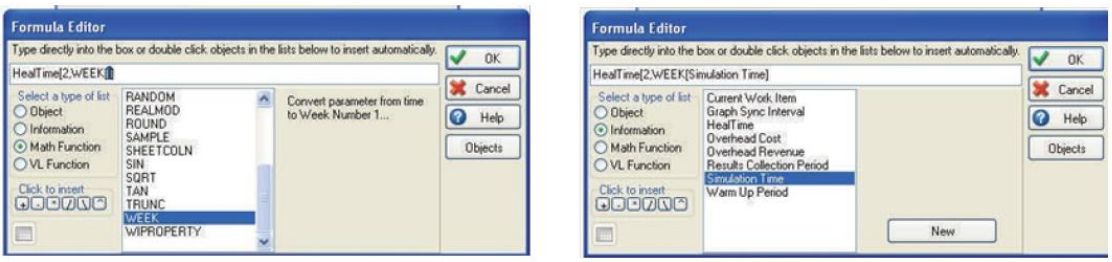


Figure 2 *How to populate the Formula Editor*

The disease model is now complete. Run the simulation for 1 year and you will see patients transition from one state to another. At the end of the run we are able to see the current state of patients and how this has changed from the initial cohort starting in an uncomplicated state.

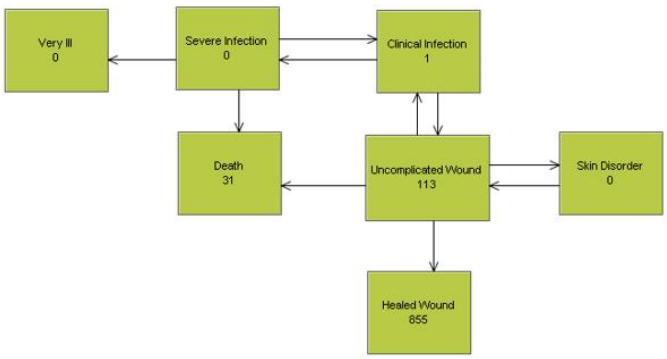


Figure 3 *Completed disease model results*

3.5 Sub States

Sometimes you may wish to have sub states within your disease model. By drawing a state wholly within another state, this means that a patient will be represented both in the main state and the sub state. In the ulcer example we have both a clinical infected state and a severe infected state but we may wish to capture information of all infected patients. By drawing a state around both the Clinical infected state and severe state this will allow us to do this. If you run the simulation you will see that patients will appear in both states.

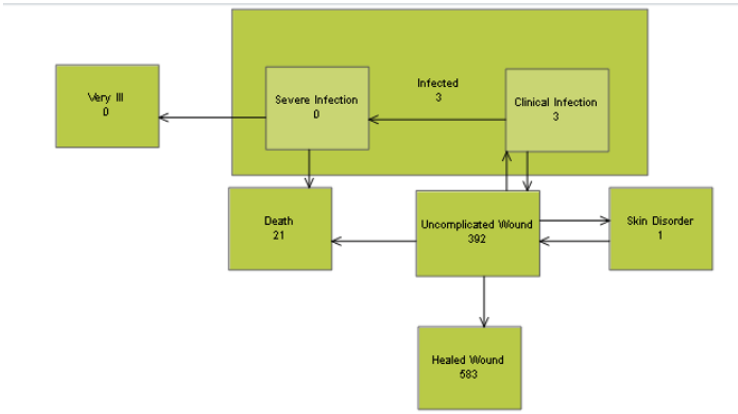


Figure 4 *Adding a sub state to the model*

3.6 Transition Decisions

These are useful when it is necessary to be able to branch a patients change in transition in a number of different paths. For instance, in the example model we are capturing how many patients are healed over the year but we may also want to capture reoccurrences and feed these back into the disease model. From the data we know that 10% of patients will have a reoccurrence in a year of being healed. Redraw the structure to include what is illustrated.

Firstly, you will need to redo the final transition rule from Uncomplicated to the new transition decision point.

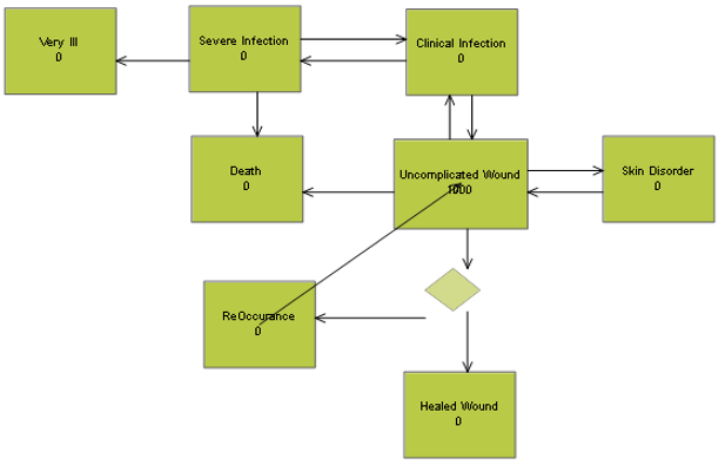


Figure 5 *Adding the new transition decision point*

By double clicking on the transition decision you can select from 3 routing options. Select percentage and enter 90% to Healed Wound and 10% to Reoccurrence. Lastly define the transition between Reoccurrence and uncomplicated wound, as this can happen at any point in the year select the 'Elapsed Time' option and use a uniform distribution between 1 and 365 days.

3.7 State Actions/Events

State events are a useful feature in being able to produce your own custom results or attaching information to individual patients. State actions/events can be accessed on the ribbon or by double clicking a state. For the ulcer example the total number of Severe Infections, Clinical Infections and Skin Disorders over the year needs to be displayed.

For this global number variables for each result need to be created. These are created in the information store, the same way as the spreadsheet was created. Once these have been created, select the State Action 'On Join' and chose the option 'Change Anything'. Now select the appropriate variable using the formula editor, global variables can be found in 'Information'. Next, choose the Action 'Increment' form the right hand side, this will then automatically increment the variable by 1 each time a patient enters the particular state.

You can also utilize SIMUL8's visual data feature so that you can see these results onscreen. To do so, go to the insert tab and select visual data. Next, click the area on the screen where you wish the data to be displayed and then select from the drop down the appropriate variable. Now run the simulation and you will be able to see on screen the total amount of patients who have entered that state.

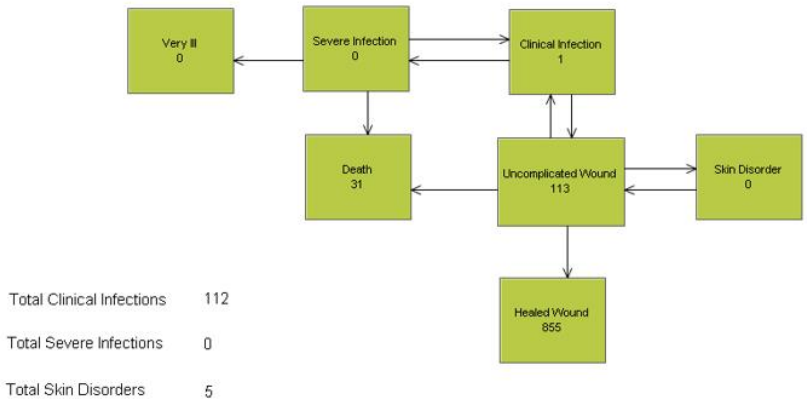


Figure 6 *Onscreen simulation results*

The same principal can be applied if you want to attach information to individual patients. For instance, if you were wanting to record utility values in order to calculate QALY's. In SIMUL8, Labels are used to do this. To create a Label, go to Data and Rules Tab and select Labels. Once you have created a Label, select a state and then as with the global variables you can select either 'Actions on Join' or 'Actions on Leave'. The value of the Label can then be changed to store particular information about that patient. In the ulcer example we can set the utility value for each patient depending on what state they are. In the image below we are setting the utility Label to 0.4 to represent the value associated with being the state 'Severe Infection'.

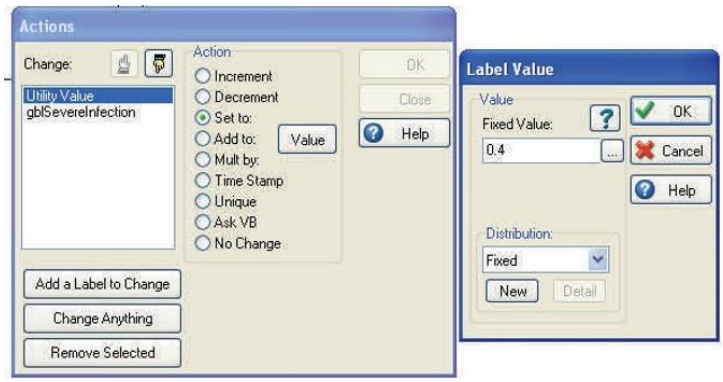


Figure 7 *Changing Label value*

In calculating QALY's, the amount of time spent in that state also needs to be recorded. Using another Label called 'Time In', the time at which a patient enters this state can be recorded using 'Action on Join'. Then on using 'Actions on Leave' we can calculate the QALY for that individual patient. First create a Label named 'QALY'. Select the 'Set to' option and chose fixed. You can then use the (…) button to open the formula editor. This allows you to enter a calculation, the below image shows the calculation used in this example. It takes the current simulation time and subtracts the Label value that contains the time that the patient entered the state. This is then multiplied by the utility value and this information is then stored in the Label 'QALY'.

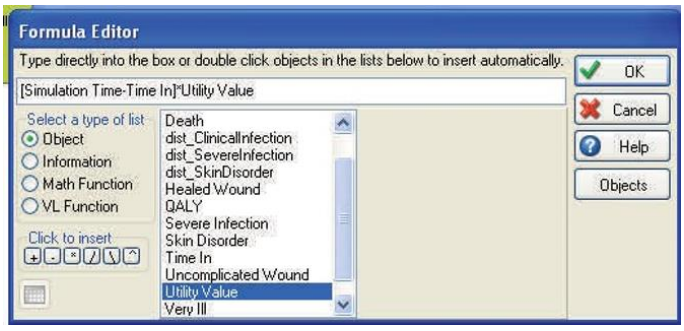


Figure 8 *QALY calculation*

3.8 Combining the Disease Model with a Discrete Process

State Charts are useful for modelling the disease pathway and how patients transition from different states but it is also useful to be able to combine these with traditional discrete process. For instance, in the ulcer example we want to simulate that the patient will have a hospital stay should they transition into Clinical Infection. To do this, first create the process outlined in *Figure 9* using the traditional SIMUL8 building blocks.

Figure 9 *Discrete process simulation model*

The next step is to trigger a patient arrival to this process. This can be done by selecting the Clinical Infection state and choosing the Events tabs. Here you can select from all start points in your model and this will cause the patient in the state to arrive at the discrete process.

Before running the simulation make sure that the check box 'Remove from all states' is unchecked. This is found by double clicking the End object in the process. If this is checked then when patients have gone through the treatment then they will also be removed from the State Charts. This is a useful feature if an outcome of the process would result in the patient being removed from the disease model. In this example, after treatment the patient will continue to transition in the disease model.

If you run the simulation as is you will get the results shown below, you will notice that the number of arrivals does not match the total number of clinical infections recorded in our results. This is because currently we only have the capacity to treat 1 patient at a time in our discrete process. As you can see from the queue we have a large build up of patients waiting for a bed. Patients are waiting so long for a bed that they have transitioned out of the clinical infection state. At some point over the year they have transitioned back to an infected state but this time an arrival will not be triggered as in reality they are still waiting on a bed.

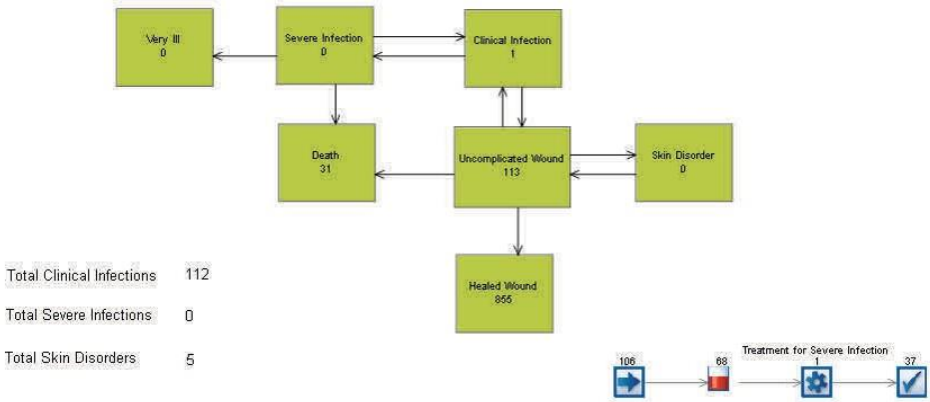


Figure 10 *Combining the disease model with the discrete simulation*

This is one of the key benefits of using simulation as you can now start to explore the effect of capacity on your outcomes. By selecting the Activity 'Treatment for Severe Infection' we can increase the capacity by using the replicate function found on the activity's 'Additional' tab.

If you increase this to 10 and run the simulation, this problem no longer occurs as we have sufficient capacity to deal with all infections. However, in reality it may be that we don't have the capacity to. To do this, reduce the replications to 5 and add the additional objects shown below to the process. Next, set the Shelf life of the queue to 3 and the routing in of 'Treatment for Severe Infection' to Expired Only. Set the Treatment for severe Infection to a dummy activity by making the activity timing a fixed 0 and ensure the end point has the remove from all states unchecked.

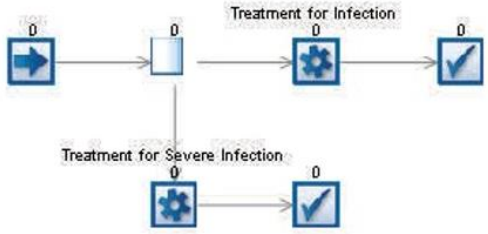


Figure 11 *The expanded discrete event simulation*

The aim of this process is to simulate what happens to patients if they cannot get a bed for treatment of an infection. By setting a shelf life this means that any patient that has to wait longer than 3 days will move down a different path. We now want the simulation to escalate the patients condition, essentially overriding the traditional state transitions, should patients not get a bed within 3 days. As states are controlled via Label values we can do this by using Label actions on 'Treatment for Severe Infection'. Set the Label Clinical Infection to 0 and the Label Severe Infection to 1.

Now, run the simulation again and we can see that as a result of lack of capacity we have an increased amount of Severe Infections. By modelling this it allows you to think about the effects of capacity on patient outcomes and how this may differ with treatment options.

As well as being able to trigger arrivals and control the State Charts, the time spent in activities can also be linked to the disease model. In the ulcer example patients will require a bed whenever they have a clinical or severe infection. The disease model controls when they will transition back to an uncomplicated state so in the discrete process we want to ensure that the patient will remain in a bed until they transition back to a healthy state.

Select the object Treatment for Infection and on the Ribbon select Additional, then Timing. Choose the bottom option and select Uncomplicated Wound from the list. This will now ensure that the patient will stay in the activity until they transition back to an uncomplicated state.

3.9 Costs

If you have connected your disease model to a discrete process you can use the traditional building blocks to apply costs to your simulation. On each tradition simulation object there is a costs option which is accessed by selecting the object, Properties and Finance. In the ulcer example we can apply a cost per day that would be incurred due to the patient being in a hospital bed.

Cost results are displayed in the income statement which is found on the home tab. This will give you a breakdown of where all the costs incurred in the model are. Like all results in SIMUL8 you can add these to the results summary by hovering over the result and right clicking. This is helpful when comparing runs.

Costs can also be applied using the method discussed in State Actions, where you can increment variables or spreadsheets when a patient enters or exits a state. If you require more complicated cost functions which require rules then you can also access Visual Logic. Each state has a number of events where you can code conditional statements in the Visual Logic editor. Select a state and on the properties tab on the ribbon chose Events, after selecting an event this will then open the Visual Logic editor.

4 CONCLUSION

Disease modeling is just one example where combining discrete event and agent simulation techniques adds value. This combination can also be applied in manufacturing, supply chain, service delivery, or anywhere where you have equipment or resources that degrades over time. With SIMUL8, you can rapidly build simulations that incorporate either individual techniques, or a combination of techniques to model and improve any type of business processes.

A APPENDICES

Table A-1 *Heal Time Data*

From State	To State
Week 1	0
Week 2	0
Week 3	0
Week 4	0
Week 5	0.04
Week 6	0.06
Week 7	0.12
Week 8	0.16
Week 9	0.23
Week 10	0.25
Week 11	0.29
Week 12	0.29
Week 13	0.41
Week 14	0.41
Week 15	0.41
Week 16	0.41
Week 17	0.43
Week 18	0.43
Week 19	0.55
Week 20	0.55
Week 21	0.62
Week 22	0.64
Week 23	0.69
Week 24	0.69
Week 25	0.72
Week 26	0.72
Week 27	0.72
Week 28	0.72
Week 29	0.73
Week 30	0.73
Week 31	0.74
Week 32	0.75
Week 33	0.75
Week 34	0.76
Week 35	0.76
Week 36	0.76
Week 37	0.76
Week 38	0.76
Week 39	0.77
Week 40	0.77
Week 41	0.77
Week 42	0.78
Week 43	0.79
Week 44	0.8
Week 45	0.81
Week 46	0.82
Week 47	0.83
Week 48	0.84
Week 49	0.84
Week 50	0.84
Week 51	0.84

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.010

USING FACILITATED SIMULATION TO EVALUATE INTEGRATED COMMUNITY-BASED HEALTH AND SOCIAL CARE SERVICES

Dr. Antuela A. Tako

School of Business and Economics
Loughborough University
Epinal Way, Loughborough LE11 3TU
a.takou@lboro.ac.uk

Professor Stewart Robinson

School of Business and Economics
Loughborough University
Epinal Way, Loughborough LE11 3TU
s.l.robinson@lboro.ac.uk

Dr. Anastasia Gogi

School of Business and Economics
Loughborough University
Epinal Way, Loughborough LE11 3TU
anagogis@gmail.com

Professor Zoe Radnor

Cass Business School,
City, University of London
Northampton Square, London, EC1V 0HB, UK
zoe.radnor@city.ac.uk

Ms. Cheryl Davenport

Leicestershire County Council
Championship Way, Glenfield, Leicester , LE3 8RA, UK
Cheryl.Davenport@leics.gov.uk

ABSTRACT

This paper introduces SIMTEGR8, which stands for "Simulation for Great Care". SIMTEGR8 is a new facilitated simulation approach, developed to evaluate the effectiveness of integrated community-based health and social care services. Working jointly with different stakeholder groups (service providers, commissioners, and service users), simulation models are developed and used as a catalyst for generating discussion about the effectiveness of the patient pathway and for identifying potential improvements to the services evaluated. Service users, a stakeholder group that can contribute to the knowledge generated in facilitated modelling sessions, have not been involved in facilitated simulation studies reported so far in the literature. The SIMTEGR8 approach is illustrated using a case study on the evaluation of the Lightbulb service, a housing support provider based in the Leicestershire area in the UK. The outcomes of the study and the challenges faced with involving stakeholders in simulation projects are discussed.

Keywords: Facilitation, discrete-event simulation, health care, SIMTEGR8 approach.

1 INTRODUCTION

The provision of integrated health and social care services is part of an NHS government policy enabled by the introduction of the Better Care Fund in 2015. This was designed to deliver new models of care ,

whereby local authorities and social care teams work with NHS partners in order to provide joined up and patient-centered services in the community (Charles et al 2018). The Better Care Fund provides a facility for a pooled budget between the clinical commissioning group and upper tier local authorities, it has a number of national conditions linked to integrated health and care and is measured nationally in terms of its impact on reducing emergency admissions, improving hospital discharge, reducing the number of people being permanently admitted to residential care and improving the number of people who are reabled and can maintain their health, wellbeing and independence at home (e.g. after illness, surgery or injury). Since 2014, across Leicester, Leicestershire and Rutland (LLR) health and care partners have been collaborating to transform health and care across this geographical footprint, in a partnership known as Better Care Together (Barber 2015), latterly called a Sustainability and Transformation partnership (STP), with the LLR STP being one of 44 such partnerships in England, led by the NHS.

The Better Care Fund are a key enabler to this wider partnership and provide a local plan and pooled budget to be applied in each part of Leicester, Leicestershire and Rutland (LLR). As part of this transformation programme, in 2015/6, partners in Leicestershire were testing several new models of care and community based services for frail and older people. The SIMTEGR8 project was set up in order to evaluate 8 of the services in the Leicestershire area, so that the results of the evaluation would be independently and systematically analysed with academic input, following which the results would inform future commissioning decisions for these services.

This paper introduces a new facilitated simulation approach, developed to evaluate the effectiveness of integrated community-based health and social care services in the Leicestershire area as part of the SIMTEGR8 project. SIMTEGR8 stands for "Simulation for Great Care" and was a research collaboration between Loughborough University, Leicestershire County Council (LCC), Healthwatch Leicestershire and SIMUL8 Corp. Computer simulation modelling, more specifically discrete-event simulation (DES), was used in facilitated workshops with groups of stakeholders in order to evaluate selected pilot services, their effectiveness in avoiding emergency admissions, and to assess ways in which the patient journey could be improved. The facilitation process and activities involved are explained. Our aim was to involve both service providers and service users in facilitated simulation workshops.

The contribution of this paper is twofold. First, we present a new facilitated DES approach that embeds the perspective of both service providers and service users in the simulation project lifecycle. To the best of our knowledge, this is the first research reported that engages both the client (commissioning body and service providers) and service users (patients and their family) in facilitated DES workshops. Secondly, it demonstrates the potential of using facilitated DES to support and evaluate the effectiveness of community-based health and social care services. Furthermore, we present a case study as an illustrative example that enables us to reflect on the advantages and limitations of involving different stakeholder groups in facilitated DES interventions.

The paper is structured as follows. Section 2 explores existing literature considering facilitated DES and stakeholder involvement in healthcare. Section 3 presents the SIMTEGR8 approach, followed by an illustrative case study in section 4, describing the context, the facilitation process, the models developed and the outcomes of using the approach to evaluate the Lightbulb (LB) service. A discussion follows considering the involvement of different stakeholder groups in simulation studies.

2 FACILITATED DES AND STAKEHOLDER INVOLVEMENT IN DES STUDIES

Research on facilitated discrete-event simulation (DES) is becoming popular, with a number of researchers reporting on building and using DES models in a facilitated mode of engagement with stakeholders. As opposed to the traditional analyst-oriented approach, in this mode the simulation analysts works jointly with the client to develop meaningful and relevant models, that can also support the stakeholder group in identifying feasible solutions (Robinson et al. 2014; Tako and Kotiadis 2015; Kotiadis and Tako 2018). The stakeholder group attends workshops, where the facilitator guides the process through a set of planned activities including: defining the problem, validating the model,

considering model findings and identifying possible solutions. A brief overview of facilitated DES studies follows.

Robinson et al. (2014) developed the SimLean approach that combines the use of simulation models and lean processes to support process improvement in healthcare. They use approximate models to understand the healthcare processes involved and to explore different solutions in facilitated workshops. The authors comment that client engagement enabled the acceptance and implementation of lean improvements identified by the study. Similarly, Baril et al. (2016) combine DES and lean principles to improve patient flows in an outpatient haematology-oncology clinic. Stakeholder involvement varies across the project, between individual and group facilitation to elicit information that informs the models which are created offline. At the end of the project, a Kaizen event was held using simulation-based games live with stakeholders, which informed subsequently the improvements implemented in the clinic.

Tako and Kotiadis 2015 developed PartiSim, a framework that supports the facilitation process in DES, consisting of six stages of which four are facilitated workshops with stakeholders. They also develop tools inspired from Soft Systems Methodology (Checkland 1999) to support the facilitation process and assembly of information in pre-model (Kotiadis et al. 2014) and post-model coding stages (Kotiadis et al. 2018). Proudlove et al. (2017) consider the technical aspect of making the model development phase more facilitated using the Business Process Model and Notation (BPMN) standard to enable stakeholder involvement. They build simulation models of two hospital settings. While the live development of models was possible for a simple model, this was not for more complex models. Further technological extensions to BPMN would be needed, to ensure that more complex models can be built jointly with stakeholders at workshops (Onggo et al. 2018).

While facilitated modelling offers a platform for involving stakeholders in simulation studies, existing studies do not explicitly include service users in facilitated DES process. There is currently an increasing interest internationally in involving patients and members of the public in health care research, recognizing the potential benefits that members of the public and service users have to offer in designing and improving health services (Pearson et al. 2013; Monks 2016). In the UK also, health and social care service providers are committed to involving service users and patients in the planning, development and evaluation of their services (Pearson et al. 2013). A similar expectation was also present when undertaking the research described in this paper.

Patient and public involvement (PPI) in healthcare modelling simulation is limited (Pearson et al. 2013). They identify a number of benefits from involving service users in the simulation study, including input into obtaining a better understanding of the context and of the objectives to be pursued, design and validation of models from the perspective of the patients and users of these services, as well as identification of acceptable and relevant to patients options for change. Alongside the benefits, Pearson et al. (2013) recognise the challenges faced when involving lay people and members of the public in technical modelling work such as simulation, which may inhibit modelers to engage more closely with such groups in their work. They discuss the lack of effective communication between researchers and patients to ensure there is shared understanding, primarily due to lack of a common language and knowledge between these groups. For example, patients have a different view of the service, limited to the part of the service they have experience of, which can affect their understanding of the models and technical terms used when considering the service as a whole. Another concern is related to the way patients and service users are identified and selected to participate, to ensure that bias is as much as possible avoided (Pearson et al. 2013). Considering that the user base of health and social care services are elderly and frail people, access and ability to participate is further impaired. Such difficulties were encountered also in the current study.

This paper presents a new facilitated simulation approach, the SIMTEGR8 approach, which combines two existing approaches, SimLean Facilitate (Robinson et al. 2014) and PartiSim (Tako and Kotiadis 2015), adapted specifically to fit the process carried out to evaluate integrated health and social care services as well as to ensure that participation of service providers and users is achieved. The approach

enables the triangulation of information between the modelling and the service provider team as well as a group of service users.

3 KEY STAGES OF THE SIMTEGR8 APPROACH

The approach consists of five main stages, of which three are facilitated workshops: project briefing, conceptual modelling workshop (W1), model development, service providers workshop (W2) and service users workshop (W3) (Figure 1). Each stage is next briefly explained.

Figure 1: Phases of the SIMTEGR8 approach.

1. **Initial Pathway Briefing.** This consists of a meeting with a smaller group of stakeholders, including members of the modelling and service provider team. The aim is to develop an initial understanding of the pathway, by discussing the aims and priorities of the service, workshop requirements, access to patient representation and data availability to inform the model.

2. **Conceptual Modelling Workshop.** The stakeholder group discuss the planned pathway and reflect on its efficiency. The discussion involves the following four phases:

    - Aims of evaluation. A brainstorming session to identify aspects of the service to be evaluated.
    - Process map. The modelling and stakeholder group work jointly to identify the main activities that take place in the real system and draw a process map of the service.
    - Pathway Effectiveness. A brainstorming session to identify performance measures used by the service. Service users' opinions about their experience are also considered.
    - Data Requirements. People responsible for providing the data required are identified, based on the process map developed.

3. **Model Development.** The qualitative conceptual diagram developed at the workshop is turned into a simulation model. The model is kept to a minimum to ensure stakeholder participation at the workshops. It shows the basic processes, the capacity and use of resources within the system.

4. **Service Providers Workshop.** This workshop uses the model to facilitate a discussion with members of the service provider team on how the service can be improved. The discussion involves the following four phases:

    - Model Understanding. The simulation model developed is presented and shown to the participants to allow them to understand how the simulation works;
    - Face Validation. The participants consider whether the simulation model reflects reality;
    - Problem Scoping. Based on model results, participants are asked to identify areas that affect pathway effectiveness.
    - Improvement. Participants identify changes that can be introduced to the service.

5. **Service Users Workshop**. The model with improved visual representation is used to help facilitate a discussion with patients and carers. The discussion involves the following three phases:

- Model understanding. The pathway and model are explained to the participants and the simulation run showing a patient moving around the system.
- Problem Scoping. Issues that have been revealed by running the model and the participants' own experiences and concerns about the service are discussed.
- Improvement. The discussion turns to discussing how the service could be improved.

4 CASE STUDY: EVALUATION OF THE LIGHTBULB SERVICE

4.1 The Lighbtbulb (LB) Service

Lightbulb (LB) is one of the services evaluated using the SIMTEGR8 approach. It is a housing support service helping elderly and frail people in the community to stay safe and longer in their homes by preventing accidents and falls and keeping them away from hospital. It provides a wide variety of housing support and advice, including minor home alterations, such as hand rail or major home adaptations such as installing a downstairs bathroom or stair lifts. A pilot service was available in some localities within Leicestershire at the time that the project was undertaken (in 2016). Next the key phases of the project and milestones are briefly explained.

4.2 Project Briefing

The modelling team met with the service managers and a patient voice agency representative. It was established that the aim of the LB service evaluation was to support the business case being developed at the time, which involved the design of a new and faster process to deliver services to patients. It quickly became clear that the focus was to test that the new process had been modelled accurately in the business case and that it could deliver the expected time scales and throughput. The existing detailed process map used for the business model was shared with the modeling team. The performance manager in charge of the business case was our main point of contact regarding data requirements for the model.

Stakeholder involvement and the workshops plan was also discussed in this meeting. Analysis of the different roles and staff involved in the LB service took place. As a result the group came up with a list of staff that would be invited to attend the workshops (conceptual modelling and project leads workshop), representing a variety of roles, to ensure that a good representation of the different aspects of the service was achieved. It was also agreed that service users involvement would be organized by one of the partners in the project, Healthwatch, a locally-based independent organization, representing the patients' voice on aspects related to health and social care. They would oversee the process of communicating with and inviting service users to attend Workshop 3. This also ensured that we were able to adhere to data protection rules and patient confidentiality.

4.3 Conceptual Modelling Workshop (Workshop 1)

The first workshop was attended by seven key staff from LB, the modeller, note taker (project manager) and facilitated by the first author. The workshop was held in a dedicated meeting room. The session was managed within a tight timeframe of 2 ½ hours to ensure it impacts minimally on service delivery. The activities that took place are next described.

We started with the aim of the evaluation. It was agreed that the main aims of the evaluation were to: 1) evaluate the utilization of the staff (occupational therapist - OT, housing support coordinator - HSC, and technical officer - TO) involved in the delivery of the service and distribution of tasks between them; 2) validate the overall expected times scales in providing services; and 3) consider the impact of an increase in demand for services. A significant part of the workshop was spent on drawing the process map. Workshop participants were invited to contribute activities that take place in the service based on their perceptions of the process on a large white paper stuck on a wall. After a few iterations, an agreed process map was produced. This was transferred into a tidied up version on the Visio software after the workshop (Figure 2).

The participants were next invited to identify the metrics used to evaluate the performance of the service. These included: total completion time for minor alterations and major adaptation cases; staff utilisation levels (in %) and the total number of cases completed by service type and staff type (throughput). The effectiveness of the pathway was then discussed from the project leads' and service users' perspective. This focused on the time taken for services to be provided both from patients' and service providers' point of view. While feedback received from the patients who had used the service during the pilot phase was generally positive, some delays in the time taken to complete the work to be done had been noted. The new redesigned pathway aimed to resolve this. It was deemed that the relevant information needed to proceed with building the model was acquired, so the workshop drew to a close.

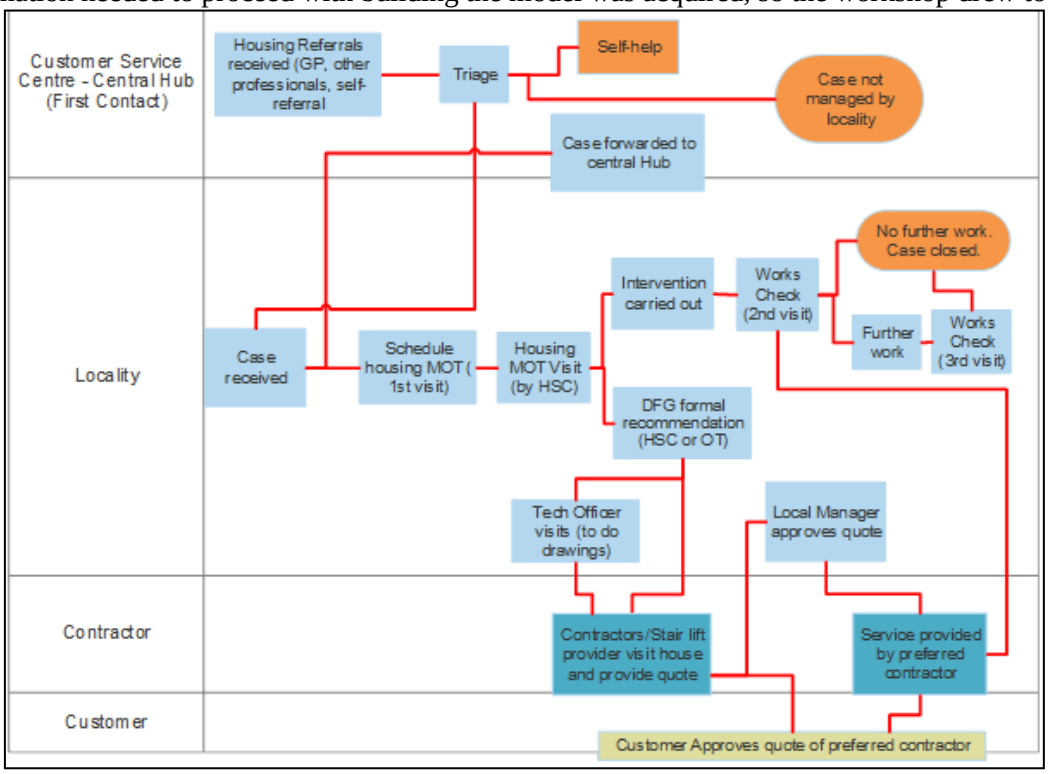


Figure 2: Process map of the Lightbulb service.

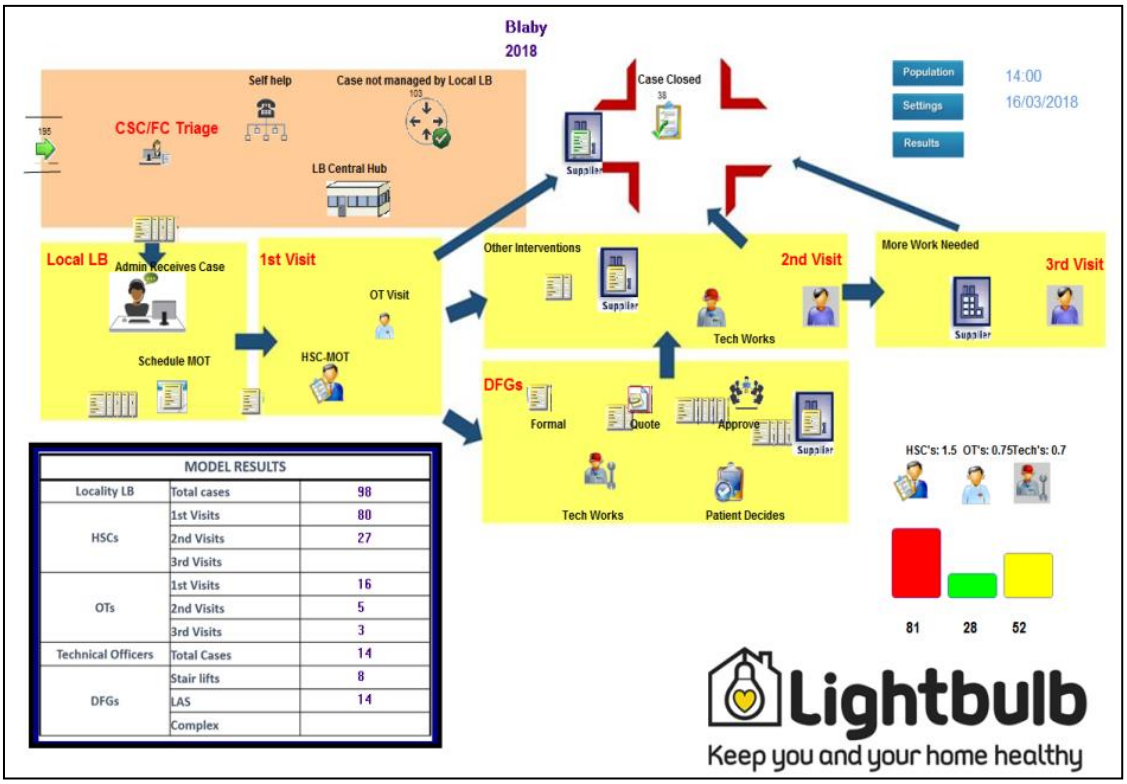
4.4 Model Building

After the workshop, the conceptual model developed was converted into a simulation model (Figure 3) that imitates the planned flow of services and user cases through the service. The model represents each locality separately as well as the overall Leicestershire service. It shows what each service would look like based on current levels of demand and projected staffing levels in the new redesigned service pathway. The model outputs include staff utilization for the three types of staff involved in providing services (Housing support coordinator, occupational therapist and technical officers) and the number of cases completed (throughput) by type of service and resource. These were visually displayed in the model (Figure 3) so that participants would be able to validate the model and its outputs at the next workshop.

4.5 Service Providers Workshop (Workshop 2)

This workshop was held in a dedicated meeting room. There were five members of staff from LB, including the service manager. Two of the participants had not attended the first workshop. The modeler, facilitator and the project manager who also was the note taker, were present. The sequence of activities that took place is next briefly described.

The modeller explained the model to the participants, including the outputs that it captured (e.g. number of 1st Visits by HSCs, number of Stairlift cases, TOs total number of cases etc.) as well as the



MODEL RESULTS		
Locality LB	Total cases	98
HSCs	1st Visits	80
	2nd Visits	27
	3rd Visits	
OTs	1st Visits	16
	2nd Visits	5
	3rd Visits	3
Technical Officers	Total Cases	14
DFGs	Stair lifts	8
	LAS	14
	Complex	

Figure 3: The Lightbulb simulation model.

assumptions made (e.g the model did not show interaction with other services). Then the simulation was run through and the participants were allowed time to absorb the model. One specific locality as chosen by the participants, was selected to run the model.

A large part of the workshop was spent on validating the model. The facilitator asked the participants to confirm whether the model reflected their understanding of the process. This wasn't intended to be a detailed validation to assess statistical accuracy, but instead for the participants to gain trust in the model, that it was performing as expected. A variety of opinions existed within the room and as a result heated discussions took place regarding the data used. This was expected as the model is based on the business case rather than on an established service. It was observed that the model showed that the time taken to complete some of the complex services and major adaptations was longer than it was anticipated. It was agreed that the model would be amended to reflect service times based district council data.

The model was next used to evaluate the service and understand the service metrics provided. Based on the insights gained from the model it was identified that there was a high reliance on HSCs, who were working close to 80% capacity. Reliance on HSC resulted also in longer case completion and customer waiting times, which were higher than what the service had planned for in the business case. On the other hand, OTs and TOs were under-utilized, ranging between 29% – 60% across the different localities, so it was clear that a further look at the distribution of work in the model was needed.

Reflecting on the model results, the participant group was next encouraged to identify changes that can be introduced to the service. In light of the disproportionate staff utilization levels, it was suggested that HSCs could not work every case through to the end. While there was no time at the workshop to look at this in detail, it was agreed that this would be looked at after the workshop.

Post-workshop, the service revisited work flows and division of work amongst staff. Later stages of disabled facilities grants (complex services) were assigned to OTs and TOs to complete instead of HSCs. It was also suggested that MOT Visits are completed by an OT when all HSCs are busy. The model was modified to reflect a re-distribution of the work between the staff. The updated model showed a significant reduction in overall case completion and customer waiting times, achieving a reduction between 19% and 38% in overall time in the system for disabled grant facilities cases. In addition, more realistic staff workload levels (60%) were achieved for all staff.

4.6 The Service Users Workshop (Workshop 3)

This workshop was held at a local District Council's offices. The modelling team included a facilitator, the modeler the patient voice agency (Healthwatch Leicestershire) and two staff members from the district LB team (service provider). The workshop was attended by nine service users, of which two were carers. For reasons that were out of the researchers' control, the service users that attended the workshops had only used the LB service for minor adaptions. Hence, the workshop focused primarily on this part of the service.

The workshop started with a brief explanation of the LB simulation model to the participants. Before the workshop, the model had been further developed into a "User Mode" model. This is a simplified version of the simulation model built for the service provides workshop, converted into a more user-friendly format with improved graphics for the service users. The model shows only one patient moving at a time until his/her journey within the service is completed and the simulation stops. This enables us to isolate different types of patients and monitor their journey within the service on a patient by patient basis. To start another patient's journey, users can press the "play" button again. Furthermore, while the model is running, explanatory text appears on screen following the progress of a patient's case through the service. The participants could watch the patient moving through the parts of the service in the model, while one LB team member explained how this related to the service they had received in real life.

Next, the participants were asked to discuss their experiences with the service and compare it to the model previously presented. This was an indirect way of assessing whether the participants understood the pathway and that it made sense to them, without restricting the discussion to the case presented. The participants commented about the good quality of work delivered by the service and the quick process, the support provided to them, highlighting how it enhanced their quality of life and helped them to be more independent. They also discussed the multi-disciplinary nature of the service, that different services are coordinated by one point of contact, which is considered important as it creates familiarity. Despite the positives, with encouragement by the facilitator, the participants commented about the fact that the service was not well-known and that there is not enough clarity of the services on offer and how they can be accessed. One of the participants reported that the handyperson assigned was not able to complete the job at one visit and several visits were required by the HSC to ensure that the work was completed.

As the users were generally happy with their experience of completion times, there were no improvements identified that related directly to the patient pathway presented. With encouragement by the facilitator, participants identified a number of possible improvements that the service could benefit from. These were mostly related to improving access to the service, which is expected as the service had been operating on a pilot basis. The service has taken these suggestions on board and better signposting is now provided on the LCC website, with further plans to increase visibility of the service in the community.

5 REFLECTIONS AND CONCLUSIONS

This paper presents the SIMTEGR8 approach through an example where the LB service, a housing support service based in Leicestershire in the UK, is evaluated. Using a computer simulation of a patient pathway in order to stimulate discussion and to identify ways to improve the service, with members of the

service provider and users team, was effective. The discussion that took place in the three workshops was lively with many contributors and engagement with the models was high.

The conceptual modelling and service provider workshops achieved a mutual understanding of the service among the participants, this in turn informed the model developed, which was then tested and used to identify further improvements in the service, respectively. Members of the service provider team participated in these workshops. Similarly, the service users workshop successfully achieved the aims originally set out. The participants demonstrated a shared understanding of the pathway, despite having had experience of only a small part of the service. There was some engagement with the simulation model. The presence of the service staff members at the workshop, helped achieve a common level of communication at the workshop as the participants were familiar with those staff members and a positive rapport had been already established. Involving a service user group into a workshop was a great achievement as they face accessibility issues due to their medical condition.

The SIMTEGR8 approach presented in this paper advances the existing practice of facilitated simulation, by developing a new facilitation process that combines the inputs of the modelling team, with that of a group of service providers and users, in using simulation models to inform service improvements. The approach adapts pre-existing facilitated simulation approaches, SimLean (Robinson et al. 2014) and PartiSim (Tako and Kotiadis 2015; Kotiadis and Tako 2018). Achieving a good level of participation of the different stakeholder groups in the facilitated sessions was a challenge. This leads us to consider whether a different sequence of activities could work better and/or whether further improvements to the 'User Mode' model layout could improve engagement of service users with the model and ultimately their understanding of how the service works on the whole. Furthermore, as we experienced difficulties in establishing PPI, further software advancements in technology could support a better engagement with patients and vulnerable groups. Similarly an equivalent participation of the members of the service team in the first two workshops could have avoided some of the disagreements that were aired in the 2nd workshop.

Engagement with service users brought a complementary perspective to the evaluation. It helped the researchers and the evaluation project overall to reach more meaningful conclusions. In this particular case, the participants confirmed that the resulting patient waiting times were acceptable to them. This indirectly confirmed the planned changes regarding the division of tasks among staff that emerged at the end of workshop 2 with the service providers. The service users' input in the case study presented was mostly to confirm our understanding of the context and aims of the evaluation. They did not make a direct input into the model or the data used in it as suggested in Pearson et al. (2013). They however identified complementary suggestions for further improvements to the service, which were not previously obvious to the service provider and modeling team. Service users' involvement in SIMTEGR8 does not necessarily aim to improve the model developed, but to use the model as a vehicle to generate discussion and insights about the service metrics and to identify potential improvements that are acceptable and satisficing.

ACKNOWLEDGMENTS

We acknowledge the support of the partners in the SIMTEGR8 project, including Leicestershire County Council (LCC), Healthwatch Leicestershire and SIMUL8 Corp. Special thanks go to the Health & Care Integration Manager at LCC, the managers and members of the Lightbulb Service, including Blaby District Council, the service users and their carers for attending the workshops detailed in this paper.

REFERENCES

Barber, M. 2015. "Outcomes Roadmap: Anticipated Changes to Health and Social Care Services across Leicester, Leicestershire & Rutland in Support of the Five Year Strategy" (Version 1, First Issue). Better Care Together. http://www.bettercareleicester.nhs.uk/EasysiteWeb/getresource.axd?AssetID=34600, accessed 2nd September 2019.

Baril, C., V. Gascon, J. Miller, and N. Côté. 2016. "Use of a Discrete-event Simulation in a Kaizen Event: A Case Study in Healthcare". *European Journal of Operational Research* 249(1):327-339.

Charles, A., C. Ham, B. Baird, H. Alderwick, and L. Bennett. 2018. "Reimagining Community Services: Making the Most of Our Assets". https://www.kingsfund.org.uk/sites/default/files/2018-01/Reimagining_community_services_report.pdf, accessed 2nd September 2019.

Checkland, P. B. 1999. *Systems Thinking Systems Practice*. Chichester, UK: Wiley.

Kotiadis, K. and A.A. Tako. 2018. "Facilitated Post-model Coding in Discrete Event Simulation (DES): A Case Study in Healthcare", *European Journal of Operational Research*, 266:3, 1120-1133.

Kotiadis, K., A.A. Tako and, C. Vasilakis. 2014. "A Participative and Facilitative Conceptual Modelling Framework for Discrete Event Simulation Studies in Healthcare", *Journal of the Operational Research Society*, 65(2): 197-213.

Monks, T. 2016. "Operational Research as Implementation Science: Definitions, Challenges and Research Priorities", *Implementation Science*, 11:81.

Onggo B.S.S., N.C. Proudlove, D'Ambrogio, S. A. Calabrese, B. Stefania, and N. L. Ghiron. 2018. "A BPMN Extension to Support Discrete-event Simulation for Healthcare Applications: An Explicit Representation of Queues, Attributes and Data-driven Decision Points", *Journal of the Operational Research Society*, 69:5, 788-802.

Pearson, M., T. Monks, A. Gibson, S. Brailsford, and K. Stein. 2013. "Involving Patients and the Public in Healthcare Operational Research-The Challenges and Opportunities", *Operations Research for Health Care*, 2:86-89.

Proudlove N.C., S. Bisogno, B.S.S. Onggo, A. Calabrese, and N.L. Ghiron. 2017. "Towards Fully-facilitated Discrete-event Simulation Modelling: Addressing the Model Coding Stage", *European Journal of Operational Research*, 263, (2):583-595.

Robinson, S., C.Worthington, N. Burgess, and Z.J. Radnor. 2014. "Facilitated Modelling with Discrete-event Simulation: Reality or Myth?", *European Journal of Operational Research*, 234 (1): 231–240.

Tako A.A. and K. Kotiadis. 2015. "PartiSim: A Framework for Participative Simulation Modelling", *European Journal of Operational Research*, 244 (2): 555–564. DOI: 10.1016/j.ejor.2015.01.046.

AUTHOR BIOGRAPHIES

ANTUELA A. TAKO is a Reader in Operational Research at the School of Business and Economics, Loughborough University. She holds a PhD in Simulation and an MSc in Management Science and Operational Research from the University of Warwick. Her research interests include comparison of simulation approaches, facilitated and participative simulation modelling, conceptual modelling and health care modeling. Antuela is an Associate Editor of the Journal of the Operational Research Society and Journal of Simulation.

STEWART ROBINSON is Professor of Management Science and Dean at the School of Business and Economics, Loughborough University, UK. He holds an honors degree in Management Science (Operational Research) (1985) and a PhD in Management Science (1998), both from Lancaster University. His research interests are in the practices and use of simulation models where he focuses on conceptual modelling, model validation, output analysis and alternative simulation methods (discrete-event, system dynamics and agent based). His home page is www.stewartrobinson.co.uk.

ANASTASIA GOGI was the Research Associate in the SIMTEGR8 project. Anastasia holds a BSc in Industrial Management (University of Piraeus, Greece), an MSc in Management Sciences (University of Southampton, UK) and a PhD (Loughborough University, UK). Her research intersts include behavioural OR and simulation.

ZOE RADNOR is Professor of Service Operations Management at Cass Business School and VP of Strategy and Planning for City, University of London. Zoe holds a PhD in Management Science from the University of Manchester, UK, an MSc in MSc in Computer Integrated Manufacturing, Cranfield University, UK and a BEng (Hons) in Production Engineering and Production Management, University of Nottingham, UK. Zoe's research interest is in performance and process improvement and service management within public sector organisations. She is a Fellow of the British Academy of Management (FBAM) and the Academy of Social Science (FAcSS).

CHERYL DAVENPORT is Director of Health and Care Integration at the Leicestershire County Council. Cheryl has 27 years' experience of working in health and care in the UK.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.011

OPPORTUNITIES FOR USING AGENT-BASED SOCIAL SIMULATION AND FUZZY LOGIC TO IMPROVE THE UNDERSTANDING OF DIGITAL MENTAL HEALTHCARE SCENARIOS

Owen Barnes

University of Nottingham
School of Computer Science
Nottingham, NG8 1BB
owenebarnes@gmail.com

Dr Peer-Olaf Siebers

University of Nottingham
School of Computer Science
Nottingham, NG8 1BB
peer-olaf.siebers@nottingham.ac.uk

ABSTRACT

Agent-Based Social Simulation (ABSS) has seen success in studying emerging behaviour in social contexts. In this paper, we explore the use of ABSS to improve the understanding of digital mental healthcare scenarios. Our exploratory study focuses on understanding how different age generations within a population react to the implementation of digital mental health therapies. Our model uses a complex social media network where agents can communicate and alter their opinions over time. In this paper we also present some technical novelty. We introduce interval type-2 fuzzy logic as an option for modelling human decision-making and accounting for the uncertainty that exists when modelling complex systems. Fuzzy Logic is a concept particularly well suited to formalise and deal with imprecise concepts. The work presented here seeks to encourage Social Simulation modellers to experiment more with interval type-2 fuzzy logic.

Keywords: Agent-Based Modelling, Social Simulation, Fuzzy Logic, Digital Mental Health, Trust

1 INTRODUCTION

With the recent emergence of data-driven technologies, the popularity of digital mental healthcare has risen significantly (Arigo et al. 2019). With this comes the opportunity to use simulation to further our understanding of how digital mental healthcare may affect the population when implemented. Since this area is recent to the past decade, there is little insight into how we can advance our abilities to model such technologies, especially when considering the social and ethical implications of the technology. In this paper we consider the following possible future digital mental healthcare scenario: Therapy sessions are inspired by chatbot therapy, which are one-to-one therapies taking place between digital chatbot therapists and patients (Health Europa 2019). After weekly sessions, the trust value and the psychological condition of patients change, and trust levels can also be altered through access to social media. Over time, some trends should be observable that indicate the impact of healthcare quality and social communication on the trust of the population as a whole, and on the trust of specific groups within the population in particular. Our study considers different generations of people (from Millennials to Baby Boomers), and explores how different groups of people responded to different scenarios, for example when the quality of the provided therapy is poor.

The goal of this paper is twofold. On the one hand, we are interested to investigate how Agent-Based Social Simulation (ABSS) can be used in the context of studying digital mental healthcare. For this purpose, we use the Engineering ABSS (EABSS) framework by Siebers and Klügl (2017). This framework supports the conceptualisation of ABSS models, using the concept of co-creation, as well as software engineering tools and methods. It drives the model development process and documents the outcome of this process. It is used to provide transparency to viewers and to improve the ability for multi-disciplinary stakeholders to understand the model without technical jargon. On the other hand, we want to look at novel ways of representing decision-making processes in agents. For this we

test whether interval type-2 fuzzy logic could provide modellers with a realistic decision-making representation when modelling these types of scenarios. Interval type-2 fuzzy logic benefits from being able to handle uncertainty with promising results, but has seen little application within ABSS.

In the remainder of this paper we first provide background information on ABSS, fuzzy logic and its use in ABSS, and digital mental healthcare (Section 2). We then present a detailed description of our conceptual model for simulating digital mental healthcare scenarios, using the EABSS framework as a documentation tool (Section 3). Next, we briefly describe the tools we used for the implementation of our conceptual model (Section 4). We then present our experimental results and findings with regards to our established hypotheses, and comment on interesting insights (Section 5). Finally, we provide a summary of achievements, limitations, and how to overcome these limitations in the future (Section 6).

2 BACKGROUND

2.1 Agent-Based Modelling and Social Simulation

In Agent-Based Modelling (ABM) a system is modelled as a collection of autonomous decision-making entities (agents) where each agent individually assesses its situation and makes decisions on the basis of a set of rules (Bonabeau 2002). Individual agents interact with each other and their environment to produce complex collective behaviour patterns at system level. Agents are designed to mimic the behaviour of their real-world counterparts; they are capable of making autonomous decisions and showing proactive behaviour. Agent-Based Simulation (ABS) is a powerful simulation paradigm that can be used for conducting what-if analysis of human centric systems (Siebers and Aickelin 2008). By developing models of complex social systems and studying their evolution through simulated time, researchers have an artificial lab where they can observe the interactions between social agents and processes and their consequences. Such artificial labs can be used for small-scale exploratory studies as well as large-scale decision support applications. ABS is a bottom-up approach and is used in situations for which individual variability between the agents cannot be neglected. It allows understanding how the dynamics of many real systems arise from traits of individuals and their environment. It allows modelling a heterogeneous population where each agent might have personal motivations and incentives, and to represent groups and group interactions. Social Simulation (SS) studies socio-economic phenomena by investigating the social macrostructures and observable regularities generated by the behaviour and relationships between individual social agents, and the environment in which they act. This is useful for policy decision support in many scenarios, including transport, housing, education, or healthcare. ABSS is a variation of ABS and SS, which looks at modelling social behaviour using agent technologies; it is commonly described as a multidisciplinary intersection between agent-based computing, social sciences, and computer simulation (Davidsson 2002).

2.2 Fuzzy Logic in the Context of Social Simulation

Fuzzy Logic is defined by Zadeh (1988) as the logic underlying approximate, rather than exact modes of reasoning, considering "degrees of truth" rather than the usual "true or false" Boolean logic. It is a concept particularly well suited to formalise and deal with imprecise concepts (Izquierdo et al 2015). When using fuzzy logic in simulation contexts, the most popular use of fuzzy logic is by using inference systems. Fuzzy inference systems are able to provide modellers with an ability to model the subjective uncertainties which arise in ABSS, such as when modelling social relationships and exchanges between agents (Raoufi and Rayek 2015). By providing the system with inputs, the system can produce an outcome based on the modelled scenario. Type-1 fuzzy inference systems model uncertainty with the use of membership functions which correspond to the value an input may have to a specific set. Interval type-2 fuzzy systems go a step further by also making the membership functions a fuzzy set. In a type-1 membership function, the set modelled must be precisely defined by the modeller, while in interval type-2 membership functions, the modeller can be vague in their definition. The modeller defines an upper and lower membership function which forms an area called the footprint of uncertainty. Within the footprint of uncertainty, an embedded set is produced which is

the equivalent of a type-1 set. This is useful when trying to model subjective concepts such as emotion, which is not the same for every agent. When modelling these concepts, the larger area of the footprint of uncertainty means more uncertainty is represented in the definition of the membership function. For a more in-depth coverage of this topic please refer to Mendel et al. (2006). The applications of type-1 fuzzy logic to ABSS are extensive, but we have only found one application of interval type-2 fuzzy logic to ABSS (Castañón-Puga et al. 2014) and none documented with regards to healthcare.

2.3 Digital Mental Health and Healthcare in the Context of Social Simulation

The rise of big data within the world means that quantitatively driven solutions can be applied to a range of problems. Examples of these applications include devices which track your health statistics by counting the amount of steps you make in a day (Kerner and Goodyear 2017), using motion tracking to analyse your sleep performance (Hamida et al. 2015), or recently, offering a way for users to communicate their emotions and feelings to act as a digital therapeutic bot (Fitzpatrick et al. 2017). Since popularity has grown, there have naturally been efforts to implement this sort of digital technology to more specialised areas of healthcare. This consists of developing AI in ways which can help provide alternative methods of treating mental health issues successfully such as depression, anxiety, and dementia (Fernandez-Sotos et al. 2019). We found very few papers during our literature review which explored digital mental health scenarios, and it was found that those papers were focused on understanding the social and ethical implications of digital mental health rather than looking at the service provision aspects. As far as we could see, the simulation of digital therapies appears novel and untouched by the simulation modelling community.

3 CONCEPTUAL MODEL

3.1 Approach

The EABSS framework provides the core of a methodology which supports ABSS model development and documentation in a structured way. Full details about the framework and its application can be found in Siebers and Klügl (2017). The EABSS framework is grounded in the concept of co-creation (Mitleton-Kelly 2003) and ideas from software engineering (Sommerville 2015), but it can also be used by individuals. In the latter case the individual needs to consider the perspective of stakeholders (i.e. slip into their roles) during each process step. The framework consists of an Analysis and a Design part, as depicted in Figure 1. For capturing different types of information, it uses the Unified Modelling Language (UML) notation (Fowler 2004) extensively.

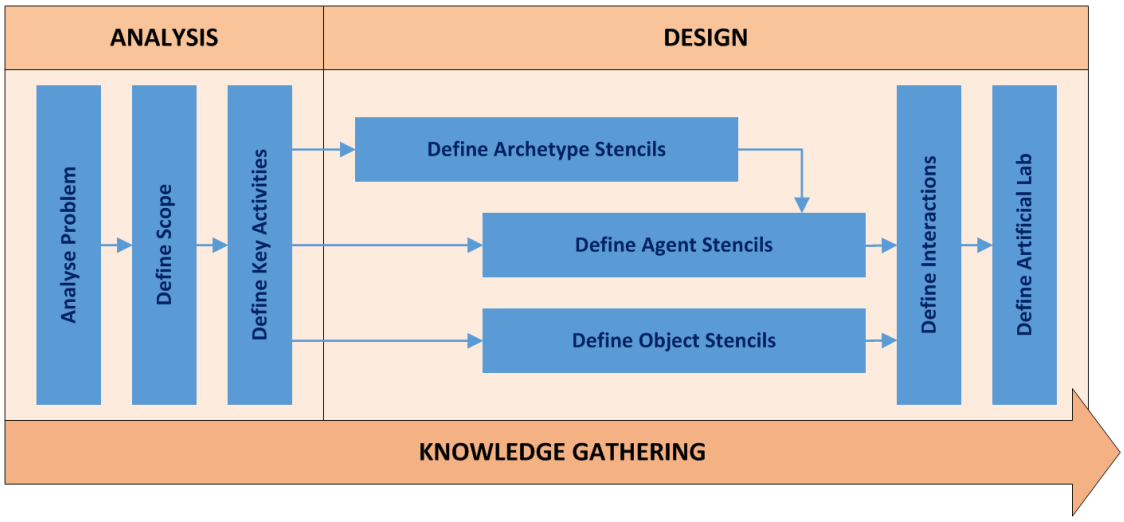


Figure 1 A *high-level overview of the EABSS (after Siebers and Klügl 2017)*

3.2 Analysis

Analysing the Problem

Experience with digital therapy is still in its infancy. The aim of our study is to better understand how people react towards digital therapy, with people modelled based on generation groups: Millennials, Generation-X, and Baby Boomers. In this paper we focus on testing two hypotheses: (1) "The more access to social interaction agents have, the more sensitive the simulation would be when certain parameters (e.g. trust) are altered" and (2) "Competence of digital therapies (i.e. our 'Session' in the simulation) will play a large role in defining the output trust of the population".

Experimental factors are model parameters that can be altered to test our hypotheses. For our purpose the experimental factors consist of (1) the proportion of different generations of agents, (2) the initial level of trust of the population tested, (3) the competence of digital therapies, and (4) the level of access agents have to the social network. By altering these model parameters, it is possible to investigate the effect they have on the dynamics of the system over time.

Responses are the outputs from the model which enable us to test our hypotheses. ABSS allows to collect micro and macro-level data in form of time series and averages over time. The trust variable was the primary response for testing our hypotheses; at the macro-level this captured the aggregated trust of the population, and at the micro-level it looked at individual agent's trust level evolution during the simulation.

Scope

This stage involves defining the level of abstraction appropriate to test the hypotheses, identifying relevant actors, relevant elements of the physical environment, as well as social and psychological aspects that might be relevant for the modelling. The agreed scope is presented in Table 1.

Table 1 *Resulting scope table*

Category	Element	Decision	Justification
Actor	Person (Patient)	Include	Subject which is being observed
	Session	Include	A virtual 'therapist' providing services to patients
	Administrator	Include	Manages waiting lists, organises and creates sessions on demand
Physical Environment	Hospital	Exclude	No benefits of representing these elements within the environment as physical entities
	Home	Exclude	
	Therapy	Exclude	
	Spatial node	Include	Used to measure distance between agents and defining the location of agents during the simulation
Social and Psych. Aspects	Person social influence	Include	Used to help determine the outcomes of social interactions
	Person condition	Include	Tracks the outcome of session treatment, and is what makes people admit themselves to care
	Person trust	Include	Our primary response (and therefore fundamental to the model)
	Person age	Include	Important for determining social media usage
	Desire	Exclude	Psychological theory was initially considered, but later not used due to time constraints
	Volatility	Exclude	
	Arousal	Exclude	
	Stereotype	Exclude	Captured by person age
Other	Social network	Include	Agents can interact with each other, in a realistic manner through using friendLink networks to send messages
	Therapy system	Include	System boundary used by people who require digital therapy

Defining Key Activities

This stage involves linking actors to use cases (which represent the key activities). Figure 2 defines how actors can interact. Bubbles represent relevant use cases. Associations between actor and use case indicate which actor is involved in which use case. Relationships between two use cases specify common functionality and simplify use case flows.

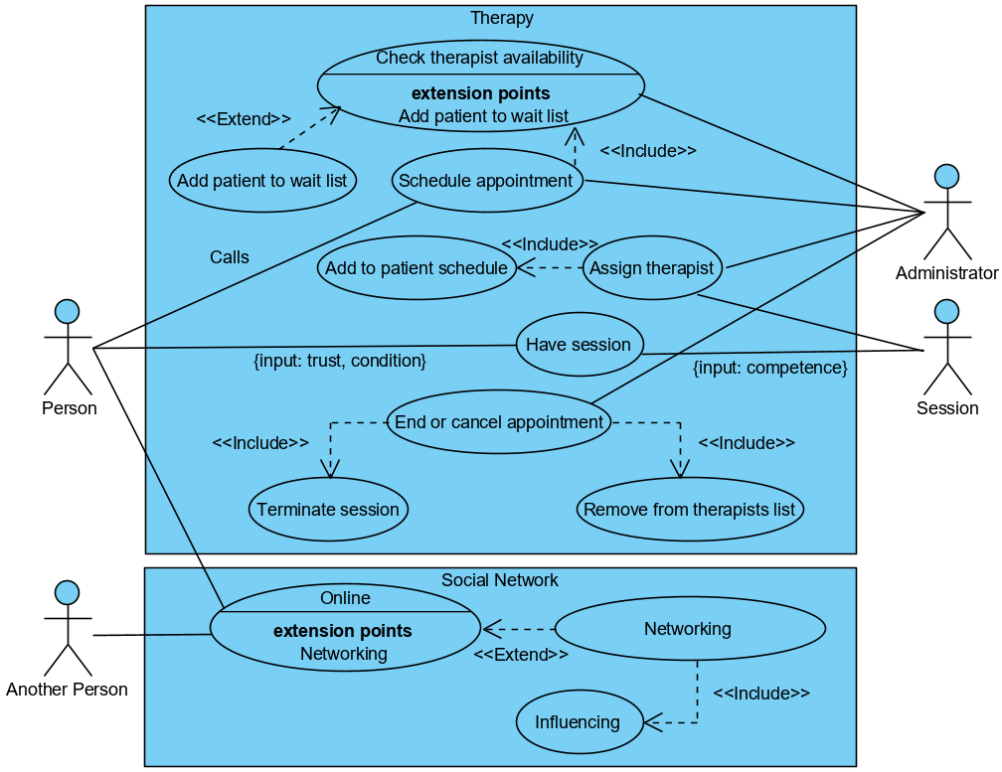


Figure 2 *Resulting use case diagram*

3.3 Design

Archetypes

This stage involves defining archetypes, based on what we found in the literature. We distinguish three generations, based on age range. From the eMarketer (2019) stats we know how many people within a specific age range use social media, and from Globalwebindex (2018) we know roughly how much time people spend online each day, using social media. A summary of the data can be found in Table 2. We use this information as a stochastic element of the model that resembles the likelihood of an agent using social media at any point during the day.

Table 2 *Resulting archetype definitions*

Name	Age range	Social media users [% of age range members]	Use of social media [hours per day]
Millennials	1980-1994	90.4	2.38
GenerationX	1965-1979	77.5	1.49
Baby Boomers	1944-1964	48.2	1.12

Agents

This stage involves defining the states that entities can be in and the dependencies between these states in form of transitions. UML state machine diagrams (statecharts) are used to present this information. Figure 3 contains the statecharts for all agents identified as actors in Table 1. There are two statecharts for capturing a Person agent's states. The first tracks the possible condition of a person. People in the 'stableCondition' state (holding a *condition* value of 50+) do not require digital therapy. If their condition value drops below 50, they will arrange a digital therapy session (via self-admission) with the administrator. The second statechart controls a Person agent's social media usage, where the rate at which a person goes online is defined by their archetype. The Admin and Session agents are secondary agents that provide services to the Person agent. When Person agents get into an unstable condition, they request a session by contacting the Admin agent. The Admin agent then

creates a session in form of an abstract agent that represents a virtual therapist providing digital therapy sessions.

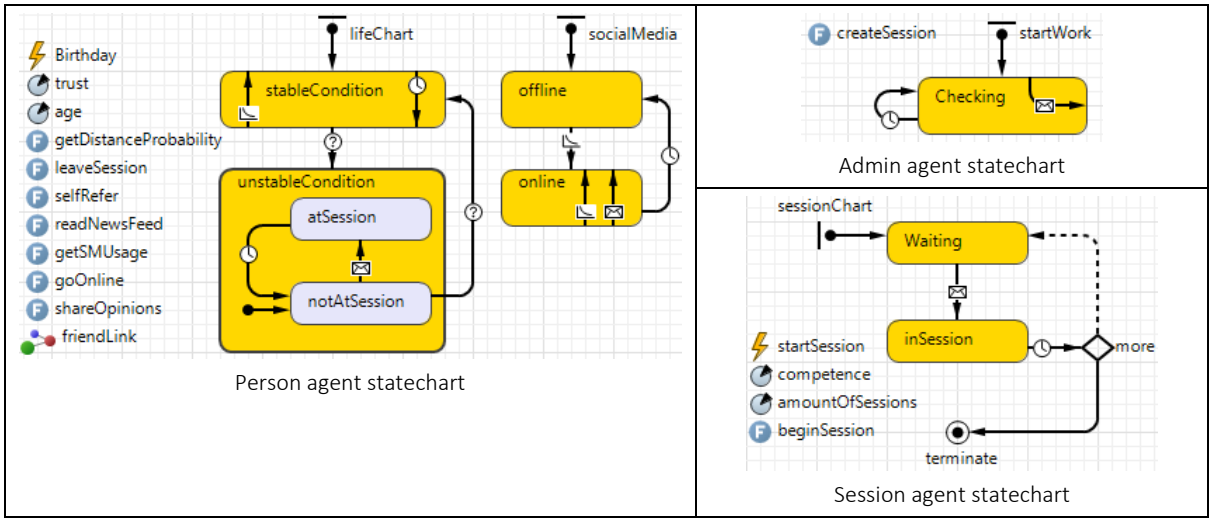


Figure 3 *Statecharts of the actors identified in the scope table*

Objects

The only objects we consider in relation to the environment for our model are spatial nodes. They allow us to measure the distance between Person agents, and to graphically represent their location and communication links within a defined space.

Interactions

The interaction between different agent types is described in Section 3.3, Subsection "Agents". Here we focus on the interaction (online communication and its effects in relation to opinion dynamics) between Person agents (friend links). We model the social media network using an algorithm that connects agents depending on their proximity and social similarity. This algorithm produces a probability that is applied once agents go online, providing realistic communication and influencing effects. More information about the algorithm can be found in Goldenberg & Levy (2009).

Artificial Lab

The artificial lab is responsible for providing methods to collect aggregated outputs of the model or provide global methods that can be accessed by all agents. Notable elements of our artificial lab are the "social feed", which emulates a social media newsfeed, and some debugging methods to ensure input parameters to fuzzy inference systems are within bounds.

4 IMPLEMENTATION

For the implementation of the conceptual model described in Section 3, we used AnyLogic 8.5 (https://www.anylogic.com/). This is a multi-paradigm simulation IDE that supports (amongst others) the ABM paradigm. It is (relatively) easy to use, yet not restrictive, as it includes a high-level graphical modelling language and allows users to extend the model with custom low-level Java code.

For the implementation of opinion dynamics (covering opinions of agents when communicating across the social media channel) and outcomes of digital therapy sessions, we used fuzzy logic, to see how it accounts for the uncertainty that exists within such systems. We embedded the fuzzy logic as follows. The first fuzzy logic system represents how somebody (p1) online may react to another person's (p2) opinion of digital therapies, which is modelled with an attribute 'Trust'. For agent p1 the system takes three inputs, the trust of agents p1 and p2, and the 'social influence' variable of p1. The system returns a value between [-0.05, 0.05] which is added to p1's trust variable. The second fuzzy logic system provides an output to reflect the reaction of a person to the digital therapy, adjusting p1's 'Trust' and 'Condition' attribute. The system takes three inputs, involving the competence of the session agent, p1's Trust, and p1's Condition. If the person has a positive experience, their values

increase, which is then reflected in future communication on social media. This produces a 'word-of-mouth' social network that offers realistic communication within the model. For fuzzy logic support within AnyLogic, we embedded the Juzzy library (Wagner 2013). It is a Java based toolkit for type-1, interval type-2 and general type-2 fuzzy logic and fuzzy logic systems. The library provides modellers with a clear and well thought through syntax which can be implemented into AnyLogic by using an external class file. It offers a user-friendly interface for less technical stakeholders, to define the fuzzy logic elements.

Our model is available for download at https://www.comses.net/. Our model has been verified with the help of visual debugging support and several external modellers. In order to ensure results accurately reflect the mean performance of our stochastic simulation model, we use the confidence interval method (α=5) to find the required number of replications. The test indicated that four replications of every iteration were required when collecting results. We also conducted a comparison test between fuzzy and crisp decision-making using a method proposed by Vu et al. (2013). The details of our comparison are omitted here due to space constraints but can be found in Barnes (2019). In general, we found that results with the fuzzy implementation corresponds better to what we expect to see in the real world, effectively modelling the hesitancy expected within the first few months of deploying the digital therapy to the population.

5 EXPERIMENTATION

We tested two hypotheses as defined in the conceptual model, one dealing with "social interactions" and one dealing with "competence". In this section we define on the hypotheses we are testing, provide the experimental setups and results of our experiments and discuss the findings.

5.1 Testing Hypothesis 1: Social Interactions

Here we explore the effect that age plays in trust dynamics and how sensitive the output of the model is for each generation (from Millennials to Baby Boomers) when we altered the initial trust level of agents. ***Hypothesis***: "The more access to social interaction agents have, the more sensitive the simulation would be when certain parameters (e.g. trust) being altered". ***Experimental setup***: In order to test this hypothesis, we run the simulation for seven iterations for each generation, changing the initial trust value from 47 to 53. We then collected our data, and tracked statistics such as the variability between runs, to understand whether we should accept or reject our hypothesis. ***Results***: We found that Millennial and Generation-X runs (who experienced more exposure to the social network) proved sensitive when the initial trust level was altered. Both showed high variance between runs, where the variance between Millennial runs was 113.15, and for Generation-X 111.25; significantly lower than when testing our older generation at 4.827. The high variance shown on these runs tells us that more access to social media is more likely to affect the final trust value of the population at the end of the simulation. Furthermore, we identified a macro-level trend which represented the hesitancy of the population during early stages of the simulation. From this experiment we learnt that slight changes to initial conditions can play a huge role in the success of implementing digital therapies; and this information could prove useful to stakeholders. Figure 4 shows us that early stages of the simulation are crucial in determining the outcome trust levels of the population, and few situations arise where population established trust once the average trust is below 47. We also found that agents tended to benefit more from using digital therapies when the trust level of the population was higher, because they possessed higher input values which provided a better outcome after every session. We *accepted our hypothesis* due to the high variance between runs, which increased as agents had more access to the social network.

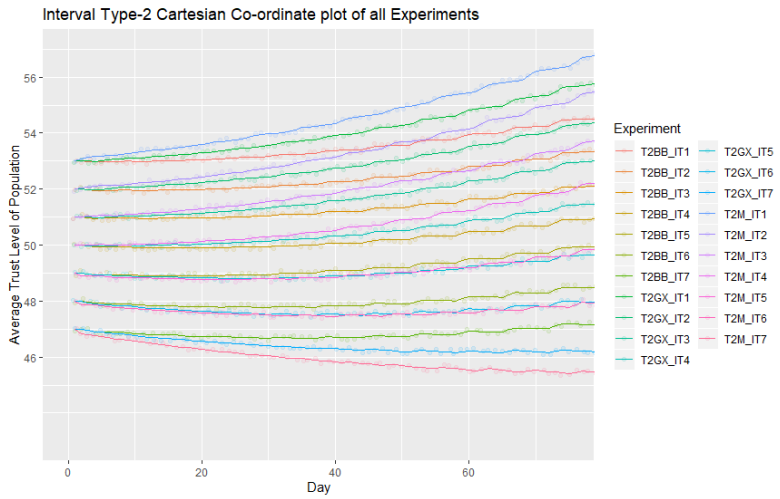


Figure 4 *A Cartesian time series plot showing the initial hesitancy of our population across all runs*

5.2 Testing Hypothesis 2: Competence

Here we test how the simulation reacts when we alter the competence of sessions, which controls the performance of chatbot therapies in the simulation. The output aggregated trust level of the population is expected to alter a lot. We check whether any extremist behaviour emerges, e.g. if the change in competence of our digital therapies cause a significant difference in the overall trust level. ***Hypothesis***: "Competence of digital therapies (i.e. our 'Session' in the simulation) will play a large role in defining the output trust of the population". ***Experimental setup***: To test this hypothesis, we ran scenarios that saw the competence variable of sessions increment from 20 to 80, in steps of 5. We tested this to explore whether the variable played a statistically significant impact in the output trust levels of the population, which would lead us to accept our hypothesis. We should expect to see our variance increasing as competence is increased, and for trust to increase consummate to the increase in competence of digital therapies. ***Results***: We found that there was very little difference across all runs, even when comparing extreme scenarios where digital therapies had poor competence, against when they were highly skilled. We can see from Figure 5 that the variance is inconsistent with Generation-X runs.

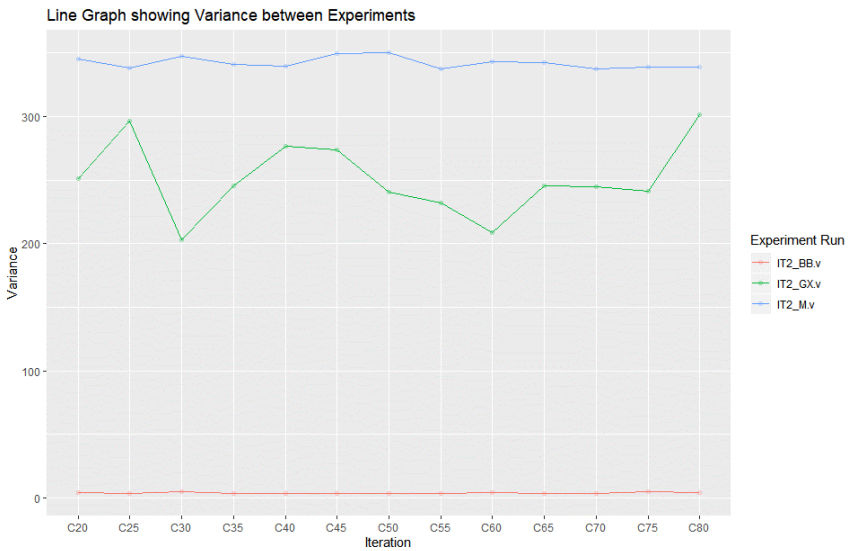


Figure 5 *Variance of output trust levels of all runs*

Further exploration into this tells us that the attitude of agents, presented in Figure 6, changed the most between day 400 to 600 of the simulation. Figure 6 also shows that there is a wide spread of results when looking at trust level over time, but that all runs establish high trust levels, regardless of

the competence of digital therapies. This was unexpected; we expected to see a lower trust level when the competence of digital therapies is low. We *rejected our hypothesis*, since the output trust level of the population stayed consistently high regardless of whether our digital therapies were more skilled.

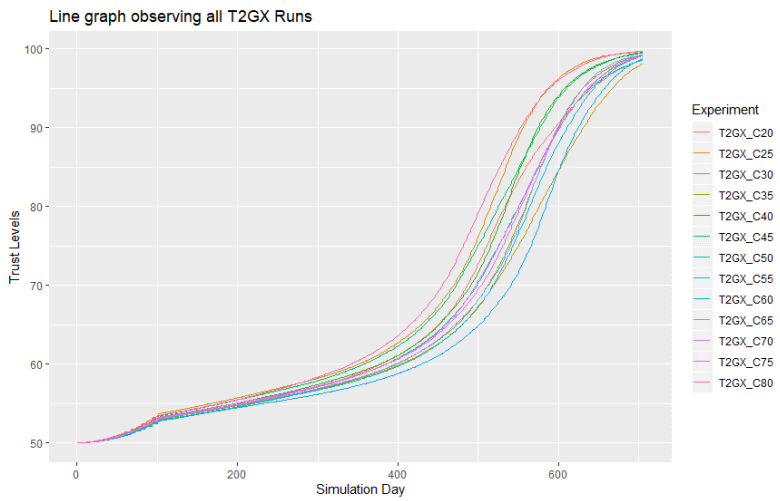


Figure 6 *Time series plot of all simulation runs with agents belonging to Generation-X*

6 CONCLUSIONS

In this paper we have created an exploratory model which introduces a way of simulating the reaction of a diverse population to different digital mental healthcare scenarios. We also showed how interval type-2 fuzzy logic can be successfully implemented to provide decision-making capabilities of agents in uncertain environments. We found patterns in both, micro and macro-level data, which provide useful insight into agents' reactions to digital therapies. This knowledge can be useful for stakeholders when considering the social implications of using digital technologies. The model was unable to be validated and was therefore classified as exploratory. This was perhaps the biggest limitation. Since this is a novel application, the extent to which one can validate the model is limited, and there were no datasets which could be used to compare with the output of the model. In the future we aim to validate the base model, which could be done by organising a workshop with relevant experts and stakeholders.

In conclusion, our project has shown the potential of simulation to investigate scenarios of digital mental health dynamics and the impact appropriate healthcare can have. We hope that in the future more resources are made available to continue the work on this novel research topic.

REFERENCES

Arigo D, Jake-Schoffman D E, Wolin K, Beckjord E, Hekler E B and Pagoto S L (2019). The history and future of digital health in the field of behavioral medicine. *Journal of Behavioral Medicine* **42(1)**: 67–83.

Barnes O (2019) Exploring the effects of using interval type-2 fuzzy logic with social simulation to help improve the understanding of digital mental health. *Dissertation*. University of Nottingham.

Bonabeau E (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*. **99**: Suppl 3.

Castañón-Puga M, Flores-Parra J M, Gaxiola-Pacheco C, Palafox-Maestre L E, Ahumada-Tello E (2014) Extending NetLogo with interval type-2 fuzzy inference systems. In: Second World Conference on Complex Systems (WCCS) 10-13 Nov, Agadir, Morocco.

Davidsson P (2002). Agent based social simulation: A computer science view. *Journal of Artificial Societies and Social Simulation*, 5(1):7.

eMarketer (2019) US Social Media Users, by Generation, https://www.emarketer.com/chart/226029/us-social-media-users-by-generation-2019-of-population [Last accessed 01/02/2020]

Fernandez-Sotos P, Fernandez-Caballero A, Gonzalez P, Aparicio A I, Martnez-Gras I, Torio I, Dompablo M, Garca-Fernandez L, Santos J L and Rodriguez-Jimenez R (2019) Digital

technology for internet access by patients with early-stage schizophrenia in Spain: Multicenter research study. *Journal of Medical Internet Research* **21(4)**.

Fitzpatrick KK, Darcy A and Vierhile M (2017) Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* **4(2)**.

Fowler M (2003) UML Distilled: A Brief Guide to the Standard Object Modeling Language (Object Technology Series). 3e. Addison-Wesley Professional.

Globalwebindex (2018) Social media across generations (Preview) https://www.globalwebindex.com/reports/social-media-across-generations [Last accessed 01/02/2020]

Goldenberg J and Levy M (2009) Distance is not dead: Social interaction and geographical distance in the Internet era. arXiv:0906.3202.

Hamida, S T B , Ahmed B, Cvetkovic D, Jovanov E, Kennedy G and Penzel T (2015). A new era in sleep monitoring: The application of mobile technologies in insomnia diagnosis. In: Mobile Health. Springer, Cham.

Health Europa. (2019) New chatbot therapist to help combat depression unveiled, https://www.healtheuropa.eu/chatbot-therapist-to-combat-depression-un/93609/ [Last accessed 01/02/2020]

Izquierdo LR, Olaru D, Izquierdo SS, Purchase S, Soutar GN (2015) Fuzzy logic for social simulation using NetLogo. *Journal of Artificial Societies and Social Simulation* **18(4):1**.

Kerner C and Goodyear V A (2017) The motivational impact of wearable healthy lifestyle technologies: A self-determination perspective on FitBits with adolescents. *American Journal of Health Education* **48(5)**.

Mendel J M, John R I and Liu F (2006) Interval type-2 fuzzy logic systems made simple. *IEEE Transactions on Fuzzy Systems* **14(6)**: 808-821.

Mitleton-Kelly E (2003) Ten principles of complexity & enabling infrastructures. In: Mitleton-Kelly E (ed) Complex Systems and Evolutionary Perspectives of Organisations: The Application of Complexity Theory to Organisations. Elsevier.

Raoufi M and Fayek A (2015). Integrating fuzzy logic and agent-based modeling for assessing construction crew behavior. In: Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) 2015 held jointly with 5th World Conference on Soft Computing (WConSC), 17-19 Aug, Redmond, Washington, USA.

Siebers P O, Aickelin U, Celia H and Clegg C W (2008) An agent-based simulation of in-store customer experiences. In: Proceedings of the OR Society Simulation Workshop (SW08), 1-2 Apr, Worcestershire, UK.

Siebers P O and Klügl F (2017) What software engineering has to offer to agent-based social simulation. In: Edmonds B and Meyer R (eds) Simulating Social Complexity - Understanding Complex Systems. Springer

Sommerville I (2015) Software Engineering. Pearson.

Vu T, Siebers P O and Wagner C (2013) Comparison of crisp systems and fuzzy systems in agent-based simulation: A case study of soccer penalties. In: Proceedings of the 2013 UK Workshop on Computational Intelligence (UKCI 2013), 9-11 Sep, Guildford, UK.

Wagner C (2013) Juzzy - A Java based toolkit for type-2 fuzzy logic. 2013 IEEE Symposium Series on Computational Intelligence, 16-19 Apr, Singapore.

Zadeh L A (1988) Fuzzy logic. *Computer* **21(4)**: 83-93

AUTHOR BIOGRAPHIES

OWEN BARNES received a MSc (with distinction) in Computer Science with Artificial Intelligence from the School of Computer Science, University of Nottingham, UK. This paper is based on his MSc dissertation.

PEER-OLAF SIEBERS is an Assistant Professor at the School of Computer Science, University of Nottingham, UK. His main research interest is the methodological advance of Computer Simulation as it is used for studying human-centric and coupled human-natural systems. He is a strong advocate of "object oriented" Agent-Based Modelling. For more see http://www.cs.nott.ac.uk/~pszps/

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.012

ANALYZING CLOSER CARE STRATEGIES FOR ELDERLY PATIENTS: EXPERIENCE AND REFLECTIONS FROM MODELING WITH SYSTEM DYNAMICS

Dr. Gary Linnéusson

School of Engineering
Jönköping University
P.O.Box 1026, SE-551 11, Jönköping, Sweden
gary.linneusson@ju.se

Dr. Ainhoa Goienetxea Uriarte

School of Engineering Science
University of Skövde
P.O.Box 408, SE-541 28, Skövde, Sweden
ainhoa.goienetxea@his.se

ABSTRACT

The expected demographic changes, and especially the rise in life expectancy, will considerably increase elderly patients' demand for healthcare. There are different strategies that can offer better care for these patients, reduce their unnecessary visits to the emergency departments, and in consequence, reduce the number of hospitalizations and days at the hospital. This study employed system dynamics to analyze the economic and quality-related effects of different closer care strategies such as investments in care coordinators and mobile health clinics, as well as to offer proactive care in the primary care facilities for elderly patients. The results indicate that a combination of the different strategies will support better care for patients, will reduce hospital costs and will reduce the existing pressure on the emergency department. The paper also reflects on the process followed to conduct the study and the lessons learned.

Keywords: System Dynamics, Elderly Care, Simulation, Healthcare

1 INTRODUCTION

There is an increasing pressure on healthcare policymakers to design systems that will be sustainable in the future (Lyons and Duggan, 2015). This is a challenging task, taking into account the expected increase of the aging population, e.g., the actual world population aged sixty years or over will be doubled by 2050 (United Nations, 2017). Additionally, the rates of chronical diseases and multimorbidity are expected to increase in elderly patients (Lindgren, 2016), which in consequence will increase the demand on the healthcare system and the economic pressure on healthcare providers.

According to Lyons and Duggan (2015), there are different factors that characterize the healthcare infrastructure: 1) exogenous factors associated with population dynamics (demographics, lifestyle, etc.); and 2) internal decision variables associated with policy measurements as well as the development of the healthcare services to respond to the existing demand by the exogenous factors. This paper focuses on the analysis of the second factor and tries to analyze the impact of establishing policies and to develop the healthcare services to offer a closer and better care for elderly patients (65 years or older), and at the same time, to minimize the care they require from the emergency departments (ED) and the subsequent hospitalizations and days staying at the hospital. Different authors have defended this approach stating that to offer timely and effective primary care (PC) can even reduce hospitalizations, and thereby, avoidable complications during hospitalizations for these type of patients (Boyd et al 2008).

There are different operational research methods and tools that can be employed to support healthcare policymakers to make better decisions, some of them are reviewed in Hulshof et al. (2012). Simulation is a popular technique, and different studies employing simulation to support healthcare system design and improvement have been reviewed by different authors (Brailsford et al 2009; Katsaliaki and Mustafee, 2011; Mielczarek and Uziałko-Mydlikowska, 2012; Salleh et al 2017). When the problem under study has a dynamic nature and there is a need to understand the

interconnections between the different parts involved in the system, System Dynamics (SD) is an appropriate tool to be employed (Linnéusson et al 2018). As discussed by Senge and Sterman (1992), it enables multiple testing with the objective to question own mental models, and at the same time, it questions the underlying values governing the system. Different authors have studied the dynamic complexity of restructuring the healthcare systems via SD, such as in Homer and Hirsch (2006), as well as in the examples provided in the reviews by Kunc et al (2018) and Chang et al (2017).

This paper presents the results of a case study using SD with the aim of analyzing the effects of the development of closer care strategies for elderly patients including multimorbidity patients and frequent attenders (FA). To the best of the authors' knowledge, this approach has still not been used to analyze the dynamics of this problem. Besides the traditional case study report, this paper also presents a reflection which analyses the process and learnings from the multiple trial and error search for a problem focus, subsequent problem behavior to be modeled, its consequences to the model building process, and how the results supported decision-making.

The article is structured as follows: Section 2 presents the background on how and why the project was done; Section 3 describes the method and steps applied to conduct the project; Section 4 elaborates on the details about the qualitative SD model; Section 5 briefly presents the quantitative SD model and overall simulation results; Section 6 includes a reflection and lessons learned during the project development; finally, Section 7 reveals the conclusions and future work.

2 BACKGROUND

As an important step towards quality in care, the region of Västra Götaland (VGR) in Sweden is working on an initiative to offer closer care to patients. This includes four areas of action: 1) to develop the organization to offer closer care; 2) to concentrate the offered care to achieve better quality and availability; 3) to develop digital care services; and 4) to focus on quality-driven improvements. Some of the main motivations to work with closer care actions are related to increase the quality of care provided, as well as to decrease existing waiting times, queues, and rising costs for hospitals (Taylor and Dangerfield, 2005). The lack of coordination and availability, as well as a reactive and non-person-centered focus which usually has characterized PC, has influenced the behavior of patients that prefer to go to the ED, sometimes unnecessarily.

Elderly patients are a specific group of patients who have continuous care need and contribute with a considerable amount of visits to the ED and hospitalizations (LaCalle and Rabin, 2010). According to data from 2016, elderly people (65 or older) in VGR were around 320.000. Of these, around 14% were patients with multimorbidity, and around 1,56% of them were FA in the ED, which means that they visited the ED at least four times in one year (the most common definition of FA, according to LaCalle and Rabin, 2010). The details of these patients are shown in Table 1.

Table 1 *Data about elderly patients' non-planned hospital services usage.*

Parameter	Multimorbidity frequent attender patients	Multimorbidity patients – non frequent attenders	Non-multimorbidity patients
Total persons in the region	5.000	41.000	274.000
Average amount of ED visits per year	5,5	0,9	0,3
Average amount of times the patient is hospitalized per year	3,3	0,7	0,1
Average amount of days per hospitalization	6	10	6
% of patients with avoidable hospitalization	15%		2%
% of patients returning to the ED and being hospitalized again after 1-30 days	5%		1%

Even if the elderly are a relatively small group of patients, the number of visits to ED, hospitalization rates and length of stay at the hospital are considerably high in comparison to other groups. These variables have been identified as one of the major causes of ED overcrowding (Moskop et al 2009). Therefore, analyzing how to offer closer care for this group of patients was prioritized.

Multimorbidity FA patients within this group have the highest number of visits to the ED, hospitalization rates, as well as the highest avoidable hospitalization rates. The data show that many of these patients were hospitalized unnecessarily and that a considerable percentage of them were coming back to the ED in the period of a month.

Different actions to offer closer care performed in specific municipalities of the region reported very good results in terms of quality of care provided and cost savings. For example, to have care coordinators in charge of coordinating the care offered to elderly patients arriving to the ED together with the community-based services and PC, as well as to introduce a survey about the status of the elderly patients, helped reduce the amount of time the patients were waiting in the ED, reduced the number of patients being hospitalized, and reduced even the number of patients returning back to the ED within a month. Another tested action, which proved to reduce the number of visits to the ED, is to have mobile health clinics including a team of a doctor and nurses who visit elderly patients in their homes or home care facilities. These patients are usually very unstable or need palliative care. The mobile health clinics try to offer proactively high-continuity and person-centered care, as opposed to the commonly reactive care offered both in the ED and PC. A third action that proved to reduce the number of visits to the ED is to proactively work with the elderly in the PC facilities. Meaning that PC visits are pre-booked systematically and more time is assigned to doctors and nurses to meet and treat these patients. In a specific region of Sweden, this proved to reduce 20% of the visits of this group of patients to the ED reducing, in consequence, the number of hospitalizations.

Therefore, the objectives pursued in this project were: 1) to analyze the dynamics of elderly patients' care-seeking behavior; 2) to analyze how different actions to offer closer care can impact the number of visits in the ED and hospitalizations; and 3) to analyze the effect in the costs.

3 WHICH STEPS DID WE TAKE?

The steps followed to conduct the project were the ones presented in Figure 1.

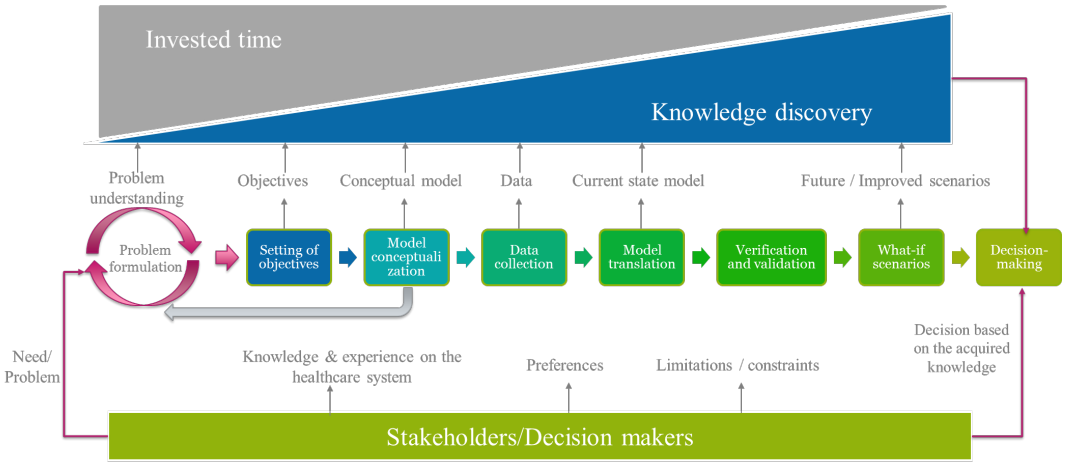


Figure 1 *The process followed to conduct the study. Updated from Goienetxea Uriarte et al (2017).*

The project started with an open problem formulation about how to define closer care for all the patients in the region. To better understand this concept a workshop was organized with the stakeholders and subject matter experts from the PC, ED, and the hospital. Additionally, many iterations were performed in group model building sessions to facilitate discussions about what does it mean to offer closer care, who needs it, what is the dynamic behavior of patients seeking care, etc. These discussions eventually helped to focus the study on elderly patients visiting the ED. After deciding the problem to be addressed, the modelers together with the stakeholders developed the conceptual model using causal loop diagramming (CLD) described in section 4. In the next step, a more concrete focus was decided where specific closer care strategies for elderly patients were chosen to be analyzed via SD, these included: 1) to have care coordinators in all the ED of the region; 2) to have enough mobile care clinics to be able to offer the service to all in need; and 3) to offer a proactive care in the PC facilities. Subsequently, data was gathered and the current state simulation model was developed, as described in section 5. The model was verified and validated together with the stakeholders, based mainly on historical data and face validation (Sargent, 2011). Then different scenarios related to the three closer care strategies explained previously were developed, compared,

and combined to analyze their impact in reducing the visits to the ED, number of hospitalizations and days at the hospital, as well as costs. The decision makers were then provided with the results which served as an input to decision-making regarding how to conduct the investments for better closer care in the region.

The initial stages of the process, especially the ones related to problem formulation, the setting of objectives and model conceptualization took a considerable amount of time from the project budget, although it was on the last stages of the project where most of the knowledge was generated.

4 ANALYZING THE DYNAMICS OF ELDERLY PATIENTS IN THE HEALTHCARE SYSTEM USING CAUSAL LOOP DIAGRAMMING

The closer care concept was considered a general fix to the healthcare system and to support reducing the care-seeking behavior at EDs in specific. However, in order to comprehend how the pressure on EDs could be released, a CLD was developed with the purpose to identify the potential feedback mechanisms involved in creating this pressure. The problem structuring process explored several potential problems eventually resulting in the CLD presented in Figure 2, which established a creative dialogue and a deeper understanding of the problem dynamics. It was the initial articulation of the central feedback loop of R1 which finally released the dialogue. The model includes several identified short- and long-term feedback of events and actions that together create the dynamic forces of central importance to better understand the complexity and inherent challenges of improving the quality of care and its related costs. The core of the CLD is composed of two central feedback loops:

- R1 – *GoTo ED 65+*: the main feedback resulting in elderly patients' care-seeking behavior, which creates a sustaining pressure on EDs.
- B1 – *Proactive Care 65+*: the main feedback which could potentially limit or enable the development of elderlies' current care-seeking behavior to the ED.

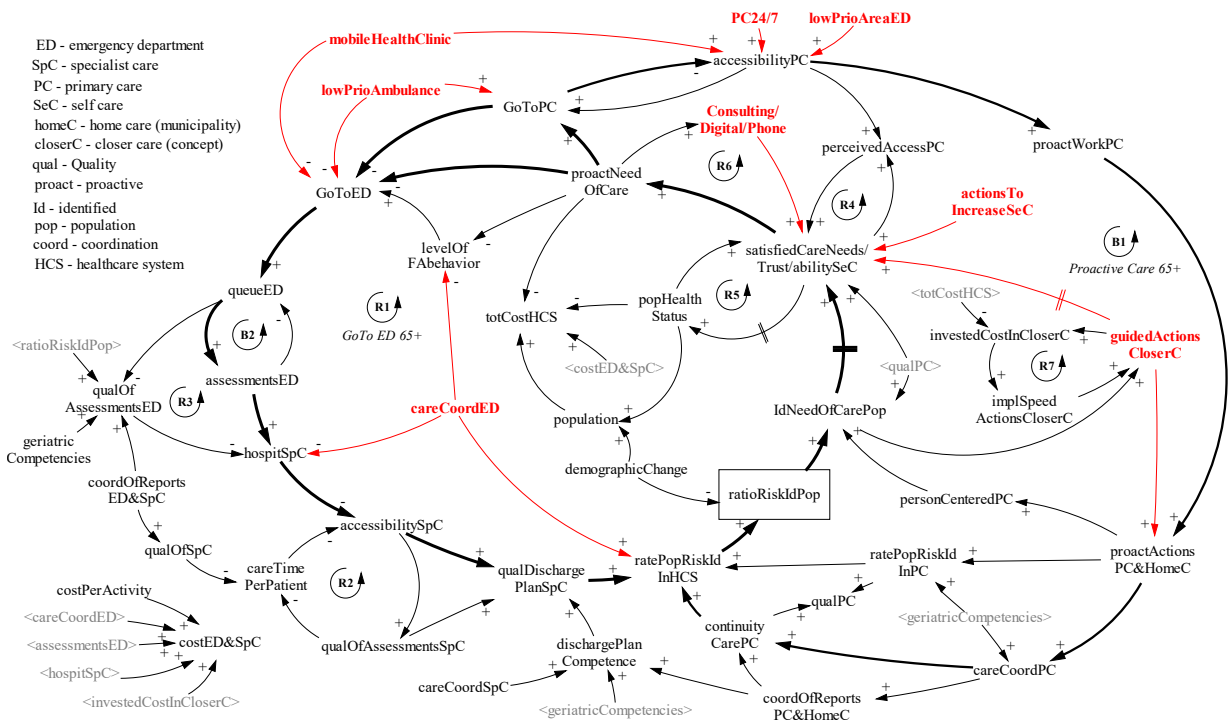


Figure 2 *The main interactions between the healthcare system and elderlies use of care, using Vensim software.*

The variables employed in the CLD are represented by abbreviations described in the legend in Figure 2. In the description below, a subset of these variables and their feedback loops are elaborated using *Italics*.

The R1-loop considers how to reduce the current care-seeking behavior to the ED (*GoToED*). R1 is a reinforcing feedback loop that can potentially lead to a vicious cycle towards a worse-and-worse

condition, or a virtuous cycle towards a better-and-better condition. Hence, if *GoToED* could be reduced, a chain-reaction of positive effects could be expected such as 1) *queueED* would be reduced, 2) lesser assessments at the ED (*assessmentsED*) would be needed, and 3) fewer people would need hospitalization and specialist care (*hospitSpC*), which would increase the accessibility of hospitals (*accessibilitySpC*) and potentially improve the quality of the discharge planning (*qualDischargePlanSpC*) leading to better care. However, the opposite is also considered to hold true, i.e., lower accessibility affects time invested in discharge planning leading to a cutting-the-corner-behavior (Repenning and Sterman, 2001).

Another key aspect brought up at the very first workshop is that in order to provide more efficient and quality care there is a dependency on the current levels of the patients' trust in the healthcare system (R1-loop). In succeeding discussions it was considered that trust was closely connected to satisfied patient care needs, to increase their health literacy, as well as their self-care. Hence, the variable with the name *satisfiedCareNeeds/Trust/abilitySeC*, is a multidimensional variable with the outcome of changing the conditions for proactiveness in the healthcare system. It is also a variable strengthened and weakened by several other feedback loops, where the desired improved outcome is to redirect people from the ED to more proactive ways to seek care. One fundamental aspect that supports raising the levels of *satisfiedCareNeeds/Trust/abilitySeC* is a parameter called risk identification. It is a term that represents when a person becomes a registered patient in the healthcare system. Meaning, that a person could have repeated interaction with the healthcare system without being risk identified, which could be the cause for becoming an FA (Kivelä et al 2018). This could be caused by the system itself through lacking quality in ED, PC, and the hospitals' discharge planning described above. Hence, actions to increase the rates of risk identifying patients in the healthcare system (*ratePopRiskIdInHCS*) are crucial to increase the number of people being identified (*ratioRiskIdPop*) and to reach its tipping point towards a more proactive state (wherever that might be). Moreover, *ratioRiskIdPop* is a level diluted by the continuous rates of *demographic change,* through deaths and new elderly people. Hence, the stock of risk identified people in the system and its connected flows are also important in order to attain the satisfied patient care needs for the population overall (*IdNeedOfCarePop*).

Another part of the healthcare system with the mission to provide proactive care found in the B1-loop is the work carried out by the PC as well as by the municipalities through home care services. However, the level of proactive work (*proactWorkPC*) depends on the applied resources and resulting accessibility to PC (*accessibilityPC*), leading to more proactive actions in PC and home care (*proactActionsPC&HomeC*). Proactive care includes three main actions: 1) higher levels of coordinated work in PC (*careCoordPC*) leading not only to better reporting between PC and home care (*coordOfReportsPC&HomeC*) and supporting discharge planning at the hospitals, but also to improved continuity care in PC (*continuityCarePC*) which leads to higher precision in the risk identification of patient needs and improved quality of PC; 2) higher rates of patients being risk identified by the PC (*ratePopRiskIdInPC*), increasing the rate of overall risk identification in the healthcare system leading to improved quality where a higher quality of PC (*qualPC*) supports improving *IdNeedOfCarePop* and directly supports the level of *satisfiedCareNeeds/Trust/abilitySeC*; and 3) improved person-centered care (*personCenteredPC*), which directly supports higher quality care for those who are risk identified through more or less continuous monitoring.

Hence, the B1-loop identifies that proactiveness in PC is dependent on that resources are available and utilized according to the above-mentioned actions. Here, as an example, a changed balance of people who are seeking care, from *GoToED* to *GoToPC,* due to a productive implementation of higher quality actions in ED and hospitals can be restricted by a limiting performance of the B1-loop. Accordingly, improvements are needed in both the R1- and B1-loops to achieve sustainable effects on peoples' care-seeking behavior (*proactNeedOfCare*).

R1 and B1 are not the only loops in the diagram. Worth mentioning is the R5-loop, including variables decoupled through a time delay, where the variable *popHealthStatus* can be identified as a slow-working buffer affecting the overall performance of the healthcare system. In a well-functioning healthcare system (high levels of quality and proactive work as well as high levels of *popHealthStatus*) short-term cost savings may for a period cause economic benefits. Meanwhile, the consequences of mistreatment in the healthcare system could be hidden for the decision makers – and

the system performance! – which on the longer term will likely reach a tipping point and the general *popHealthStatus* will be pushed to its threshold value where an escalating vicious cycle of a worse-and-worse behavior in the R1-loop could be activated. While, in an already dysfunctional system, the power of inertia from the R5-loop may diminish the results from actions to improve the R1-loop. Hence, the *popHealthStatus* is potentially the memory of the healthcare system performance, which creates inertia between actions and their effects toward the desired proactive development.

Finally, different actions to offer closer care are represented by the red variables in the CLD, where some were recently implemented in part of the studied region as described in the background, and others are still in their design process. Identifying their interaction with the current system dynamics in the CLD visualizes where these actions may provide support to the desired proactive behavior. At some local hospitals, to have care coordinators in the ED (*careCoordED*) has provento reduce unnecessary hospitalizations, support the rate of risk identifications, and to identify pre-risk patients with potential escalating FA behavior. *MobileHealthClinics* also reactively absorb some of the care users which would otherwise *GoToED,* and provide closer care as a substitute to *GoToPC* having the effect of improving *accessibilityPC*. At the same time, *lowPrioAmbulances* go to patients' home offering basic emergency care which reduce people *GoToED* and send patients on follow-up-controls in PC increasing people who *GoToPC*. Moreover, to offer closer care could include using more local walk-in centers to increase the accessibility of service for low acuity emergency care, as well as having PC open 24/7 in order to improve the *accessibilityPC*. Both are actions to improve the *perceivedAccessPC* and relieve the pressure on the daily work of PC. These support maintaining the required levels of *practWorkPC*. Similarly, the *actionsToIncreaseSeC* would seemingly have a direct effect on increasing *abilitySeC*, yet, having in mind its complex interaction with, e.g., the *popHealthStatus* which potentially can limit peoples' receptiveness, and thus, the success of such strategy.

Besides several more potential consequences, when analyzing the qualitative CLD using mental simulation, we can conclude that pulling the healthcare system into a more proactive balance through the aforementioned improvements will lead to a redistribution of how the total costs sum up (*totCostHCS*); which is a function of how care users seek care (*costED&SpC*), strongly affected by the current level of *proactNeedOfCare*, the amount of care users (*population*), and the *popHealthStatus* which altogether affect care users to live longer healthier lives.

5 CALCULATING THE EFFECTS OF THE DIFFERENT CLOSER CARE STRATEGIES WITH SD

The CLD provided with a systemic common view of the dynamics of elderly care users in the healthcare system and supported processing potential focuses of the SD model. The project group considered that all aspects could not be included in a simulation model due to insufficient data and time. One critical group of care users were elderly with multimorbidity and an FA behavior. To include the appropriate dynamics of this group, all the elderly population in the region was considered as part of the model population. Initially, the focus was on identifying the dynamics of how the elderly got into an FA behavior, but data was hard to identify and the project team experts considered current knowledge and research to be insufficient to its support. Instead, a stock and flow structure was developed of the elderly and their multimorbidity behavior, creating a dynamic theory that challenged the limited one-year perspective that mere statistical data provided. Moreover, three target groups could be identified from studying the data: Gr1) elderly without multimorbidity; Gr2) elderly with multimorbidity; and Gr3) elderly with multimorbidity having an FA behavior which is a subset of near 11% of all the elderly with multimorbidity. The elderly without multimorbidity with an FA behavior were so few they could be completely omitted as a group in the model.

The base structure (BS) of the model includes three parts as shown in Figure 3. These are: BS1) care demand based on elderlies multimorbidity behavior; BS2) calculations of the number of assessments at the ED, subsequent hospitalizations, and days at the hospital for the three defined target groups; and BS3) the estimated costs from the care usage. The model includes three additional parts to incorporate the scenarios of the identified closer care improvement actions: S1) to introduce care coordinators in the ED which identify elderly risk patients coming to the ED; S2) to promote the use of mobile health clinics which reduce unnecessary visits to ED and also hospitalizations; and S3)

to offer a proactive PC service in terms of more thorough consultations to elderly patients which was expected to lead to reductions in the visits to the ED.

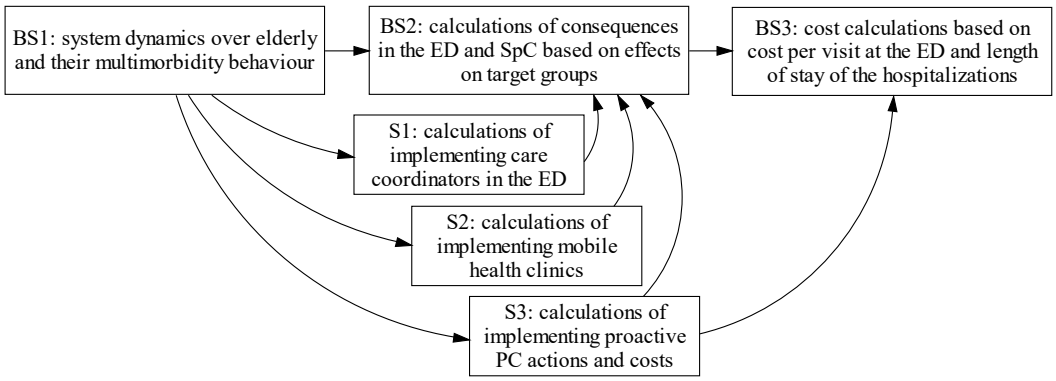


Figure 3 *Simplified overview of the parts in the SD model, using Vensim software.*

Additional information is found in table 2, in which the target groups included in the different scenarios are presented, as well as the results on the employed approximate consequences to the number of visits to the ED, the subsequent hospitalizations, and effects on the length of stay. Regarding costs to implement the tested scenarios, S1 and S2 were already funded by the region in previous years, thus no further investments were needed, while S3 included trading the added resources and costs in PC to the calculated cost benefits.

Table 2 *Overview of data and results of the tested scenarios.*

	Target group (TG)	Visits to the ED	Hospitalizations	Length of stay	Increased cost
S1	11% of (0.1*Gr1+Gr2+Gr3)	no direct effect, reduction of revisits	~30% reduction of which 83% had no hospitalization	same	already funded
S2	Gr2	~0.9→0.8 /person&year	follows reductions of the ED visits	~7% increase	already funded
S2	Gr3	~40% reduction		same	already funded
S3	0.1*Gr1+Gr2+Gr3	20% reduction		same	2* PC visits for TG

Figure 4 displays the simulation results in two inclusive parameters, to the left totaling the costs of the healthcare system and to the right the sum of all the visits to the EDs. An implementation period of three years was utilized. The graphs present the current scenario (line 4) together with the tested scenarios S1 (line 1), S2 (line 2) and S3 (line 3), and the combination of the scenarios (line 5). With respect to costs (to the left), the combination of scenarios S5 is the most beneficial, followed by S3. With respect to quality in care (to the right) leading to lower demand on the EDs, the effect of S2 and S3 is similar, while the combination of scenarios in line 5 is again the most beneficial.

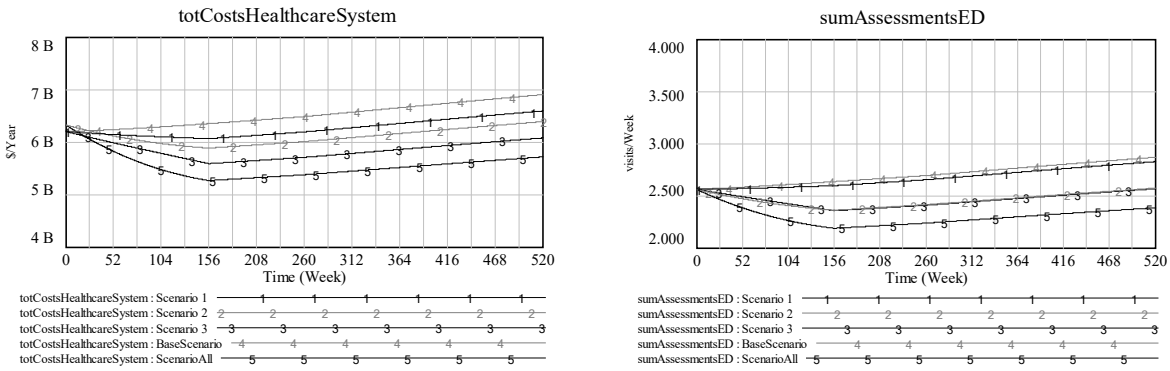


Figure 4 *Result graphs of the tested scenarios on the total cost outcome and care-seeking behavior.*

6 REFLECTING UPON THE PROCESS AND RESULTS. WHAT DID WE LEARN?

This section describes some reflections about the problems encountered during the development of the simulation model, as well as some lessons learned that may serve as tips to those working with SD in similar projects in the healthcare domain. These are:

- *Working with stakeholders with no experience with simulation*: This was the first time the stakeholders worked in an SD simulation study (although they had previous experience with discrete event simulation - DES). This impacted mainly on the problem definition and structuring phase increasing the difficulty to focus the problem to be studied. This was due to the complexity of the problem, but also to the knowledge gap on defining dynamic feedback problems using SD. It was also evident that the expectations about the capabilities of simulation to support decision-making were unrealistic, as a decision maker pointed out: "I hope that with this study you can solve all the problems we have at the ED". Probably a good idea before being involved in a simulation study with non-experienced stakeholders is to provide education to decision makers on the capabilities of the method.
- *Is SD too abstract to be understood?:* SD was chosen for its capabilities of including feedback, studying short- and long-term dynamics, and due to the system-wide perspective of the problem at hand. However, when building the CLD and later on the quantitative model, it was evident that working with SD requires abstraction capabilities from the stakeholders that are not needed with, e.g., DES. This is probably something that should be explained and presented to stakeholders before starting an SD study.
- *Qualitative vs. quantitative*: The systemic nature of SD which allows the inclusion of qualitative and quantitative parameters makes it a very complete and flexible method, however, the experience from this project demonstrated that the stakeholders were not comfortable of including qualitative parameters that lacked proved evidence or statistical data to support the assumptions. This greatly limited the completeness of the model.
- *Time invested vs. knowledge gained*: As Figure 1 shows, the biggest gain in knowledge acquisition comes in the latter stages of the model development process. Especially when different what-if scenarios can be tested to see their potential benefits and drawbacks. However, due to the extensive amount of time taken to define the problem to be studied, as well as continuous changes in the focus of the scenarios to be analyzed, very little time was left to build complete what-if scenarios and make a deeper analysis of the results obtained. Therefore, how the model was used was very limited in comparison to the potential the model could have had to support decision-making.
- *Trying to understand the problem vs. wanting a specific solution:* An additional benefit of using SD is to provide a base for rich discussions including a system perspective. This provides a deep understanding of the problem under study and variables involved, which may be a good base for decision-making. However, the experience in this project showed that the decision makers' priority was not on understanding the problem but on getting a specific solution to the problem instead. In consequence, the CLD was very useful for the project team to understand the problem and the dynamics of the demand for care of the elderly, however, it was not employed by decision makers in their decision-making process. Instead, the decision makers just pursued specific results regarding a reduction in the number of visits to the ED and the number of hospitalizations, as well as the economic gain and loss depending on the scenario tested. This could have been done probably with advanced calculation worksheets instead of using an SD model.
- *Healthcare domain specifics*: There are still many barriers to overcome for extended use of simulation in the healthcare context. More experiences like the one presented in this paper are surely needed as an addition to courses or training to healthcare personnel, decision makers, and policy makers to show the potential of the method to support decision-making.

7 CONCLUSIONS

This paper describes a case study where SD has been employed as a method to model and analyze the demand for care of elderly patients. The objectives pursued with the study included an evaluation of different scenarios to offer closer care to these patients in order to increase the quality of care

provided, to minimize the number of visits to the EDs, as well as to minimize the subsequent hospitalizations. An additional objective involved the analysis of the cost of these scenarios. In order to achieve these objectives, a qualitative model was designed which was valuable to open up discussions and to define and limit the scope of the project. Additionally, a quantitative model was also developed to test the economic effect of applying these closer care scenarios, which were 1) the implementation of care coordinators in the ED; 2) the implementation of mobile health clinics; and 3) employing proactive care in PC. It showed that the best results were provided by combining all three scenarios into a forth scenario. However, this combined scenario had less total benefit than adding the separate results from the scenario 1 to 3 due to the overlap effects from the closer care actions on the identified target groups of the elderly applied in the model.

In addition to the description of the process and results of the case study, this article also reflects on the journey of developing the SD model as well as the lessons learned, which may serve as an input to other simulation modelers working in similar projects in the healthcare domain or with stakeholders without experience with simulation projects.

The project results are being analyzed and employed as a support for decision-making on where to invest to offer closer and better care for elderly patients, which is even economically sustainable.

Further collaboration to analyze other existing problems and improvement areas in the regional healthcare system are nowadays under analysis, and simulation will surely be one of the methods considered to be employed to support the decision makers in these tasks.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the region of Västra Götaland (VGR) for funding the project. The authors are especially thankful to the project group members from the Data analysis department of VGR Catarina Karlberg, Anna Norman Kjellström, Karin Fröjd, and Malin Lönnbark for their engagement during the whole development of the project.

REFERENCES

Boyd C M, Ritchie C S, Tipton E F, Studenski S A and Wieland D (2008). From Bedside to Bench: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Comorbidity and Multiple Morbidity in Older Adults. *Aging Clin Exp Res* **20(3)**:181-8.

Brailsford S C, Harper P R, Patel B and Pitt M (2009). An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation* **3(3)**:130-140. doi: 10.1057/jos200910.

Chang A Y, Ogbuoji O, Atun R and Verguet S (2017). Dynamic modeling approaches to characterize the functioning of health systems: A systematic review of the literature. *Social Science and Medicine* **194**:160-167. doi: 10.1016/j.socscimed.2017.09.005.

Goienetxea Uriarte A, Ruiz Zúñiga E, Urenda Moris M and Ng A H C (2017). How can decision makers be supported in the improvement of an emergency department? A simulation, optimization and data mining approach. *Operations Research for Health Care* **15**:102-122. doi: 10.1016/j.orhc.2017.10.003.

Homer J B and Hirsch G B (2006). System dynamics modeling for public health: background and opportunities. *American journal of public health* **96(3)**:452-458. doi: 10.2105/AJPH.2005.062059.

Hulshof P J H, Kortbeek N, Boucherie R J, Hans E W and Bakker P J M (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems* **1**:129-175.

Katsaliaki K and Mustafee N. (2011). Applications of simulation within the healthcare context." Journal of the Operational Research Society **62(8)**:1431-1451. doi: 10.1057/jors.2010.20.

Kivelä K, Elo S and Kääriäinen M (2018). Frequent attenders in primary health care: A concept analysis. *International Journal of Nursing Studies* **86**:115-124. doi: https://doi.org/10.1016/j.ijnurstu.2018.06.003.

Kunc M, Mortenson M J and Vidgen R (2018). A computational literature review of the field of System Dynamics from 1974 to 2017. *Journal of Simulation* **12(2)**:115-127. doi: 10.1080/17477778.2018.1468950.

LaCalle E and Rabin E (2010). Frequent Users of Emergency Departments: The Myths, the Data, and the Policy Implications. *Annals of Emergency Medicine* **56(1)**:42-48. doi: https://doi.org/10.1016/j.annemergmed.2010.01.032.

Lindgren B (2016). The Rise in Life Expectancy, Health Trends among the Elderly, and the Demand for Health and Social Care. Lund, Sweden: National Institute of Economic Research.

Linnéusson G, Ng A H C and Aslam T (2018). Towards strategic development of maintenance and its effects on production performance by using system dynamics in the automotive industry. *International Journal of Production Economics*. **200**:151-169. doi: https://doi.org/10.1016/j.ijpe.2018.03.024.

Lyons G J and Duggan J (2015). System dynamics modelling to support policy analysis for sustainable health care. *Journal of Simulation* **9(2)**:129-139. doi: 10.1057/jos.2014.15.

Mielczarek B and Uziałko-Mydlikowska J (2012). Application of computer simulation modeling in the health care sector: A survey. *SIMULATION* **88(2)**:197-216. doi: 10.1177/0037549710387802.

Moskop J C, Sklar D P, Geiderman J M, Schears R M and Bookman K J (2009). Emergency Department Crowding, Part 1—Concept, Causes, and Moral Consequences. *Annals of Emergency Medicine* **53(5)**:605-611. doi: https://doi.org/10.1016/j.annemergmed.2008.09.019.

Repenning N P and Sterman J D (2001). Nobody Ever Gets Credit for Fixing Problems that Never Happened: Creating and Sustaining Process Improvement. *California Management Review* **43(4)**:64-88. doi: 10.1109/EMR.2002.1167285.

Salleh S, Thokala P, Brennan A, Hughes R and Booth A (2017). Simulation Modelling in Healthcare: An Umbrella Review of Systematic Literature Reviews. *PharmacoEconomics* **35(9)**:937-949. doi: 10.1007/s40273-017-0523-3.

Sargent R G (2011). Verification and validation of simulation models. In: Jain S, Creasey R, Himmelspach J, White K P and Fu M C (eds). *Proceedings of the 2011 Winter Simulation Conference.* Winter Simulation Conference: Phoenix, Arizona, US, pp 11-14.

Senge P M and Sterman J D (1992). Systems thinking and organizational learning: Acting locally and thinking globally in the organization of the future. *European Journal of Operational Research* **59(1)**:137-150. doi: 10.1016/0377-2217(92)90011-W.

Taylor K and Dangerfield B (2005). Modelling the feedback effects of reconfiguring health services. *Journal of the Operational Research Society* **56(6)**:659-675. doi: 10.1057/palgrave.jors.2601862.

United Nations (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. United Nations.

AUTHOR BIOGRAPHIES

GARY LINNÉUSSON received a Ph.D. degree in Industrial Informatics (2018) from the University of Skövde in Sweden. His background is in Mechanical Engineering (2003) and has both an industrial and academic background from the fields of production engineering and maintenance using system dynamics and discrete event simulation. He is currently a researcher and lecturer at Jönköping University. His research interest is mainly on the problem structuring process using simulation and simulation-based optimization to its support in areas such as industry and healthcare. https://www.researchgate.net/profile/Gary_Linneusson

AINHOA GOIENETXEA URIARTE received a Ph.D. degree in Industrial Informatics (2019) from the University of Skövde in Sweden. Her background is in Computer Engineering (2004) and Industrial Management Engineering (2007), from the University of Mondragon in Spain. She has both an industrial and academic background in the areas of production engineering, innovation, and discrete event simulation. She is currently a researcher and lecturer at the University of Skövde. Her research interest is mainly on simulation and simulation-based optimization for decision support in healthcare and manufacturing system design and improvement. https://www.his.se/en/about-us/Facts-and-figures/staff/ainhoa_goienetxea/

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.013

A TWO STAGE ALGORITHM FOR GUIDING DATA COLLECTION TOWARDS MINIMISING INPUT UNCERTAINTY

Drupad Parmar
Dr. Lucy E. Morgan
Dr. Andrew C. Titman
Dr. Richard A. Williams

STOR-i Centre for Doctoral Training, Lancaster University
Lancaster, LA1 4YW, UK

Prof. Susan M. Sanchez

Operations Research Department, Naval Postgraduate School
Monterey, California 93943, USA

ABSTRACT

In stochastic simulation the input models used to drive the simulation are often estimated by collecting data from the real-world system. This can be an expensive and time consuming process so it would therefore be useful to have some guidance on how much data to collect for each input model. Estimating the input models via data introduces a source of variance in the simulation response known as input uncertainty. In this paper we propose a two stage algorithm that guides the initial data collection procedure for a simulation experiment that has a fixed data collection budget, with the objective of minimising input uncertainty in the simulation response. Results show that the algorithm is able to allocate data in a close to optimal manner and compared to two alternative data collection approaches returns a reduced level of input uncertainty.

Keywords: Input Uncertainty, Data Collection, Budget Allocation

1 INTRODUCTION

The randomness in stochastic simulation is caused by input models which are often represented by probability distributions or processes. When the real-world processes can be observed, samples of data can be collected and used to estimate the input models. The samples of data from which to estimate the input models are finite and thus the input models will never be truly representative of reality. The uncertainty in the estimated input models is propagated through the simulation model resulting in an error in the simulation response known as input uncertainty. Input uncertainty must be quantified, along with stochastic estimation error, to measure the variability around the simulation response and ensure that decisions are made with an appropriate level of confidence; Barton (2012) illustrates the significant risk of ignoring input uncertainty. A reduction in input uncertainty can only be achieved by collecting additional observations from the real-world processes. One way this is done is by studying the contribution made to input uncertainty by each of the input models and specifying how best to allocate a budget for additional data collection amongst the input models.

Here instead of looking at additional data collection to reduce input uncertainty we introduce the idea of guiding the initial data collection process in a manner that minimises input uncertainty. We consider the case of parametric input models and by assuming some knowledge of what values the parameters may take, we develop a two stage algorithm that allocates observations amongst the input models with the objective of minimising input uncertainty. Collecting data in this way is likely to reduce input uncertainty, and hence the level of variability, in the simulation response compared to alternative approaches, thus increasing the level of insight that can be derived from experimental results. This will lessen the need for additional data collection in order to reduce input uncertainty and may also

reduce unnecessary data collection more generally, both of which are particularly beneficial when data collection is expensive and time consuming.

We discuss background literature in Section 2 and detail the existing methodology we build upon in Section 3. We present a new breakdown of the existing methodology in Section 4, which allows us to form and solve optimisation problems which minimise input uncertainty. We describe the two stage algorithm for data collection in Section 5 and illustrate the algorithm with some experiments in Section 6. We discuss some open research questions in Section 7 and then conclude in Section 8.

2 BACKGROUND

Various methods have been proposed to quantify input uncertainty, for an overview of existing techniques see Barton (2012) or Song et al. (2014). We focus on the methodology developed by Cheng and Holland (1997) for the case of parametric input distributions. Here, input model uncertainty reduces to parameter uncertainty and is modelled using a first order Taylor series expansion around the true input parameters. A recent development to this approach was made by Lin et al. (2015), who exploit the gradient estimation method of Wieland and Schmeiser (2006), to estimate input uncertainty in a single experiment. Although initially restricted to the case of stationary input distributions, further work by Morgan et al. (2016) has since extended this input uncertainty quantification method for simulation models which utilise piecewise-constant non-stationary Poisson arrival processes.

The problem of allocating resources for extra data collection was considered by Ng and Chick (2001), who use asymptotic normality properties to approximate the posterior distribution of each parameter. By considering the expected information of additional observations and propagating uncertainty through the simulation using a linear metamodel, they provide sampling plans for further data allocation which aim to reduce input uncertainty effectively. Alternatively Freimer and Schruben (2002) use an ANOVA test to detect whether a parameter has a significant effect on the expected simulation response as the parameter varies over its confidence interval. If the effect is significant then more data should be collected to narrow the confidence interval until the effect is no longer significant. Finally Song and Nelson (2015) model the expected simulation response as a function of the mean and variance of each input model, and use the sample size sensitivity of each distribution to recommend how to collect further data. These methods aim to guide data collection based on input models that have been estimated using real-world observations however our method aims to guide data collection before any real-world observations have been collected. We now describe an existing input uncertainty quantification technique that we will utilise within our approach for guiding data collection.

3 TAYLOR SERIES APPROXIMATION

Consider a simulation driven by L random processes which follow known independent parametric distributions with unknown parameters. Let the unknown but true parameters be denoted by $\boldsymbol{\theta}^c = (\theta_1^c, \ldots, \theta_p^c)$, where $p \geq L$ as some distributions may require more than one parameter. Suppose that real-world data can be collected from each input distribution and that parameters are estimated via their maximum likelihood estimators (MLEs). Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)$ denote the MLEs of the input parameters given the observed data. In this parametric setup the simulation response can be thought of as a function of the input parameters and the output of replication j of the simulation can be denoted by

$$Y_j(\hat{\boldsymbol{\theta}}) = \eta(\hat{\boldsymbol{\theta}}) + \varepsilon_j(\hat{\boldsymbol{\theta}}),$$

where $\eta(\hat{\boldsymbol{\theta}})$ is the expected value of the simulation output given the estimated parameters and ε is a random variable with mean 0 representing stochastic noise.

The goal of the simulation experiment is to estimate $\eta(\boldsymbol{\theta}^c)$, the expected value of the simulation output given the true input parameters. This can be estimated via the sample mean of the simulation output over n replications

$$\begin{aligned}\bar{Y}(\hat{\boldsymbol{\theta}}) &= \frac{1}{n}\sum_{j=1}^{n} Y_j(\hat{\boldsymbol{\theta}}),\\ &= \eta(\hat{\boldsymbol{\theta}}) + \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j(\hat{\boldsymbol{\theta}}).\end{aligned}$$

As n increases the error caused by stochastic noise tends towards 0 however the impact of $\hat{\boldsymbol{\theta}}$ on the expected simulation response is not affected by the choice of n. The variance of this estimator breaks down into two distinct terms, stochastic estimation error and input uncertainty. The former arises from the random variates generated in each replication and can be easily estimated via the sample variance. The latter measures the variability in the expected output due to having estimated the input parameters, that is

$$\sigma_I^2 = \mathrm{Var}[\eta(\hat{\boldsymbol{\theta}})].$$

Using a first-order Taylor series approximation around the true input parameters $\boldsymbol{\theta}^c$, Cheng and Holland (1997) provide the following estimate of input uncertainty

$$\sigma_I^2 \approx \nabla\eta(\boldsymbol{\theta}^c)\mathrm{Var}(\hat{\boldsymbol{\theta}})\nabla\eta(\boldsymbol{\theta}^c)^\top,$$

where $\nabla\eta(\boldsymbol{\theta}^c)$ is the gradient of the expected value of the simulation output with respect to the input parameters $\boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta}^c$. This estimate of input uncertainty depends on how sensitive the simulation output is to the input parameters and how well the input parameters have been estimated. Neither of these terms are known and so both have to be estimated. Note that this method for quantifying input uncertainty has been extended for the case of non-stationary input models by Morgan et al. (2016).

3.1 Variance Estimation

As the parameters are estimated via maximum likelihood estimation, the variance matrix can be approximated by

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}) = \mathbf{I}(\hat{\boldsymbol{\theta}})^{-1},$$

the inverse Fisher information matrix evaluated at the MLEs $\hat{\boldsymbol{\theta}}$. This follows since the asymptotic distribution of the MLEs is multivariate normal with covariance matrix $\mathbf{I}(\boldsymbol{\theta}^c)^{-1}$, and $\mathbf{I}(\boldsymbol{\theta}^c)^{-1}$ can be consistently estimated by $\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}$.

3.2 Gradient Estimation

As the true parameters $\boldsymbol{\theta}^c$ are unknown, Cheng and Holland (1997) approximate $\nabla\eta(\hat{\boldsymbol{\theta}})$ instead of $\nabla\eta(\boldsymbol{\theta}^c)$, however the simulation effort for the method they propose increases linearly with the number of input parameters. Lin et al. (2015) improve upon this by providing a method for estimating $\nabla\eta(\hat{\boldsymbol{\theta}})$ which is independent of the number of input parameters. This method, which extends the work of Wieland and Schmeiser (2006), requires running simulation replications using the fitted input parameters and recording the simulation output along with internal parameter estimates for each replication. Internal parameter estimates are obtained using the realisations generated from the input distributions during a simulation replication, for example inter-arrival times observed within a replication can provide an internal estimate of the arrival rate. Fitting a least-squares regression model, with the simulation output as the response variable and the internal parameter estimates as the explanatory variables, gives a regression model whose coefficients $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}})$ provide an estimator to $\nabla\eta(\hat{\boldsymbol{\theta}})$.

3.3 Contributions to Input Uncertainty

Input uncertainty can then be approximated by combining the estimates for the variance matrix and the gradient vector as follows

$$\sigma_I^2 \approx \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}})\,\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}\,\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}})^\top.$$

This approximation also provides us with an estimate of the contribution made to input uncertainty by each input distribution. Let $\boldsymbol{\theta}_l$ denote the parameter vector for input distribution l, note that this could be a scalar or a vector depending on the distribution. Since the input distributions are independent the variance matrix has a block diagonal form with elements consisting of individual variance matrices $\mathrm{Var}[\hat{\boldsymbol{\theta}}_l]$ for each input distribution. Let $\hat{\boldsymbol{\delta}}(\boldsymbol{\theta}_l)$ denote the gradient vector for the parameters belonging to input distribution l, then

$$\sigma_I^2 \approx \sum_{l=1}^{L} \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}}_l)\mathbf{I}(\hat{\boldsymbol{\theta}}_l)^{-1}\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}}_l)^\top,$$

where the lth term in the sum represents the contribution made to input uncertainty by input distribution l. This breakdown can be used to show which input distributions should be targeted for further data collection in order to reduce input uncertainty, for example see Lin et al. (2015).

4 DATA COLLECTION FOR MINIMISING INPUT UNCERTAINTY

We now present a new breakdown of the Taylor series approximation to input uncertainty which we propose as a tool for guiding data collection. The Fisher information for an i.i.d. sample of size m is simply m times the Fisher information for a single observation. Let m_l denote the number of observations used to estimate the parameters for input distribution l. The Fisher information matrix of $\hat{\boldsymbol{\theta}}_l$ is then given by

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_l) = m_l \mathbf{I}_0(\hat{\boldsymbol{\theta}}_l),$$

where $\mathbf{I}_0$ represents the Fisher information of a single observation. Let $m = \sum_{l=1}^{L} m_l$ denote the total number of observations used to estimate all input distribution parameters. For each input distribution l, we can write $m_l = r_l m$, where $r_l \in (0,1)$ represents the proportion of all observations which are from input distribution l, and $\sum_{l=1}^{L} r_l = 1$. Input uncertainty can then be written as

$$\sigma_I^2 \approx \frac{1}{m} \sum_{l=1}^{L} \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}}_l)[r_l \mathbf{I}_0(\hat{\boldsymbol{\theta}}_l)]^{-1} \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}}_l)^\top. \tag{1}$$

Initially let us consider a set of specific parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. For these parameters the Fisher information matrix can be calculated and the gradient vector can be estimated using the method outlined above. Plugging these into (1) would give an approximation of input uncertainty at $\boldsymbol{\theta}$ in terms of the total number of observations m, and the proportions r_l in which they are allocated to each input distribution. If $\boldsymbol{\theta}$ were the set of true input parameters we could use this to guide data collection by finding the proportions in which to collect data such that input uncertainty is minimised. Note that the proportions which will minimise input uncertainty are invariant to the total number of observations.

We are therefore interested in solving the following optimisation problem

$$\{\min \sum_{l=1}^{L} \frac{a_l}{r_l} \quad \text{s.t.} \quad \sum_{l=1}^{L} r_l = 1 \quad \text{and} \quad r_l > 0 \text{ for } l = 1, \ldots, L\}, \tag{2}$$

where $a_l = \hat{\boldsymbol{\delta}}(\boldsymbol{\theta}_l)\mathbf{I}_0(\boldsymbol{\theta}_l)^{-1}\hat{\boldsymbol{\delta}}(\boldsymbol{\theta}_l)^\top$ and r_l are the proportions to be optimised. This problem can be converted to an inequality-constrained nonlinear programme and solved to optimality by studying the first-order KKT conditions proved by Karush (1939) and Kuhn and Tucker (1951). The optimal proportions are given by

$$r_l = \sqrt{\frac{a_l}{\left(\sum_{l=1}^{L} \sqrt{a_l}\right)^2}}.$$

Alternatively suppose that input parameters $\hat{\boldsymbol{\theta}}$ have been fitted via a collection of real-world data and that input uncertainty has been quantified via the Taylor series approximation. If input uncertainty is a cause for concern then it may be of interest to collect more real-world data in a manner which effectively reduces input uncertainty. This is often done by considering a sampling budget B for extra data collection. If we wish to collect data in a manner such that the overall proportions minimise input uncertainty then there is an extra consideration to be made. Given the budget for collecting extra data and the existing data collected, there is a restriction on how small each proportion can be, that is, each proportion will have a lower bound.

We are now interested in solving the following optimisation problem

$$\{\min \sum_{l=1}^{L} \frac{a_l}{r_l} \quad \text{s.t.} \quad \sum_{l=1}^{L} r_l = 1 \quad \text{and} \quad r_l \geq b_l \text{ for } l = 1, \ldots, L\}, \tag{3}$$

where $a_l = \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}}_l)\mathbf{I}_0(\hat{\boldsymbol{\theta}}_l)^{-1}\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}}_l)^\top$, $b_l = m_l/(m+B)$ and r_l are the proportions to be optimised. Again this problem can be converted to an inequality-constrained nonlinear programme and solved to optimality

using first-order KKT conditions. The optimal proportions are given by finding a partition (I,J) of $\{1,\ldots,L\}$ such that

$$\text{for } l \in I: \; r_l = \sqrt{\frac{a_l}{\mu}} \geq b_l, \qquad \text{for } l \in J: \; r_l = \sqrt{\frac{a_l}{\mu - \lambda_l}} \geq b_l, \quad \lambda_l = \mu - \frac{a_l}{b_l^2} \geq 0,$$

where

$$\mu = \left(\frac{\sum_{l \in I} \sqrt{a_l}}{1 - \sum_{l \in J} b_l} \right)^2.$$

We use the solutions to these two optimisation problems to develop an algorithm that aims to guide the initial data collection process in a manner that minimises input uncertainty. Note that additional constraints could be added to either formulation to incorporate features of the data collection procedure.

5 TWO STAGE ALGORITHM FOR DATA COLLECTION

When modelling some systems, for example medical practices or manufacturing processes, collecting data to estimate the input models may be an expensive and time consuming task. In these scenarios, one may wish to consider a strategy for data collection rather than taking some arbitrary approach. Here we introduce a two stage algorithm to guide data collection. We assume that each input parameter lies in some known interval and study how data might be optimally collected to minimise input uncertainty within these intervals. In the first stage of the algorithm data is collected to hone in towards an optimal collection whilst relaying information about the true parameters values. In the second stage extra data is collected to achieve the proportions that minimise input uncertainty based on the parameter values from the first stage data collection.

Suppose there has been no data collection for a system. Although we have no data from which to estimate the input parameters we shall assume that each input parameter is known to lie in some interval, $\theta_i^c \in [l_i, u_i]$ for $i = 1,\ldots,p$, where the lower and upper bounds, l_i and u_i, are known. For example in a medical practice the number of patients arriving may be known to be between 15-20 per hour, but the exact arrival rate is unknown. The true input parameters could lie anywhere in this p dimensional space and in order to collect data in a way that minimises input uncertainty we need to understand how the optimal proportion changes for each input parameter across the space. An intuitive design that can be used to explore the input parameter space is a 2^k factorial design, which is used to study the effects of k factors each at two different levels (usually high and low) by considering every possible combination of factors and levels. Since there are p input parameters and each is known to be between a lower and upper bound, this naturally lends itself to a 2^p factorial design where each factor is an input parameter with low level l_i and high level u_i.

At each design point we can solve (2) to find the optimal proportions in which to collect data should the parameters at that design point represent the true parameters. Computing the optimal proportions at each design point will give us an idea of the behaviour of the optimal proportion for each input parameter over the specified parameter space. Rather than studying the effects of each parameter, we are instead interested in the minimum and maximum optimal proportion across the design points for each parameter. We use these to form an approximate interval for the optimal proportion for each parameter at the true parameter values. For example, suppose that $p = 2$ so the parameter vector is $\boldsymbol{\theta} = (\theta_1, \theta_2)$. A 2^p factorial design gives $2^2 = 4$ design points which enumerate every combination of factors and levels. Suppose that a 2^2 factorial design gives the optimal proportions shown in Table 1. From this we approximate that the optimal proportions for the true parameters will fall within the following intervals: $r_1 \in [0.3, 0.5]$ and $r_2 \in [0.5, 0.7]$.

Table 1: Example optimal proportions for a 2^2 factorial design

Design Point i	θ_1^i	θ_2^i	r_1^i	r_2^i
1	l_1	l_2	0.5	0.5
2	u_1	l_2	0.3	0.7
3	l_1	u_2	0.4	0.6
4	u_1	u_2	0.4	0.6

Suppose we have a budget B for collecting observations which is to be allocated amongst all of the input distributions. In stage one we aim to allocate as much of the budget as possible without ruling out the true optimal proportions which could occur anywhere within the limits of our parameter space. By allocating the budget according to the minimum optimal proportion for each parameter we can find out information regarding the true parameter values without ruling out any proportions which lie within the approximate intervals. For the example under discussion this would mean allocating 0.3 of the budget to estimating θ_1 and 0.5 of the budget to estimating θ_2, utilising 0.8 of the budget. Collecting data according to this allocation would give us information about the parameters whilst ensuring that any proportions within the intervals can still be achieved by allocating the remainder of the budget. Using the data collected in stage one we can calculate the MLEs, Fisher information matrix and estimate the gradient vector. We can then solve (3) to find the optimal proportions according to the parameter estimates gained from the first stage data collection, using the existing data to set the lower bounds. The remaining budget can then be allocated in order to achieve these proportions and guide the second stage data collection. Putting all these steps together gives us the following algorithm.

Algorithm 1: Two Stage Algorithm for Data Collection

Result: First stage data allocation $m_{\boldsymbol{\theta}_1,1}, m_{\boldsymbol{\theta}_2,1}, \ldots, m_{\boldsymbol{\theta}_L,1}$;
Second stage data allocation $m_{\boldsymbol{\theta}_1,2}, m_{\boldsymbol{\theta}_2,2}, \ldots, m_{\boldsymbol{\theta}_L,2}$;
Initialise two factorial design;
for *each design point i* **do**
  Compute $I_0(\boldsymbol{\theta}^i)$ and $\hat{\boldsymbol{\delta}}(\boldsymbol{\theta}^i)$;
  Find $r_1^i, r_2^i, \ldots, r_L^i$ by solving (2);
end
for *each input model l* **do**
  $r_{l,\min} = \min_i r_l^i$;
  $m_{\boldsymbol{\theta}_l,1} = B \times r_{l,\min}$;
end
Collect data according to first stage allocation;
Compute $\hat{\boldsymbol{\theta}}$, $I_0(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\theta}})$;
Find $r_1, r_2, \ldots, r_L$ by solving (3) using lower bounds $r_{1,\min}, r_{2,\min}, \ldots, r_{L,\min}$;
for *each input model l* **do**
  $m_{\boldsymbol{\theta}_l,2} = B \times r_l$;
end
Collect remaining data to achieve second stage allocation;

6 EXPERIMENTS

In this section we illustrate the algorithm on two examples. We first use an $M/M/1$ queueing model to compare the final allocation of data from the two stage algorithm with the true optimal allocation. Secondly using a more realistic simulation model we compare input uncertainty estimates given by the two stage algorithm against two commonly used approaches for data collection.

6.1 $M/M/1$ Queueing Model

To experiment with the two stage algorithm we use first use an $M/M/1$ queueing model since closed-form expressions can be found for many performance measures. We measure the mean queueing time and since we are able to derive the gradient measures analytically we can calculate the true optimal proportions in which to collect data such that input uncertainty is minimised. To evaluate the performance of the two stage algorithm we can compare the final proportions in which the data is allocated to the true optimal proportions which minimise input uncertainty.

To implement the two stage algorithm let us assume that the input parameters are known to fall within the following intervals: $\lambda^c \in [3,6]$ and $\mu^c \in [9,12]$. We run the simulation for 1000 time periods and the budget for data collection is set to $B = 1000$ observations. Within this controlled experiment we

can set true parameters and use these to generate synthetic data, here we use a Latin hypercube sample with 10 intervals to generate 10 sets of true parameters within the parameter space. At each of set of parameters we run the two stage algorithm 100 times, recording the final recommended proportions in which the data is allocated in each of the experiments. Figure 1 shows boxplots of the final proportion of data allocated to λ (r_1) for each set of input parameters. The red dots indicate the true optimal proportion for each set of parameters calculated using the analytical gradient measures. For each set of parameters the boxplot of proportions from the two stage algorithm is concentrated around the true optimal proportion, demonstrating that the two stage algorithm is able to hone in towards an optimal collection of data. We expect to see some variation around the true optimal proportions for two reasons. Firstly the final proportion of data allocated to λ is based upon the parameter estimates from the first stage data collection and these will not to be equivalent to the true parameters since we only have a finite budget. Secondly the gradient terms are estimated and hence will differ from the true gradient measures. Although variability is evident these results are promising.

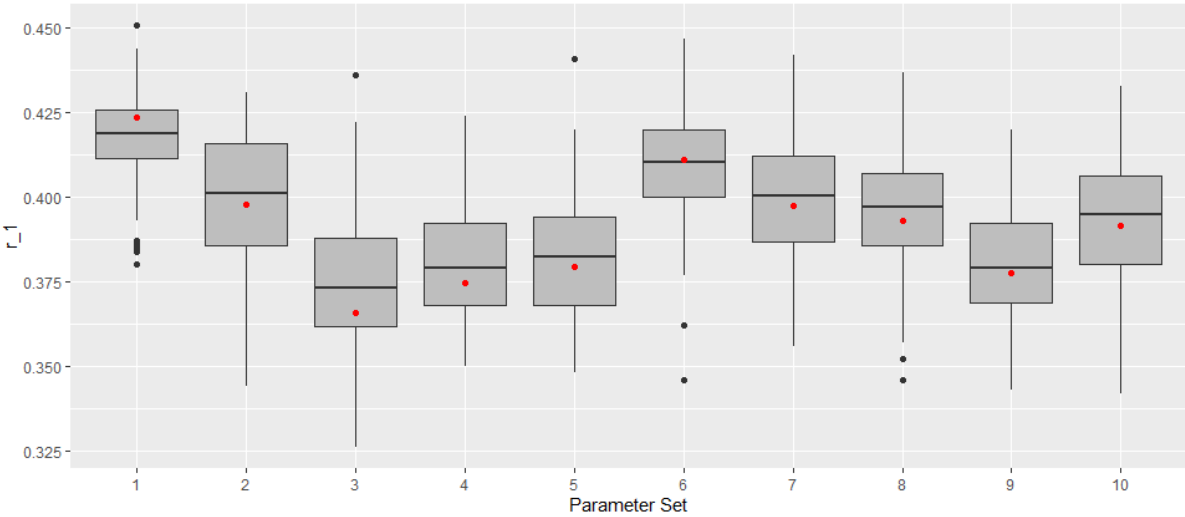


Figure 1: Boxplots showing 100 final proportions for λ given by the two stage algorithm compared to the true optimal proportion at 10 different sets of parameters

6.2 Network Queueing Model

We now consider a more realistic simulation model, a network queueing model consisting of three consecutive multi-server queues. Entities arriving at the system join the queue at node 1. After receiving service at node 1 an entity may leave the system or join the queue at node 2, and similarly after receiving service at node 2 an entity may leave the system or join the queue at node 3. After service at node 3 an entity departs the system. This model could represent a medical centre for example and may be used to solve problems relating to capacity planning and resource allocation. Systems such as this are referred to as operational models of healthcare units in Brailsford (2007).

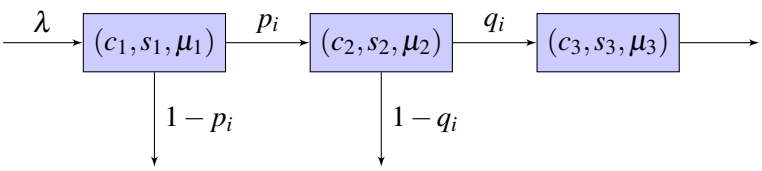


Figure 2: A graphical representation of the network queueing model

The network queueing model is set up as follows. At nodes 1, 2 and 3 there are c_1, c_2, and c_3 servers respectively each of which have a shifted exponential service distribution with parameters (s_1, μ_1), (s_2, μ_2) and (s_3, μ_3). Arrivals to the system follow a stationary Poisson process with rate λ. To represent different demographics of the population the arrivals are split into three different types; 50%

of arrivals are of type A, 30% are type B, and 20% are type C. Each type refers to how likely an entity is to travel through the system and so each is defined by a set of probabilities (p_i, q_i) representing the probability of continuing to node 2 and node 3 respectively. Finally let us suppose the performance measure of interest is the average queueing time weighted by type. For simplicity, we shall assume the the shift parameter of each service distribution and the routing probabilities for each type are known. The unknown input model parameters that require estimation are therefore the arrival rate λ and the three service rates μ_1, μ_2, μ_3. To implement the two stage algorithm we assume that the input parameters are known to fall within the following intervals: $\lambda^c \in [12, 16]$, $\mu_1^c \in [18, 22]$, $\mu_2^c \in [8, 12]$, and $\mu_3^c \in [6, 10]$. The remaining information about the system is known and is as follows: there are two servers at each node $(c_1, c_2, c_3) = (2, 2, 2)$, the shift parameters for the service distributions are $(s_1, s_2, s_3) = (0.05, 0.05, 0.1)$, and the routing probabilities for types A, B, and C are $(p_A, q_A) = (0.4, 0.4)$, $(p_B, q_B) = (0.7, 0.7)$, and $(p_C, q_C) = (0.9, 0.9)$ respectively.

We now evaluate the performance of the two stage algorithm against other data collection approaches by comparing the input uncertainty passed to the simulation response. One alternative we consider is the equal observations approach, where the same amount of data is collected to estimate each input model. This may be the case in a simple service system where arrivals and services are recorded consecutively. To implement this approach we can simply split the budget equally amongst the input models and generate observations from each true input distribution. The second alternative we consider is the timed observation approach, where data is collected by observing the true system over some set period of time. An example of this can be found in Griffiths et al. (2005) where an intensive care unit model was developed using data taken over the course of a year. By using the true parameters to run our simulation model for some chosen period of time we can imitate collecting data from a timed observation of the real-world system.

Within our experiment we wish to compare input uncertainty estimates given by the three approaches when using the same budget. Since the timed observation approach has no fixed number of observations we run this first for 250 time periods. The total number of observations gathered from this approach is then used as the budget for the two stage algorithm and for the equal observations approach. We generate 5 random sets of true parameters so we can compare the approaches across the parameter space when the optimal proportions of the true parameters vary. For each of set of parameters we run the three approaches 100 times. Input uncertainty estimates for each approach are estimated using the same amount of simulation effort and are recorded in the boxplots in Figure 3.

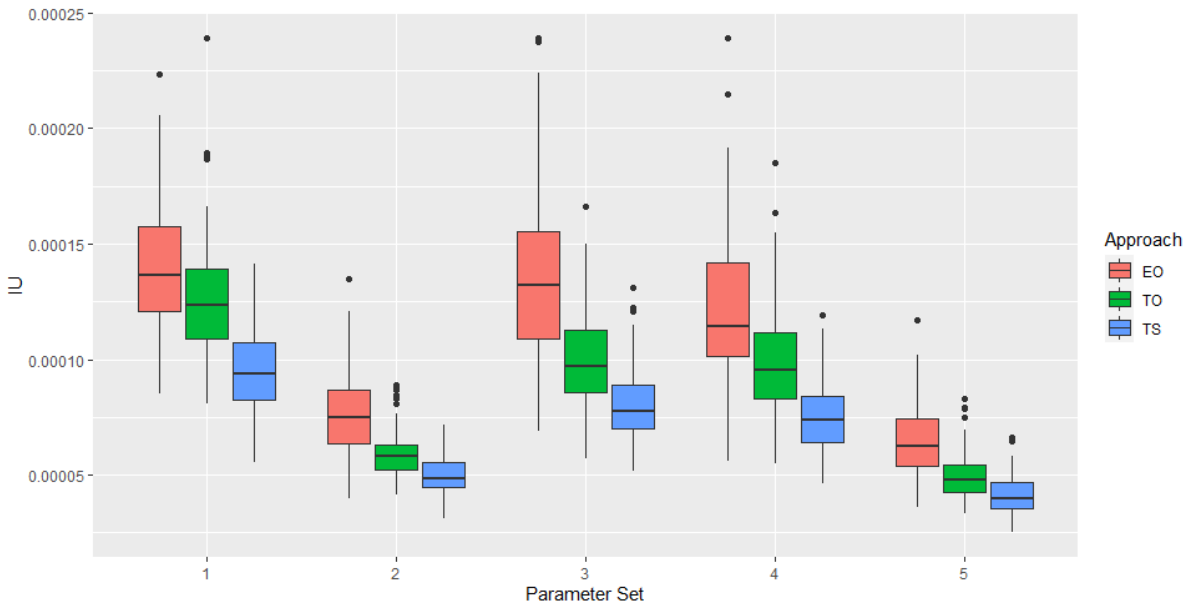


Figure 3: Boxplots comparing 100 estimates of input uncertainty given by the two stage algorithm (TS) compared to equal observations (EO) and timed observation (TO) at 5 different sets of parameters

For each parameter set the two stage algorithm reduced the mean input uncertainty between 31.44% and 40.81% compared to the equal observation approach, and between 16.05% and 24.23% compared

to the timed observation approach. For each set of parameters the two stage algorithm allocates more observations to the arrival rate compared to the other two approaches and by doing so reduces input uncertainty despite using the same overall number of observations. The two stage algorithm shows that by using some knowledge of what values the input parameters might take we can collect data for our input models in a manner that effectively reduces the input uncertainty of our performance measure. A caveat here is that both the equal observation and timed observation approach require no prior knowledge of input parameter values and can be completed in a single collection.

7 FUTURE RESEARCH

The two stage algorithm assumes that each input parameter lies within some known interval however in reality it is unlikely that such an interval is known with complete confidence. Although experts and practitioners may be able provide estimates for such intervals these will only be approximations and therefore we cannot be certain that the true parameters will lie within the intervals. It is worth noting however that the two stage algorithm works regardless of the values that the true parameters take. The second stage allocation minimises input uncertainty for the MLEs calculated from the first stage collection regardless of whether the MLE for each parameter lies in its interval or not. The minimisation however is constrained by the first stage data allocation which is based upon the optimal proportions found from the two factorial design over the parameter space. The issue with true parameters taking values outside their specified interval is that the optimal proportions for these parameters without any constraints may not be achievable as the first stage data collection could rule them out. Preliminary results show that the two stage algorithm still provides a reduction in input uncertainty when parameters lie up to half an intervals width outside their interval, however further research is required here.

Another area which requires further investigation relates to the width of the parameter intervals. Although very wide intervals are more likely to contain the true parameter, in queueing style simulation models they may lead to design points in which the system doesn't reach steady state. In general the design points will represent extreme scenarios in which certain parameters may require barely any data collection and hence have very small optimal proportions. Consequently the minimum optimal proportion for some parameters may be very small which can lead to a small first stage data allocation.

We currently impose no constraint on how large the first stage data allocation needs to be for each parameter. If the minimum optimal proportion for a parameter is small, or the budget multiplied by the minimum optimal proportion is small, then the first stage data allocation will suggest collecting very few observations for that parameter which is likely to lead to an inaccurate parameter estimate. If this is the case then the proportions calculated using the first stage data collection, which are the target proportions, will minimise input uncertainty at a point in the parameter space which may be far away from where the true parameters lie. Consequently the proportions in which the data is collected may differ greatly from the optimal proportions at the true parameters. A possible solution is to introduce a minimum allocation level, meaning that at least a minimum number of observations for each parameter must be allocated in the first stage in order to obtain reasonable parameter estimates.

8 CONCLUSION

In this paper we introduced the novel idea of allocating an initial budget for data collection in a manner that minimises input uncertainty. In particular we have developed an algorithm that by collecting data in two different stages aims to hone in on an optimal allocation of data across the input models. Using an $M/M/1$ queueing model we have demonstrated that the algorithm achieves an allocation of data that is close to the true optimal allocation. On a more realistic simulation model we have shown that the two stage algorithm results in a reduced level of input uncertainty compared to two other viable approaches for data collection. Further experimentation needs to be conducted to gain a deeper understanding of the performance when parameters lie outside their specified intervals, as well as when the parameter intervals are extremely wide.

ACKNOWLEDGEMENTS

This paper is based on work completed while Drupad Parmar was part of the EPSRC funded STOR-i Centre for Doctoral Training (EP/L015692/1).

REFERENCES

Barton, R. R. 2012. "Tutorial: Input uncertainty in outout analysis". In *Proceedings of the 2012 Winter Simulation Conference*. IEEE.

Brailsford, S. C. 2007. "Tutorial: Advances and challenges in healthcare simulation modeling". In *Proceedings of the 2007 Winter Simulation Conference*. IEEE.

Cheng, R. C., and W. Holland. 1997. "Sensitivity of computer simulation experiments to errors in input data". *Journal of Statistical Computation and Simulation* 57 (1-4): 219–241.

Freimer, M., and L. Schruben. 2002. "Collecting data and estimating parameters for input distributions". In *Proceedings of the 2002 Winter Simulation Conference*. IEEE.

Griffiths, J. D., N. Price-Lloyd, M. Smithies, and J. E. Williams. 2005. "Modelling the requirement for supplementary nurses in an intensive care unit". *Journal of the Operational Research Society* 56 (2): 126–133.

Karush, W. 1939. "Minima of functions of several variables with inequalities as side constraints". *M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago*.

Kuhn, H. W., and A. W. Tucker. 1951. "Nonlinear Programming". In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*: University of California Press, Berkeley.

Lin, Y., E. Song, and B. L. Nelson. 2015. "Single-experiment input uncertainty". *Journal of Simulation* 9 (3): 249–259.

Morgan, L. E., A. C. Titman, D. J. Worthington, and B. L. Nelson. 2016. "Input uncertainty quantification for simulation models with piecewise-constant non-stationary poisson arrival processes". In *Proceedings of the 2016 Winter Simulation Conference*. IEEE.

Ng, S. H., and S. E. Chick. 2001. "Reducing input parameter uncertainty for simulations". In *Proceedings of the 2001 Winter Simulation Conference*. IEEE.

Song, E., and B. L. Nelson. 2015. "Quickly assessing contributions to input uncertainty". *IIE Transactions* 47 (9): 893–909.

Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced tutorial: Input uncertainty quantification". In *Proceedings of the 2014 Winter Simulation Conference*. IEEE.

Wieland, J. R., and B. W. Schmeiser. 2006. "Stochastic gradient estimation using a single design point". In *Proceedings of the 2006 Winter Simulation Conference*. IEEE.

AUTHOR BIOGRAPHIES

DRUPAD PARMAR is a PhD student at the Statistics and Operational Research Centre for Doctoral Training in Partnership with Industry at Lancaster University. His research interests are stochastic simulation and input uncertainty quantification. His email address is d.parmar1@lancaster.ac.uk.

LUCY E. MORGAN is a Lecturer in Simulation and Stochastic Modelling in the Department of Management Science at Lancaster University. Her research interests are input uncertainty in simulation models and arrival process modelling. Her e-mail address is l.e.morgan@lancaster.ac.uk.

ANDREW C. TITMAN is a Senior Lecturer in Statistics in the Department of Mathematics and Statistics at Lancaster University. His research interests include survival and event history analysis and latent variable modelling, with applications in biostatistics and health economics. His e-mail address is a.titman@lancaster.ac.uk.

RICHARD A. WILLIAMS is a Senior Lecturer in Management Science at the Department of Management Science, Lancaster University. His research focuses on complex systems science, with particular emphasis on cybernetics and agent-based modelling and simulation to further our understanding of complex dynamical social systems. His email address is r.williams4@lancaster.ac.uk.

SUSAN M. SANCHEZ is a Professor of Operations Research at the Naval Postgraduate School with a joint appointment in the Graduate School of Business & Public Policy. Her interests include the design and analysis of large-scale simulation experiments, robust design, and applied statistics, with application to military operations, manufacturing, and health care. Her email address is ssanchez@nps.edu.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.014

SIMULATION OPTIMISATION FOR IMPROVING THE EFFICIENCY OF A PRODUCTION LINE

Mr. Joseph Calverley
Dr Christine Currie
Dr Bhakti Stephan Onggo

University of Southampton
Highfield Road
Southampton
SO17 1BJ
jac1g17@soton.ac.uk
christine.currie@soton.ac.uk
B.S.S.Onggo@soton.ac.uk

Dr. Thomas Monks

University of Exeter Medical School
Heavitree Road
Exeter
EX1 1TE
t.m.w.monks@exeter.ac.uk

Mr. Michael Higgins

PowerTrain Manufacturing Engineering
Ford Motor Company
Dunton Research Centre
Basildon
SS15 6EE
mhiggi37@ford.com

ABSTRACT

Finding the best set up for a production line is a traditional simulation problem. Here, we apply and adapt Optimal Computing Budget Allocation (OCBA), a well-known method for optimisation via simulation to find the best design for a production line. Typically OCBA is implemented in a sequential fashion with the results of one (or a small number) of replications being used to adapt suggest how the sampling should be allocated in the next step. In this paper we change that format to fit in with the typical experimental set up at Ford and instead work in five main stages. Each stage is allocated a set amount of simulation time and we use OCBA to determine how long to run the simulation for with each system configuration. The results show that using OCBA can substantially increase the efficiency of selecting the best out of a number of designs.

Keywords:

Optimisation via simulation, production line, simulation

1 INTRODUCTION

Simulation models are frequently built with the aim of choosing between a number of different system designs and this is the situation we consider here. When there are a large number of designs to choose

between and/or the runtime of the simulation model is very long, using methods that optimally allocate simulation time to the different designs can have a significant impact on the efficiency of the experimentation. This is especially important when optimising in real time using a digital twin or symbiotic simulation (Xu et al. 2016).

Ford are keen to implement efficient simulation experimentation to both increase the productivity of their simulation group and move to implementation within a symbiotic simulation. Here, we define a symbiotic simulation to be a model that is automatically fed with data from the real system, and is used to test out different system configurations. We assume in this case that the change in configuration needs to be carried out regularly, and that this kind of problem is encountered in, for example, job shop production systems and flexible manufacturing systems. Our definition of symbiotic simulation in this paper is very similar to the digital twin idea but with a focus on developing optimisation methods that are carried out frequently during production (and hence need to be computationally efficient) rather than simply replicating the system and forecasting future behaviour.

Optimisation via simulation (OvS) methods aim to find either a single optimal solution or a set of optimal solutions using stochastic simulation for evaluation. Solving these problems is difficult because of the randomness of the simulation output and OvS is an active field of research (see (Fu 2015) for an overview).

The problem we consider here is described as *ranking and selection*, where we are choosing between a finite number of discrete options. Branke et al. (2007) classify ranking and selection algorithms into three categories: (i) the indifference zone approach; (ii) expected value of information procedures (VIP) and (iii) optimal computing budget allocation (OCBA). Indifference zone approaches such as Kim and Nelson (2006) aim to guarantee a probability of correct selection (PCS) of the optimal configuration while OCBA and VIP procedures look for the best allocation of a finite simulation budget. The latter philosophy fits much better with the implementation of simulation optimisation within a symbiotic simulation and we choose to adapt OCBA in what follows.

In the remainder of the paper we introduce OCBA (Section 2), before going on to formulate the problem (Section 3) and describe the methodology (Section 4). We implemented the optimisation routine on a case study from a manufacturing line and we describe this in Section 5 alongside some preliminary results.

2 OCBA

For readers with a particular interest in finding out more about OCBA, there is an excellent book by Chung-Hun Chen and Loo Hay Lee that describes its use (Chen and Lee 2010). Here, we give an overview of how it works.

OCBA is a heuristic method that aims to efficiently allocate simulation time to each of a number of competing designs. Its basic premise was first introduced in the 1990s (Chen 1996) and is centred around the fact that allocating more simulation time to a design allows a more precise answer to be obtained about its output. As a result, OCBA will tend to allocate more time to designs which are critical to the decision over which is most important; for example, designs which are definitely not contenders for the best will be allocated little simulation time.

Sequential OCBA algorithms will allocate the next simulation replication to the design that most increases the estimated approximate probability of correct selection (EAPCS) relative to the current estimate of the probability of correct selection (PCS) in each time step. Improvements to the efficiency of the initial OCBA heuristic have mainly resulted from refining the assumptions made when estimating EAPCS, allowing a better estimate of how allocating more simulation time to a design will help to solve the problem.

OCBA begins by simulating each design a limited number of times, to obtain information about their means and variances. OCBA then determines the most critical designs, those which are both close to being the best and/or highly variable, and assigns further simulations to these critical designs, until the full computing budget is realized. Simpler sampling methods, such as Equal Allocation (equal number of observations of each system) and Proportional to Variance (allocate more observations to systems with

a higher variance) are highly inefficient. Given a limited computing budget, these algorithms will fail to consistently find the true best.

We assume that we are choosing between k designs and have a computational budget of T, meaning that we can run a maximum of T replications during the experiment. Note that for simplicity we assume each replication involves the same amount of computational effort. In the most general OCBA algorithm that we describe below, we assume that a total of Δ replications are allocated at each step of the algorithm. As we run $n_0 \geq 5$ replications for each of the k designs during the initialisation phase, $T - kn_0$ must be a multiple of Δ. Algorithm 1 describes the basic process of OCBA where we aim to maximise the PCS for a fixed number of replications.

Algorithm 1: OCBA

Given k, T, Δ, n_0 (and let N_i^ℓ be the number of replications of design i at stage $\ell = 0, 1, \ldots$).
  Assume that the output of interest for design i in replication j is $L(\theta_i, \omega_{ij})$.
Perform n_0 simulations for all k designs $N_1^0 = \ldots = N_k^0 = n_0$; $\ell = 0$
while $\sum_{i=1}^k N_i^\ell < T$ **do**
  Calculate sample means $\bar{J}_i = \frac{1}{N_i^\ell} \sum_{j=1}^{N_i^\ell} L(\theta_i, \omega_{ij})$;
  Calculate sample standard deviations $s_i = \sqrt{\frac{\sum_{j=1}^{N_i^\ell} (L(\theta_i, \omega_{ij}) - \bar{J}_i)^2}{N_i^\ell - 1}}$;
  Find $b = \arg\min_i \bar{J}_i$
  Set $\delta_i = \bar{J}_b - \bar{J}_i, \forall i \neq b$
  Allocate observations:
  if $b = 1$ **then**
    $N_2^{\ell+1} = 1$; $N_i^{\ell+1} = \frac{(s_i/\delta_i)^2}{(s_2/\delta_2)^2}$
  else
    $N_1^{\ell+1} = 1$; $N_i^{\ell+1} = \frac{(s_i/\delta_i)^2}{(s_1/\delta_1)^2}$
  $N_i^{\ell+1} = \frac{\left(\sum_{i=1}^k N_i^\ell + \Delta\right)}{\sum_{i=1}^k N_i^{\ell+1}} N_i^{\ell+1}$
  Simulate each design a further $\max\left\{N_i^{\ell+1} - N_i^\ell, 0\right\}$ more times
  $\ell = \ell + 1$
Output current best b.

3 PROBLEM FORMULATION

We consider the minimisation of the time engines spend on a production line, so-called dock-to-dock times. A set of 15 different system designs are considered. What is particularly interesting for this article is that Ford place a restriction of 5 on the number of replications that can be run for each design during a simulation experiment. However, the length of each of those replications can be varied independently. In the example considered in Section 5, we measure the length of a replication by the number of engines that enter the simulation, using this as a proxy for simulation duration.

By restricting the number of replications to 5 but allowing variability in runtime, we first need to rephrase the OCBA algorithm to become one of allocating simulation durations to different designs rather than allocating simulation replications. This is a relatively straightforward change but raises an interesting question of how much total duration we should allocate to each of the 5 replications, which we discuss in the following section.

4 METHODOLOGY

We provide an overview of the method in Algorithm 2.

Algorithm 2: Implementing OCBA with a fixed number of replications

```
1 Input total simulation runtime T, number of replications N, number of designs P, runtime in each
    stage t_n, n = 1,...,N
2 for n = 1 to N do
3     if n = 1 then
4         Set the duration of each design to t_1/P
5     else
6         Set the duration of design p to t_np, where t_np = N_p^n / t_n where we find N_p^n using the method
            described in steps 10-14 of Algorithm 1
7     Run one replication of the simulation model for each design with the duration set to t_np
8     Update estimates of the mean and variance of the output for each design
9 Output the design with the lowest mean
```

As we show below in the results, we can improve the efficiency of the algorithm through a careful choice of the allocation of runtime to each replication, $t_n, n = 1, \ldots N$, where $N = 5$ for the Ford example. Intuition suggests that setting the t_n such that $t_1 < t_2 < \ldots < t_N$ is likely to result in a more efficient allocation because we spend longer on replications where we have a better understanding of which designs are critical to making the correct selection. In the case study, we test setting $t_n = \frac{n^q}{\sum_{n=1}^{N} n^q} T$ for q in the range $[1,2]$.

5 CASE STUDY

When investigating the method, we resample from simulation output that has previously been produced by a Ford simulation model. When implementing this method in practice, the algorithm would instead be used to determine the runtime of each design in each replication. We compare 15 different designs and the simulation data we resample from consists of 75 data files, each containing data from one replication for one design, excluding the warm up period. Each file contains approximately $65,000$ data points.

Figure 1 displays a box plot of the results for each design, using all of the simulated replications. This shows that designs 3, 11 and 12 have very similar mean dock-to-dock times.

5.1 Initial Analysis and Checks

We carry out two statistical checks on the data before beginning the optimisation.

- **Check for autocorrelation in the output**
  As the simulation data are in the form of a time series, there is a possibility that earlier data points can influence the values of later data points, a phenomenon described as *autocorrelation*. If this is present then instead of using the raw data, it is necessary to batch adjacent points together and use the means of these batches as our input data to the optimisation algorithm. For the data set we consider here, there is no evidence of autocorrelation in any of the data files; therefore we have not used batching.
- **Check that the output follows a normal distribution**
  The method underlying OCBA assumes that outputs follow a normal distribution. If the data are not normal, the approximations of the expected improvements in the probability of correct selection are less accurate and the convergence results no longer hold. Nonetheless, the algorithm is still expected to provide a good guide of how to allocate observations. We find that none of the data follows a normal distribution.

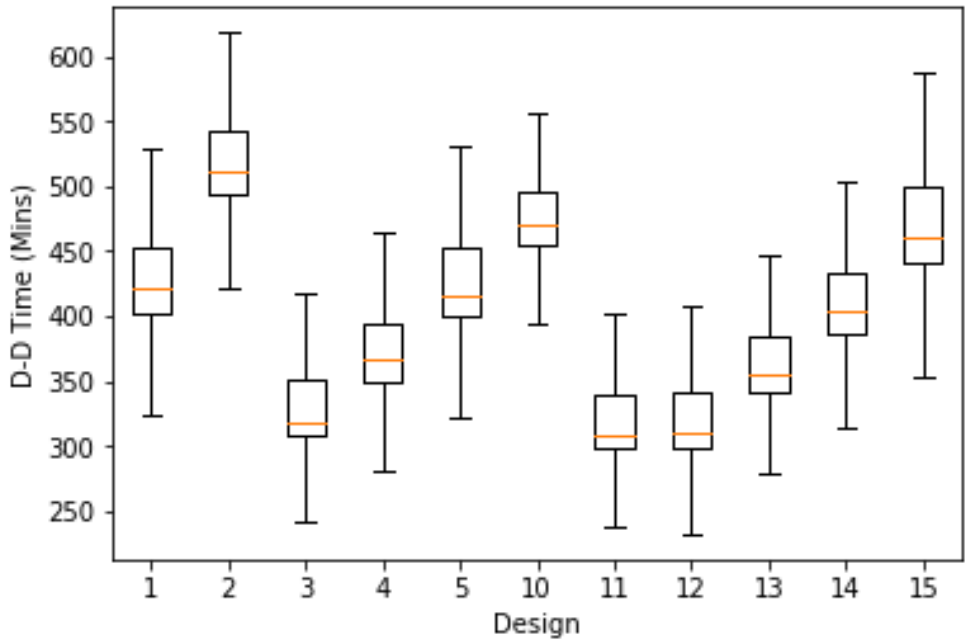


Figure 1: Box Plot of the Dock to Dock Times for each Design. Values are taken from all replications.

5.2 Implementing OCBA-PCS

OCBA-PCS aims to maximise the Probability of Correct Selection PCS. As a baseline, we test the method using the standard sequential algorithm (rather than with limited replications), deciding which design to sample next at each timestep. The number of timesteps in the first stage is $t_1 = 5P$, 5 per design; and the simulation budget $T = 11,000$. We ran the method 100,000 times, selecting the correct design 95.37% of the time, using approximately 0.003% of the total data points provided. We use this as the baseline for comparison with our method using the Ford set up of 5 replications. Figure 2 describes how the sequential OCBA algorithm works.

When implementing OCBA across 5 stages, we first consider how to allocate the runtime across the different stages such that the ratio of runtimes is $1^p : 2^p : 3^p : 4^p : 5^p$ across the 5 stages. We measure PCS where $T = 5000$ for different values of p and preliminary results are presented in Figure 3.

Preliminary results suggest that the algorithm with 5 replications performs well and significantly reduces the number of observations Ford need to make decisions about the structure of their lines.

6 CONCLUSION

Preliminary results presented here suggest that when using a small number of replications but allowing the runtime to vary between each design, it is possible to implement OCBA effectively. This has the potential to drastically reduce the duration of experiments and in the Ford example, we used less than 0.003% of the original simulation output data. This makes the simulation optimisation feasible to incorporate in a symbiotic simulation.

There are several avenues for future research. While maintaining the structure implemented by Ford in which we vary the runtime per replication rather than the number of replications, one consideration is the optimal number of replications. Initial numerical results suggest that 5 replications works well but we need to perform a proper comparison with a full sequential algorithm to better gauge the quality of its performance.

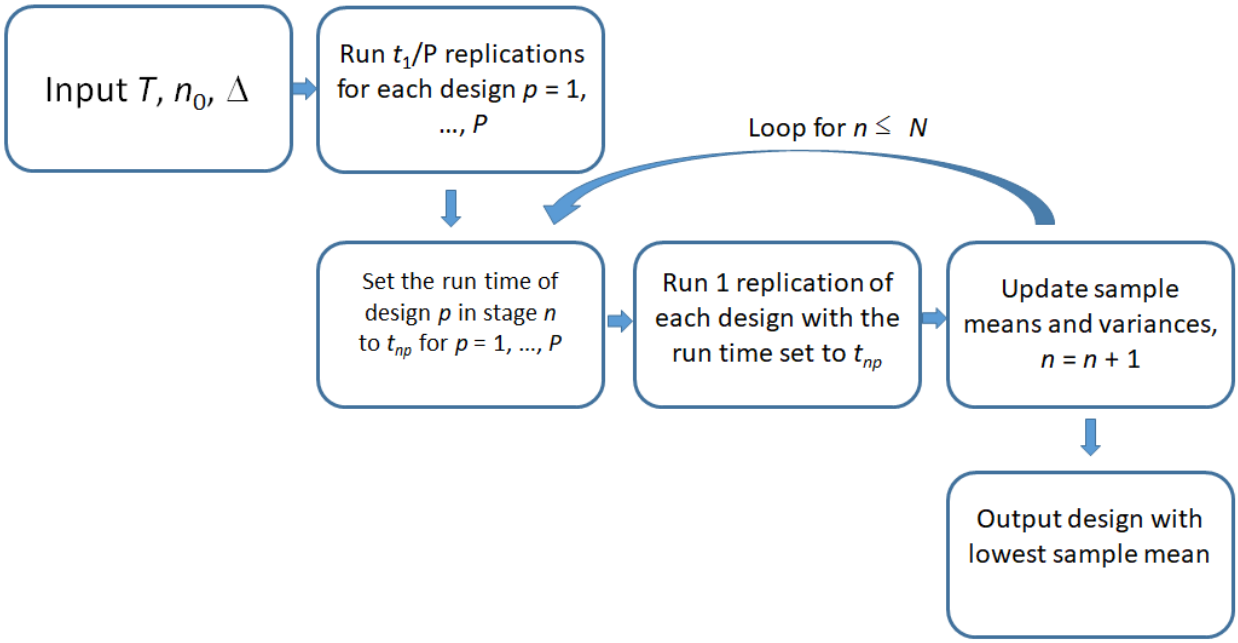


Figure 2: Graphic representation of OCBA based on PCS

In the work we describe here, we aim to maximise the probability of correct selection but a stream of research exists in which the aim is instead to minimise the *Expected Opportunity Cost* (EOC). PCS methods know only if a design is wrong but EOC methods provide a measure of the penalty in making an incorrect selection. The EOC idea was first introduced by (Chick and Wu 2005) and has an intuitive appeal for problems such as this because it directly relates to the economics.

Applying these ideas to other data sets would also be beneficial to determine whether the conclusions hold in other situations.

ACKNOWLEDGMENTS

Joseph Calverley was funded by an EPSRC vacation grant at the University of Southampton (EP/R513325/1).

REFERENCES

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure". *Management Science* 53:1916–1932.

Chen, C.-H. 1996. "A Lower Bound for the Correct Subset-Selection Probability and its Application to Discrete Event Simulations.". *IEEE Transactions on Automatic Control* 41:1227–1231.

Chen, C.-H., and L. H. Lee. 2010, 07. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*.

Chick, S., and Y. Wu. 2005. "Selection procedures with frequentist expected opportunity cost bounds". *Operations Research* 53:867–878.

Fu, M. C. 2015. *Handbook of Simulation Optimization*. 1st ed. New York, USA: Springer.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Elsevier Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. Nelson. Waltham, MA: Elsevier.

Xu, J., E. Huang, L. Hsieh, L. H. Lee, Q.-S. Jia, and C.-H. Chen. 2016. "Simulation optimization in the era of Industrial 4.0 and the Industrial Internet". *Journal of Simulation* 10 (4): 310–320.

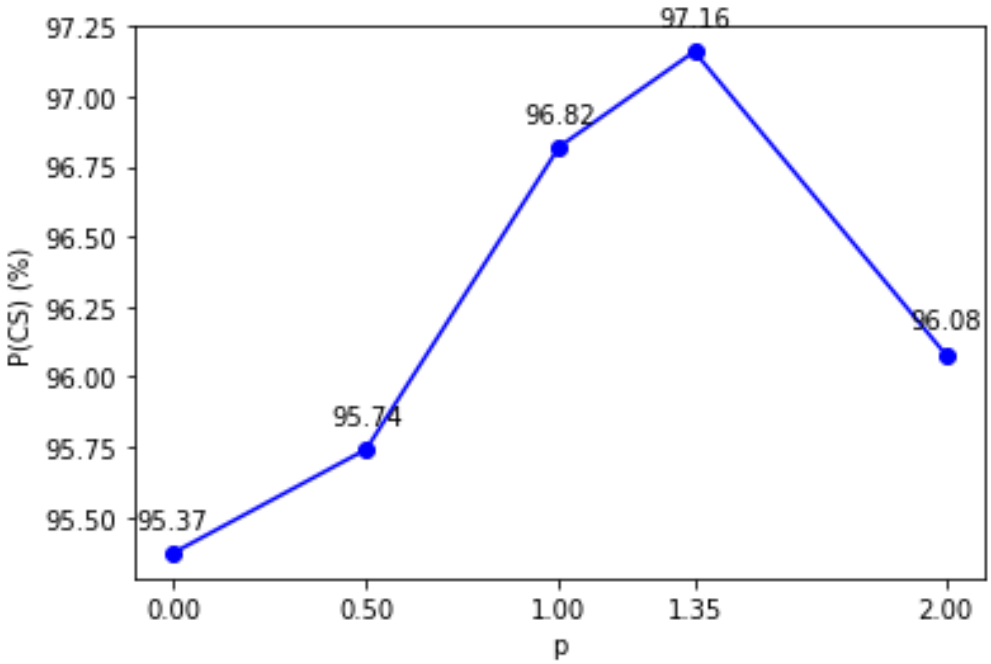


Figure 3: Effect of change in p on P(CS)

AUTHOR BIOGRAPHIES

JOSEPH CALVERLEY is an undergraduate student at the University of Southampton studying for a BSc in Mathematics, Operational Research, Statistics and Economics. He worked on this project during his summer vacation.

CHRISTINE CURRIE is Associate Professor of Operational Research in Mathematical Sciences at the University of Southampton, UK, where she also obtained her Ph.D. She is Editor-in-Chief for the Journal of Simulation. Christine was co-chair of the Simulation Special Interest Group in the UK Operational Research Society until September 2013. Her research interests include simulation optimisation, mathematical modelling of epidemics, optimal pricing and applications of simulation in health care. See http://www.southampton.ac.uk/maths/about/staff/ccurrie.page for more details.

THOMAS MONKS is an Associate Professor of Health Data Science at the University of Exeter Medical School and The Institute of Data Science and AI, University of Exeter. He is also Turing Fellow at The Alan Turing Institute and Honorary researcher at the Clinical Operational Research Unit, UCL. Between 2014 and 2019 he led an applied healthcare Data Science team funded by the NIHR. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His speciality is simulation of healthcare systems of unscheduled and emergency care. He was co-chair of the UKs Simulation Workshop 2016 and 2018. His email is t.m.w.monks@exeter.ac.uk webpage is: https://www.turing.ac.uk/people/researchers/thomas-monks

STEPHAN ONGGO is an Associate Professor of Business Analytics at Southampton Business School, the University of Southampton. He is a member of the Centre of Operational Research, Management Sciences and Information Systems (CORMSIS). His research interests lie in the areas of simulation modeling methodology (symbiotic simulation, hybrid modeling, agent-based simulation, discrete-event simulation) with applications in operations and supply chain management and health care. He is the associate editor for the Journal of

Simulation and the chair of The OR Societys Simulation SIG. His website is https://bsonggo.wordpress.com

MICHAEL HIGGINS graduated with a Bachelor of Science in Mathematics and Physics in 2011 and in 2014 gained an MPhil (Eng) through his research into the application of Discrete Event Simulation within the automotive industry. He has worked with Ford Motor Company since 2012, where his current role is within PowerTrain Manufacturing Engineering (PTME) as Lead Simulation Engineer. He has gained a holistic view of powertrain manufacturing and has utilized simulation as a tool across several functions within PTME including plant and production engineering roles. He is currently researching and developing methods to increase usability and widen the application of the Ford Simulation Toolset. His email address is mhiggi37@ford.com.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.015

Analysing the Effect of Food Supply Chain Traceability on Product Waste

Mr. Rob Glew

Insitute for Manufacturing
University of Cambridge
rg522@cam.ac.uk

Dr. Marco Perez Hernandez

Institute for Manufacturing
University of Cambridge
mep53@cam.ac.uk

Prof. Duncan McFarlane

Institute for Manufacturing
University of Cambridge
dm114@cam.ac.uk

ABSTRACT

This paper presents initial results from an agent-based simulation study into the impact of supply chain traceability information sharing on food waste reduction in the fresh food supply chain. Based on data collected during a 2019 study of a multi-tier citrus fruit supply chain, a conceptual model of information sharing relationships and their impact on food waste was created. This model formed the basis of an agent-based simulation where the actors in the supply chain share time-and-temperature storage information for the fruit. The model is successfully verified against the case study data, initial findings show that the benefits of traceability are realised further downstream in the supply chain. We further comment on the environmental and social implications of our results.

Keywords:

Supply Chain Visibility, Food Waste, Agent-based Model

1 INTRODUCTION

Food waste is a pressing economic, environmental, and social concern. In the UK alone, approximately 10 million tonnes of food was wasted in 2016 (WRAP 2017). Over 60% of fresh food waste could have been avoided by better practices in the supply chain and more sensible purchasing by consumers (Barratt and Oke 2007). Globally, total food waste is over 1.3 billion tonnes and is growing (Gustavsson et al. 2011). Higher income countries waste a higher proportion of total food production than lower income countries, leading to social inequality (Gustavsson et al. 2011). All told, avoidable food waste in the UK contributes as much as 3600Kg of CO2 equivalent environmental impact per tonne of waste, or approximately 36 billion Kgs of CO2 equivalent per year (Tonini et al. 2018).

Given this, there is strong political and industrial interest in interventions to reduce food waste (de Moraes et al. 2020). However, the breadth of academic studies into food waste reduction in the food supply chain (FSC) is limited, focusing primarily on perishability-related pricing practices (Grillo et al. 2017) or on consumer action (Morone et al. 2018). Recently, firms in all industries have begun to consider the sustainability of their whole supply chain, rather than solely their own activities (Villena and Gioia 2020). This has led to calls for greater supply chain traceability, transparency, and information sharing (Gardner et al. 2019).

The optimisation of a perishable food product supply chain requires information sharing to match supply with demand and to ensure products are sold according to their quality and residual shelf-life (RSL) (Spada et al. 2018).

With this simulation we seek to understand the relationship between food waste and traceability information sharing in a perishable food supply chain. We ask: What is the impact of traceability on

food waste in a perishable food supply chain? and address the question with a quantitative, agent-based simulation of a citrus food supply chain. Empirical data obtained from a case study is used to validate the model. The use of an agent-based approach allows the model to consider the competitive impact of supply chain partners sharing information, as it is acknowledged that firms are often unwilling to share data they perceive as commercially sensitive (Bartlett et al. 2007).

2 FOOD WASTE MODELLING APPROACHES

This study uses the definition of traceability from Olsen and Borit (2013), "The ability to access any or all information relating to that which is under consideration, throughout its entire life cycle, by means of recorded identifications" (pp. 147). In this case, specifically regarding storage time and temperature traceability information. Following Barratt and Oke (2007) we define visibility as "the extent to which actors within a supply chain have access to or share information which they consider as key or useful to their operations" (pp. 1218) and define information sharing as an activity intended to produce or increase visibility.

Food waste is the most pressing issue in global food supply chains (Gustavsson et al. 2011). It has, therefore, attracted significant industrial and academic attention. Food waste is generally defined as food lost throughout the supply chain from farm to fork, including damage of crops during harvest, damage arising from transport, incorrectly processed or handled food, and expired food (Griffin et al. 2009).

The studies by de Moraes et al. (2020), Liljestrand (2017), and Kaipia et al. (2013), indicated that more information sharing and better IT integration would reduce food waste. Improved inventory management practices and order picking result in maximising the residual shelf life (RSL) of products in the supply chain (Kaipia et al. 2013).

The application of simulation approaches to food waste reduction practices is limited. Such studies have focused solely on the relationship between food waste, food perishability, and pricing. Authors approach the problem from the economic perspective, rather than the environmental or social sustainability perspectives. As such, the goal of these studies is to maximise profit for the retailer by reducing the amount of product they dispose of, ignoring the multi-tier perspective.

Wang and Li (2012) developed a simulation where pricing decisions are based on dynamically identified RSL of the food product. Quality is modelled as a dynamic state that decays linearly over time, until it passes a threshold where the product can no longer be sold. The objective is to estimate as closely as possible the quality and RSL of the products on the retailer's shelves and adapt the price to maximise sales. Yu and Nagurney (2013) adopted a similar approach, where the objective is to maximise the RSL at the retail stage.

More recent studies of the relationship between pricing strategies, perishability, and food waste include Chang et al. (2016), where the authors included an agent-based simulation of consumer preferences. Chen et al. (2019) used a linear decay rate of quality and RSL in a game theoretic simulation of a two-echelon food supply chain. Here the goal was to maximise the profits for the both the supplier and the retailer. Grillo et al. (2017) considered the relationship in the context of Spanish fruit supply chains, a similar context to the case study of this paper. Integer non-linear programming was used to maximise the RSL of products arrive with the consumer.

In a literature review of food supply chain modelling approaches, Zhu et al. (2018) concluded that food waste simulation had not been explored beyond the objective of maximising profit through pricing strategies. In pointing to future research directions, the authors highlighted the need to understand the relationship between traceability, visibility, and sustainability in food supply chains. To date, no models have addressed the value of information sharing to food waste reduction. Zhu et al. (2018) also highlighted the lack of models that account for the increasing role of digital technologies such a supply chain traceability systems in supply chain management, which enable further collaboration and coordination. As a consequence, this study sets out to address the specific issue of the potential impact traceability could have on food waste.

3 MODEL DEVELOPMENT

3.1 Case Background

This study is based on data collected during a case study research project of a multi-tier fresh food supply chain. Supply chain data was collected and in-depth interviews were conducted with a leading fresh fruit distributor in the UK. This firm supplies a major supermarket chain, has an annual turnover of over £200 million and sources from over 500 farms in 11 countries. In their largest product line, *soft citrus*, over 65% is sourced from Spain, equating to 3500 truck loads annually. This modelling study focuses solely on the Spanish soft citrus supply chain for the partner firm.

This firm was selected as they had recently implemented a full chain traceability system for storage time and temperature information. Figure 1 shows a schematic representation of the actors in the supply chain. Growers (farm cooperatives who are also responsible for packing the fruit) supply a UK-based distributor via 3rd party logistics. A second transport stage takes fruit from the distributor to the retailer's locations around the UK, where they are stored before being displayed for sale to consumers.

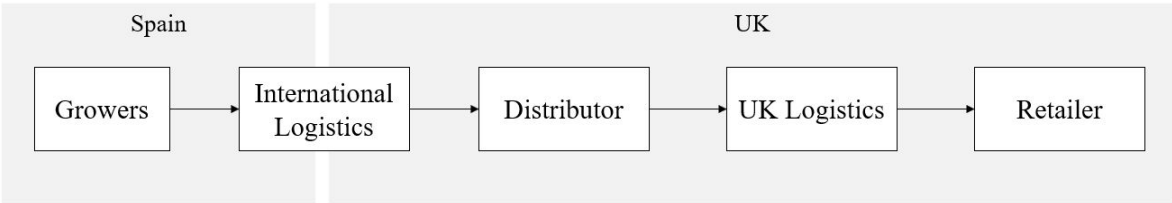


Figure 1: Schematic supply chain map

In 2015, before the traceability system was implemented, the waste level for Spanish soft citrus was 2.54%. In 2019, once the traceability was in place and part of business-as-usual operations, the waste level at reduced to 1.86%. In this supply chain, waste primarily occurs due to the inability of actors to accurately estimate the residual shelf life of the products. The most desirable inventory management policy is *"First to Decay, First Out"*, however, without detailed traceability about the life of the product it can be impossible to estimate RSL. RSL is a function of product age and storage conditions (Rong et al. 2011). A smaller portion of waste occurs due to mishandling of products during transit, this is estimated to be between 10 and 15% in this case.

3.2 Selection of Modelling Methodology

An agent-based model was selected for this study due to its ability to simulate the dynamics and interactions between the self-interested agents in a complex system such as the supply chain (Herrera et al. 2020). Simulation is growing methodology in supply chain research and agent-based approaches have been used to study supply chain risk aversion (Cannella et al. 2019), technology adoption (Basingab 2019), and many other phenomena. In the context of this study, traceability information can be competitively sensitive and firms may be hesitant to share it with partners. This effect can be included in agent behaviour to increase the accuracy of the model.

3.3 Key Elements of Simulation Model

The simulation model incorporates four types of agents that represent the key actors in the soft citrus supply chain presented in Figure 1. The model focuses in the operation of the distribution echelon when dealing with different configurations of visibility of products with the aim of minimising the product wastage along the chain.

A key agent type is the *Distributor* agent that gathers products coming from growers and dispatches them to the retailer. The distributors receive orders from retailers and breakdown these orders to place individual orders to growers. The order periodicity can be parameterised. Once the product is received, the processing time of the distributor is also parameterised and represents the time for quality control and handling, previous to the dispatching. The distributor uses a *"First to Decay, First Out"* inventory management policy by default. Hence, the more visible the information about actual remaining shelf life of the product, the better the dispatch decisions the distributor makes to reduce product wastage.

Upstream, the *Grower* agents encapsulates the activities between harvesting and packing of product, until the produce is ready to circulate along the chain for fulfilling the orders. The production is based on a normal distribution, whose mean is based on the production capacity and can be parameterised together with its standard deviation. The growers dispatch product following a FIFO policy, based on the orders received from the distributor. Each grower has different processing times and also different levels of information sharing. Hence, a grower might deliver full traceability information whereas other do not. On the other end of the chain, the *Retailer* agent incorporates the behaviour for periodically placing orders to distributor according to the demand forecasts. The demand follows a normal distribution, whose mean and standard deviation can be parameterised. Once product is received, this is offered to customer and sold within a given user-defined time.

Finally, the *Transporter* agents enable the flow of product along the chain. Each transporter covers a distance for delivery of products based on real-world map and route. The time it takes to deliver products depends on the parameters of velocity, the distance and the maximum number of hours that can be driven per time step.

One important feature of the model is the product decay dynamics. We adopt the methodology from Rong et al. (2011) which uses a linear decay rate dependent on the environmental temperature and the time. The decay, for a time period of length τ and a temperature T, is given by the equation 1.

$$\Delta q(\tau, T) = -\kappa_0 \tau \cdot \exp\left[\frac{-H}{T}\right] \tag{1}$$

The constants H and $-\kappa_0$ depend on the fruit type. In the agent model, each agent has a temperature for storing the products which is applied for as long as the agent is in control of the product. Then, the product wastage along the chain depends on the temperatures at which each agent stores the product and the storage duration. If temperature were stable along the chain the decay rate would be always the same, however, the model incorporates more realistic random variability of the temperature, especially in the transporter agents, as these are the ones that are more susceptible to experience variations in the storage temperature of the product. Hence, transporters might store the product at the standard expected temperature or at a random temperature, given a user-defined parameter (more details in section 4.2).

Another key aspect of the model is the traceability information flow. This behaviour intends to replicate the dynamics of the supply chain when there is a traceability system in place that enables information sharing among the agents. We model this situation by incorporating two components that represent delays and noise in the traceability information. The delay is defined in time units and represents how long it takes for an agent to have access to RSL information after receiving the product. Likewise, the noise is a random variable that enables the simulation of information accuracy. The noise is normalised to 1 and follows a normal distribution whose standard deviation can be defined for each specific scenario.

4 SIMULATION STUDY

In this section we present the simulated scenarios studied with the model described in the previous section and present the main results of the study.

4.1 Scenarios

We are interested in the effect of information sharing in the total wastage along the chain. So we identify two key scenarios, as follows:

- *Low traceability*: This is the case where real-time information about temperature and storage time for each product was not available. Hence, the distributor does not have accurate nor timely information about the residual shelf-life of the product.
- *High traceability*: In this case, the traceability system is in place and the distributor can base its inventory management policies on complete, accurate, and updated residual shelf-life information of the product.

4.2 Modelling Parameters

The agent-based model is parameterised to replicate the case study supply chain. The Spanish soft citrus season runs from 1st September to 31st July (302 days). Over this period, 7.5 million cases of fruit enter the supply chain at the growers end. Cases are transported in trucks from Spain to the UK. The desired storage temperature throughout is 3°C. Information in the traceability system is shared in parallel to the flow of products.

There are seven agents as detailed in Table 1. In this case there is only one single retailer and a distributor agent as we are not studying differences among these actors. However, we incorporate two growers and their corresponding international transporters to account for the differences at these stages of the chain as they provide the information inputs that the distributor uses for dispatching the product to retailer.

Table 1: Agents and Simulation Parameters.

Agent	Parameter Name	Value	Units
Grower 1	default storage time	2	time steps
	production capacity	12	product cases
Grower 2	default storage time	4	time steps
	production capacity	12	product cases
Transporter 1 (International)	prob temperature	0.9	
	distance	1200	miles
	velocity	60	miles/hour
Transporter 2 (International)	prob temperature	0.5	
	distance	1200	miles
	velocity	60	miles/hour
Transporter 3 (UK)	prob temperature	0.9	
	distance	80	miles
	velocity	40	miles/hour
Distributor	processing time	1	time steps
Retailer	max shelf time	1	time steps

All agents have a standard temperature parameter set to 3°C. Grower agents have starting stock, a default storage time (representing the minimum time required for processing), and the production capacity. One time period represents a day. Transporter agents have distance (in miles) and velocity parameters (miles per hour). Transports also have a temperature probability(`prob_temperature`), which indicates the "reliability" of that agent in storing fruit at the desired temperature. For example, an agent with a `prob_temperature=0.9` has the standard temperature 90% of the time and 10% of the time a randomly selected temperature between 3°C and 10°C. One informant in the case study estimated that sometimes over 40% of readings from trucks can be erroneous. The distributor has a processing time parameter to cover the time required for quality control and handling. Finally, the retailer has a maximum shelf time product that represents their stock turns.

For the fruit decay model, E_a and $-\kappa_0$ constants were obtained from data provided by Snart et al. (2006) with a life of 21 days at 8°C and 56 days at 3°C. R is the universal gas constant.

$$H = \frac{-E_a}{R} = 15200 \quad \textit{and} \quad \kappa_0 = 1.55 \times 10^{22} \tag{2}$$

These parameters remain stable for the two scenarios analysed. To consider the variations between both scenarios, we used the traceability information parameters `info_delay` and `info_noise`. As we have products with two different origins (grower 1 and 2), in the *Low Traceability* scenario, products originated from grower 1 were set to an `info_delay` sufficiently high that information would not be shared until after products had decayed; for grower 2, the delay was set to 1 day. In the latter case, the standard deviation of the noise was set to 0.02, meaning that reported product RSL would randomly be +/- 2% of its actual RSL. This value was obtained after model calibration to case study

data. For the *High Traceability* scenario, both parameters were set to 0 for all products to simulate perfect information sharing.

We ran the model 100 times for each scenario. A burn-in period of 200 time steps allowed for transients to decay and for the retailers and distributors to build up some starting stock.

4.3 Model Validation

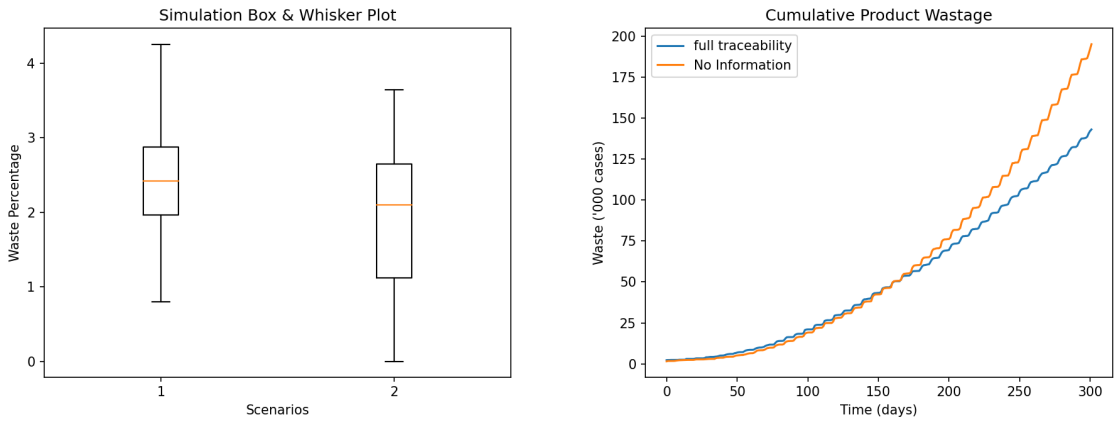
Granular empirical data to validate every aspect of the model was not available as this information is not shared transparently within the supply chain. However, given that the simulation parameters were derived from the case study data, it is sufficient in this case to ensure that the waste levels in the model match those from the case study. The waste levels of the simulation were fine tuned to match the figures from the empirical data by adjusting the rate of consumption (or demand) in the model to ensure products flowed through the supply chain. This demand rate was not available from the case study data as the retailer considered it sensitive information.

Table 2: Validation of Results to Case Data (given assumption of constant demand)

Data	Simulation Result	Empirical Data
High Traceability Waste	1.87%	1.86%
Low Traceability Waste	2.56%	2.54%

4.4 Simulation Results

In this section we present the results from the study. Table 2 shows the that the waste levels in the simulation very closely match those of the case study, with a maximum percentage error of 0.8%. In the Figures 2, 3 and 4 the time = 0 is 200 days after the start of the simulation due to the burn-in period. The results are the average after 100 runs of the simulation. The simulation configuration and resulting data are available in GitHub[1].



(a) Boxplot Comparison of Scenario 1 (Low Traceability) and Scenario 2 (High Traceability) - Aggregation of 100 Runs

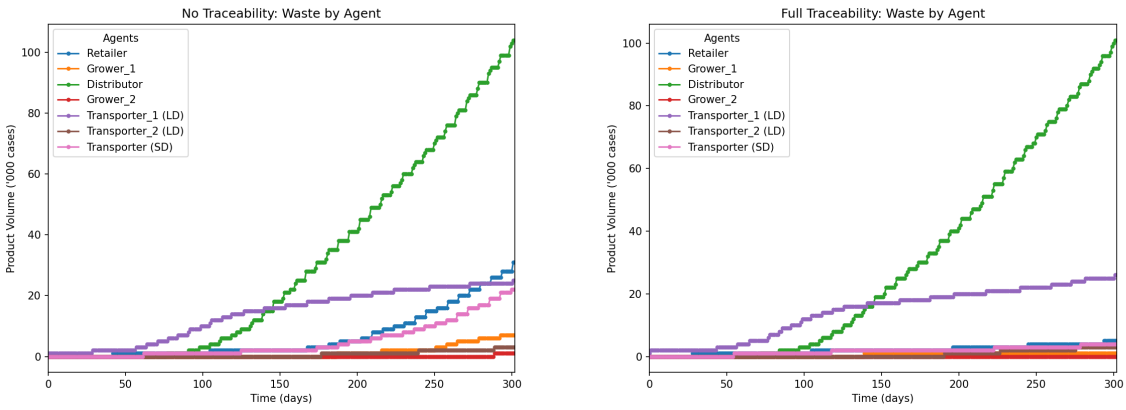
(b) Cumulative Waste

Figure 2: Cumulative Plots (all agents)

5 DISCUSSION

In this section of the paper, we discuss the results presented above and their implications. We begin with the main findings of the study by comparing the two scenarios. We then discuss the economic and environmental implications of these findings. Finally, we consider the limitations to this simulation.

[1] https://github.com/mperhez/scvis-plots/tree/master/sw21



(a) Low Traceability Case

(b) High Traceability Case

Figure 3: Waste by Agent

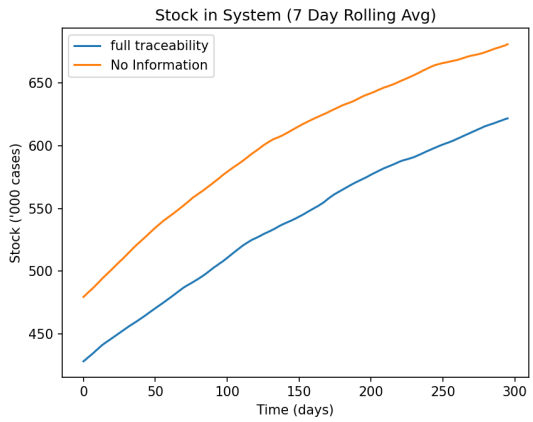


Figure 4: Stock Level in the Supply Chain

5.1 Traceability System Analysis

Based on the results of this simulation experiment, we develop four key insights regarding the relationship between traceability systems and food waste. First, we are able to observe where in the supply chain the food waste reductions occur when a traceability system is implemented. The reduction can be explained as the product traceability enables complete, timely and accurate information sharing along the chain, that in turns makes it possible to implement a *"First to Decay, First Out"* inventory management policy that has effect in the food waste within the observed scenarios.

Secondly, in figure 3b we observe that the waste level in the two scenarios is closely matched in the first half of the season, then the results diverge in the second half. Figure 3 suggests that this acceleration in waste level is caused by an increase in waste at the retailer and their logistics provider (`TransporterSD`). Without traceability information, the error in the guessing for inventory management builds up throughout the season. The downstream end of the supply chain receives the product at its oldest and therefore its highest perishability risk. Given this, the retailer and their transporter would be expected to see the greatest reduction in food waste and this is supported by the simulation experiment.

Next, we find it interesting to consider this experiment against the body of qualitative evidence that retailers are usually the actor that forces the development of a supply chain traceability system (Aung and Chang 2014). Our simulation study provides clear evidence for the reasons behind this trend; in terms of food waste, retailers see the greatest gains from implementing the traceability system. The distributor estimates that each wasted case costs £15.40 to the retailer, so a reduction from 38,000 cases

to 4,000 cases in a season is a cost saving of £520,000 per season for the retailer alone. Extending this cost-per-case-analysis, the results of the simulation study suggest the growers and international logistics providers will have little motivation to invest in the traceability system. The distributor sees an 8% reduction in waste, so will be somewhat more motivated that the growers to invest in the system.

Finally, when making the comparison between Figures 2 and 4, we observe that stock levels have reduced as well as waste levels. This indicates there may be relationship between stock levels and product waste. This matches expectations based on common sense. Higher stock increases the product volume that is likely to expire while in stock. More investigation is needed to understand the nature relationship between stock levels and waste at each agent in the supply chain. Further research is planned in this area.

5.2 Economic, Environmental, and Social Impact

Wang and Li (2006) called for more research into the "value" of supply chain traceability and data for traceability beyond the so-called "traditional" metrics. This simulation experiment indicates a number of fruitful avenues for further exploration along the triple-bottom-line perspective. Firstly, in the previous sub-section we considered the economic value of supply chain traceability in terms of cost of a wasted case. We noted that the economic benefits of traceability accrue closest to the consumers. Turning to the environmental impact of product waste, consider that product wasted at the retailer in the UK has already resulted in carbon emissions relating to its transport and storage up to that point. Therefore, the environmental impact of food waste is lower in the earlier stages of the supply chain. Traceability data could be applied to ensure that waste in the downstream supply chain is reduced, for example, by setting a residual shelf-life threshold.

Thirdly, considering social impact of product waste, fruit growing has historically been associated with poor labour practices. The production of food that eventually goes to waste could be considered to be incurring an ethical cost. This simulation study provides the insight that supply chain traceability can provide social value by reducing over-production to account for food waste and therefore the amount of labour required.

5.3 Limitations of the Study

This study presents several limitations that will be addressed in future works. The model does not incorporate a separate component for representing product waste due to handling. Although the case study shows this figures are low, incorporating this component will enable to study better the effect of traceability system performance.

The wastage in the system is already a very small amount that even small variations will look high, as a result the figures that we obtain, for the standard deviation of the relative waste among runs, look high. Although these values are consistent, we plan to incorporate new empirical data so the calibration of the model can be improved.

We have assumed a constant level of supply and demand throughout the season, however this is not an accurate reflection of reality. Both production of citrus fruit and demand follow cyclical patterns throughout the season. We plan to incorporate further empirical data to simulate these trends.

6 CONCLUSIONS AND FUTURE WORK

In this paper we have presented the development of a model to quantify the relationship between supply chain traceability information sharing and food waste. The model was trialled and validated using empirical data obtained from a European citrus fruit supply chain. We demonstrated the model was able to replicate the food waste seen in the case study data to an accuracy of better than 1%. The simulation experiment proved its value by presenting four insights into supply chain traceability systems:

- Supply chain traceability systems can reduce the food waste from perishability
- The retailer and final stage logistics provider see the greatest reduction in food waste
- The retailer is likely to be most motivated to invest in supply chain traceability, whilst the growers least so, as the reduction in wastage is greater further downstream in the supply chain
- Higher stocks are associate with a higher wastage level

We further considered how the simulation experiment provided insights into the economic, environmental, and social value of a supply chain traceability system. Finally, we point to some possible extensions to this model which the authors consider it valuable to explore:

1. **Information Quality** - Miller (1996) noted 10 dimensions of information quality. The simulation model could be extended to assess the impact of changes in these dimensions of quality in traceability information on the level of food waste. For example, by reducing the "completeness" of the information.
2. **Further empirical validation** - As noted in the discussion of the limitations of this study, the collection of more empirical data from citrus fruit supply chains would enable the validation of the model at a more granular level.
3. **Application to other supply chains** - A further extension of the simulation model would be the application to other produce supply chains by changing the parameterisation. For example, the length of the season, the perishability of the produce, and the locations of growers. Through this, more generalisable conclusions can be obtained.

ACKNOWLEDGEMENTS

The authors would like to thank AMT Fresh and Dr. Mukesh Kumar for their support in the case study for this project. We would also like to thank GS1 for their ongoing support in this field of research.

REFERENCES

Aung, M. M., and Y. S. Chang. 2014. "Traceability in a food supply chain: Safety and quality perspectives". *Food Control* 39:172 – 184.

Barratt, M., and A. Oke. 2007. "Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective". *Journal of Operations Management* 25:1217–1233.

Bartlett, P. A. et al. 2007, August. "Improving supply chain performance through improved visibility". *The International Journal of Logistics Management* 18:294–313.

Basingab, M. 2019. "Investigating the adoption of Internet of Things technology using agent-based simulation". *Proceedings of the International Conference on Industrial Engineering and Operations Management* 2019 (MAR): 1539–1550.

Cannella, S. et al. 2019. "An exploratory study of risk aversion in supply chain dynamics via human experiment and agent-based simulation". *International Journal of Production Research* 57 (4): 985–999.

Chang, X. et al. 2016. "Agent-based simulation of pricing strategy for agri-products considering customer preference". *International Journal of Production Research* 54 (13): 3777–3795.

Chen, X. et al. 2019. "Optimal pricing strategy for the perishable food supply chain". *International Journal of Production Research* 57 (9): 2755–2768.

de Moraes, C. C. et al. 2020. "Retail food waste: mapping causes and reduction practices". *Journal of Cleaner Production* 256:120124.

Gardner, T. A. et al. 2019, September. "Transparency and sustainability in global commodity supply chains". *World Development* 121:163–177.

Griffin, M. et al. 2009, December. "An analysis of a community food waste stream". *Agriculture and Human Values* 26:67–81.

Grillo, H. et al. 2017. "Mathematical modelling of the order-promising process for fruit supply chains considering the perishability and subtypes of products". *Applied Mathematical Modelling* 49:255 – 278.

Gustavsson, J. et al. 2011. "Global food losses and foodwaste". Technical report, FAO.

Herrera, M. et al. 2020. "Multi-agent systems and complex networks: Review and applications in systems engineering". *Processes* 8 (3): 1–29.

Kaipia, R. et al. 2013, April. "Creating sustainable fresh food supply chains through waste reduction". *International Journal of Physical Distribution and Logistics Management* 43:262–276.

Liljestrand, K. 2017. "Logistics Solutions for Reducing Food Waste". *International Journal of Physical Distribution & Logistics Management* 47:318–339.

Miller, H. 1996. "THE MULTIPLE DIMENSIONS OF INFORMATION QUALITY". *Information Systems Management* 13 (2): 79–82.

Morone, P. et al. 2018, June. "Does food sharing lead to food waste reduction? An experimental analysis to assess challenges and opportunities of a new consumption model". *Journal of Cleaner Production* 185:749–760.

Olsen, P., and M. Borit. 2013. "How to define traceability". *Trends in Food Science & Technology* 29 (2): 142 – 150.

Rong, A. et al. 2011. "An optimization approach for managing fresh food quality throughout the supply chain". *International Journal of Production Economics* 131 (1): 421 – 429. Innsbruck 2008.

Snart, J. et al. 2006. "Oranges: Safe Methods to Store, Preserve, and Enjoy". Technical report, University of California: Division of Agriculture and Natural Resources.

Spada, A. et al. 2018. "The influence of shelf life on food waste: A model-based approach by empirical market evidence". *Journal of Cleaner Production* 172:3410 – 3414.

Tonini, D. et al. 2018, June. "Environmental impacts of food waste: Learnings and challenges from a case study on UK". *Waste Management* 76:744–766.

Villena, Veronica and Gioia, Dennis 2020. "A More Sustainable Supply Chain". Available at https://hbr.org/2020/03/a-more-sustainable-supply-chain Accessed 31 August 2020.

Wang, X., and D. Li. 2006. "Value Added on Food Traceability: a Supply Chain Management Approach". In *2006 IEEE International Conference on Service Operations and Logistics, and Informatics*, 493–498.

Wang, X., and D. Li. 2012. "A dynamic product quality evaluation based pricing model for perishable food supply chains". *Omega* 40 (6): 906 – 917. Special Issue on Forecasting in Management Science.

WRAP 2017. "Estimates of Food Surplus and Waste Arisings in the UK-1 Estimates of Food Surplus and Waste Arisings in the UK". Technical report, The Waste and Resources Action Programme.

Yu, M., and A. Nagurney. 2013. "Competitive Food Supply Chain Networks with Application to Fresh Produce". *European Journal of Operational Research* 22:273–282.

Zhu, Z. et al. 2018. "Recent advances and opportunities in sustainable food supply chain: a model-oriented review". *International Journal of Production Research* 56 (17): 5700–5722.

AUTHOR BIOGRAPHIES

ROB GLEW received a MEng in manufacturing engineering and management from the University of Cambridge in 2019. He is currently a Doctoral Researcher at the Institute for Manufacturing, University of Cambridge, where his research focuses on the strategic value of traceability systems in food and medical supply chains.

MARCO PEREZ HERNANDEZ is Research Associate at the Institute for Manufacturing of the University of Cambridge. He received his PhD in Computer Science from the University of Leicester in 2018. His research interests include the use of multi agent systems and services architectures to engineer complex industrial systems.

DUNCAN MCFARLANE is Professor of Industrial Information Engineering at the Cambridge University Engineering Department, and head of the Distributed Information & Automation Laboratory within the Institute for Manufacturing. He has been involved in the design and operation of industrial automation and information systems for twenty years.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.016

IS FLIPPED CLASSROOM ENOUGH? TEACHING SIMULATION USING IN-CLASS FLIP MODEL

Fahim Ahmed

Department of Management science
University of Strathclyde
Glasgow, G4 0QU
United Kingdom
Fahim.ahmed@starth.ac.uk

ABSTRACT

The teaching of the simulation courses in Higher Education Institution (HEI) are mostly guided by the traditional method of lectures and lab tutorials. At some places, this style of teaching has been mixed with flipped-classroom where courses are delivered in the form of blended or distance learning environment. Despite advantages of the flipped-classroom, one of the key issue is the time students are expected to spend before the class. In our experience, students usually find it difficult to spend time preparing in advance of the class. We therefore propose to use a variant of flipped-classroom model called "in-class flip" for teaching simulation in HEI. A study with the second year undergraduate class on simulation is reported using in-class flip model and results are discussed. Initial findings from the experiment shows higher level of appreciation and satisfaction from the students compared to the traditional mode of teaching.

Keywords: In-class Flip, Flipped classroom, Simulation education, Higher Education Institution

1 INTRODUCTION

In this paper, we propose the adaptation of "in-class flip" model (Gonzalez, 2014) for teaching simulation in HEI which is a variant of the flipped-classroom model but significantly different in practice. To the best of our knowledge, the in-class flip model has not been used for teaching simulation (or similar courses) at HEI. We even struggle to find studies reporting its use in other broader disciplines within HEI. The key advantage of this model is that the students can learn at their own pace and within the class time which saves them from spending time on pre-class activities at home (which is a normal practice in a flipped classroom model).We present an experiment based on in-class flip model for a second year UG class and report the outcomes.

The contribution of this paper is to highlight the importance of in-class flip model that could bring a higher satisfaction and engagement level for students. In particular, this paper advocates significance of in-class flip model in teaching simulation (or courses containing similar structure) and also provides an opportunity to contribute further research in pedagogical innovations within HEI.

The rest of the paper is structured as follows:

Section 2: In-class flip vs flipped-classroom
Section 3: Adaptation of in-class flip model
Section 4: In-class flip model in practice
Section 5: Results from the experiment
Section 6: Conclusion and future work

2 IN-CLASS FLIP VS FLIPPED CLASSROOM

The idea of the flipped classroom roughly dates back to 1990s when Harvard Professor Eric Mazur used this technique to ask students prepare in advance of attending the class **(**Mazur, 1997; Crouch and Mazur, 2001). He called his model "just in time" teaching. Salman Khan in TED talk **(**Khan, 2011) was probably the next most prominent work to have this approach gaining fame in recent times. Salman talks about the significant use of videos in his organisation, Khan Academy (Khan, 2019). Some other works around this time from **(**Bergmann and Sams, 2012) help this model to gain formal recognition and probably considered as the pioneering work in this area. Bergman and Sams were also the founder of flipped learning network among others (flippedlearning, 2019). There are several studies that discuss flipped classroom and it is hard to reflect on all this **(**see for e.g. Lage et al. ,2000; Chen et al., 2017; Lui et al., 2017; Schmidt and Ralph, 2016; DeLozier and Rhodes, 2017; Karabulut-Ilgu et al. , 2018)

The main requirement of this model expect students to prepare, read or reflect some material assigned to them in advance of the class. The teacher can then use this pre-class activity as a basis to develop an effective discussion in the class. The pre-class content does not need to be a video although video is generally considered to be the most important part of this model which students can watch in advance. The idea of any sort of visual enhancement **(**over text based illustration) has been shown to increase student learning in many early studies **(**Mayer and Gallini, 1990).

In all of the above work, there is some portion of self-study outside the classroom that students are expected to complete. We believe that with the changing life-style, work-life issues, and engagement with multiple roles, students find it difficult to spend any significant time outside their regular contact time in class. Our other argument is despite the flexibility for students to complete certain preparation before class, there is less motivation to complete them as no formal credit hours could be assigned to the amount of time they spent before the class. We report some studies here that evaluates the effectiveness of flipped classroom when teaching courses at HEI.

In one of the studies, Bergfjord and Heggernes (2016) reports that during the teaching of a management science course in a flipped-class room setting, most videos were watched before the exam rather than before the lectures. Amresh et al. (2013) used flipped-classroom for teaching computer science courses but reports on various pitfalls where how-to-do type videos were used. One of the findings was that the students find this method overwhelming **(**probably when learning unsupervised). As mentioned in Bergfjord and Heggernes (2016), technology can be a bottleneck to adapt this model especially when lecturers have to consider all possible technical aspects of students learning remotely including management of Virtual Learning Environment (VLE), internet, software support etc. Forsey et al. (2013) believe that flipped-classroom could devalue the significance of the lecture, organisation and the culture of the university. Some studies in engineering even mentioned that students had negative perceptions toward the course and felt unprepared for the exams because they had to manage their own learning. The final grades for traditional model of learning were even higher than a flipped-classroom in this study **(**Karabulut-Ilgu et al., 2018).

By no means we assert that this is an exhaustive list of evaluation of flipped class room model in HEI, but at least it gives the basis of our motivation for using in-class flip model. In our own experience, majority of students do not watch the pre-loaded videos on VLE and leave it till few days before the exam. This even takes more time to go through the content again and revise the topics of videos during the class time. One of the reason of less student engagement outside the class **(**especially when teaching simulation course) might be that student need a quick response to the query while they watch a video on various topics. In such cases, we feel that there is a need of supervised learning within the class time **(**to answer any prompt questions) rather than asking students to prepare the content before the class.

2.1 In-class flip model

The in-class flip model appeared over the internet few years back when Gonzalez (2014) wrote about this model of teaching in one of her blog at educational learning website (Edutopia, 2019**).** To the best of our knowledge, there is no prior work that introduces this variant of the flipped classroom. This model is shown in figure 1 where different works stations are introduced to be used in the class.

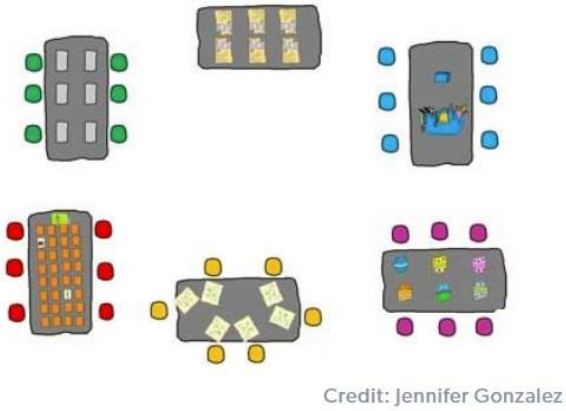


Figure 1: The in-class flip model **(**Gonzalez, 2014)

Each stations will have an activity assigned to it and a group of students will complete the activity on these stations in a pre-defined sequence. This means that students will need to keep revolving at different stations through the class time. Stations can have computer based activity **(**e.g. video, audio etc.) or a non-computer based activity **(**e.g. discussion, drawing, cards matching etc.) The teacher will be available most of the time supervising each student and answering any questions. Stations sequence are assigned in a way that students can start and progress with the activities on their own. The main station is where the video of the content is available to watch. This is recorded by the same teacher delivering that class **(**Gonzalez, 2014). Later on, Barnes and Gonzalez (2015) wrote about in-class flip where the idea of bringing the lecture part in the class room using different stations was used. Ramirez (2017, 2019) uses various station configurations adapted from Gonzalez (2014). Tucker (2016) used similar approach using station rotations. Ramirez (2018) used this model for teaching grade 7 English class.

3 MODIFICATION OF IN-CLASS FLIP MODEL

All of the above studies were proposed and demonstrated for school learning environment where the flexibility of moving groups of students over different station is manageable. Mostly, these works are based on the idea of various stations in the class **(**viewing station, independent station, feedback station, etc.) which requires students to move from one station to other in a pre-defined sequence **(**Gonzalez, 2014; Ramirez, 2018) .This setup makes it unsuitable for teaching in HEI where labs and lecture halls usually do not have that much capacity to create several rotating stations. Moreover, teaching of subjects like simulation is not similar to the teaching of history or language classes where activities can easily be planned for different stations **(**e.g. vocabulary exercise using flash cards etc.). In HEI teaching, the disruption from ever-rotating stations **(**students) would have significant effect on student experience as class teaching time for a particular subject is relatively small compare to school level teaching.

To adapt this model for teaching large classes at higher education institute, we propose the idea of integrated-activity station in a computer based classrooms. This modification of in-class flip is shown in figure 2

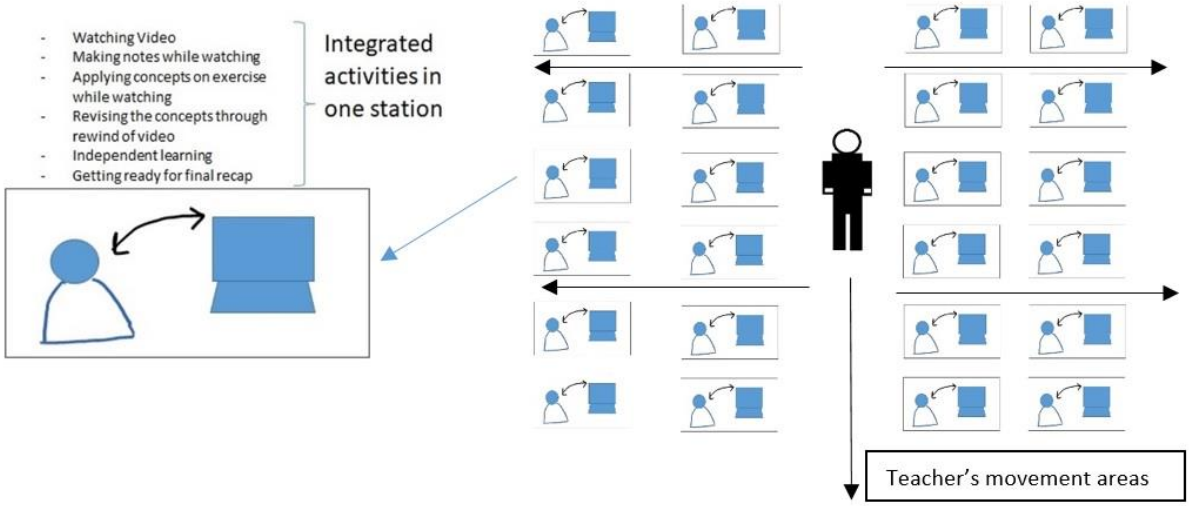


Figure 2: Modification of in-class flip model for teaching in HEI

In this model, students do all the activities at the same time **(**on individual basis) and do not revolve physically on different stations. All activities are integrated on single station and students are expected to cover them progressively **(**some sample activities are shown in figure 2). This helps to minimise the disruption in a large class and is perfectly suitable where individual learning needs to be enforced. A layout of such a classroom along with the teacher's movement area is also shown in figure 2. The activities are designed in such way that students are involved in a self-directed but supervised learning environment while teacher is available most of the time to answer any queries.

4 IN-CLASS FLIP MODEL IN PRACTICE

Next, we present the experiment of the modified in-class flip model conducted for second year UG simulation module. The simulation package used was Simul8. The course was conducted as a half semester module **(**part of a full module) at the department of management science, University of Strathclyde. Some background information is as follows:

Total number of students = 127
International Students = 35
Duration = 4 Weeks

Topics Covered= Basics of simulation modelling, application of different concepts in Simul8, simulation model verification/validation, data modelling in Simul8, conceptual modelling

4.1 Arrangements for the class

A computer lab was booked on request for four hours per week spread over two days. All students were asked to bring their headphones and this instruction was re-iterated many times through class notice page and reminder messages the day before the class.

Videos for each topic were recorded in various small parts **(**3-8 mins) with some videos longer in length as continuation was necessary **(**10-12 mins). The style of the video recording was kept similar to the scenario as if students are listening to the live lecture. All video lectures were screen-recorded **(**with audio) as it was not necessary for lecturer to be present in the video. This also reduced the time

and complexity of the videos required for recording. An exercise related to each sub-video was designed which students could attempt while watching the video or at the end of video (by going back and forth the important parts of the video). Students were expected to take notes while they watch the video and carry on the task assigned to them once they are ready to do so.

Lecturer was walking around in the class room, passing by the students throughout the entire time of video learning and during any individual exercise assigned to them (which lasts to around roughly an hour in total). During this time, lecturer answers any specific questions student may have (whether related to the content or any technical issue e.g. location of the video, sequence of the videos and exercise etc.). Students were expected to follow a complete lesson plan uploaded in advance on VLE before the class. A sample plan is shown in figure 3.

Lesson plan for simulation class

1- Watch videos 1 and 2 (10:00 am)
2- Start exercise 1 (10:15 am)
3- Watch video 3 (10:30 am)
4-Start exercise 2 (10:50 am) - optional
5- Join the discussion group (11:00 am)- compulsory
6-Get ready for reflections and conclusion (11:30)

Figure 3: A two-hour lesson plan for teaching simulation with in-class flip model

It is important to understand the sequence of events in this plan. The first three steps are compulsory and students are expected to start them by the time indicated. However, keeping in view the different needs and learning pace of the individuals, some students would want to revise the video more than once in order to catch up with the rest of the class. To smooth out this variation, an optional activity was inserted after first three activities and only those students were expected to work with it who already have completed all the previous steps. The next activity (5) is again compulsory and requires everyone to join together in the form of a discussion group and complete the given task (this can be any formation of group; random, pre-assigned etc.). Those students who could not start on the optional activity also had to join the group work now. This creates an opportunity for every student to utilize their time effectively by not sitting and waiting for an activity to start (as fast learners can always do an optional activity in the middle). This optional activity needed to be designed in a way that it does not affect the balance of the class learning objectives and the group discussion. Finally, the lecturer sums up the class with important reflections and take away points from the session.

It is worth mentioning that in few instances, a large computer lab was not available and hence the students were split in two different labs and similar model was applied which ran smoothly without any disruption. This was possible as the only time where all students were needed together in one place was the conclusion or reflection part. As no station was required at this stage, students were asked to gather compactly in one place without acquiring one-to-one stations. The only requirement was that the two computer labs must not be booked too far from each other so lecturer could easily move between them. At times, a help from an extra tutor (usually postgraduate research student) was taken so one supervisor is present at all times with the students but the main lecturer was still supervising teaching in both rooms and was consulted anytime if required during his periodic rounds in both labs.

5 FINDINGS OF THE EXPERIMENT

The results of experiment were recorded through questionnaire at the end of the course using paper-pencil method (for a higher response rate). There were 102 students present at the time of questionnaire reporting. Three questions were asked from the students stated below:

Q1) Is this model better than the traditional lecture style?
Q2) Did this model increase your satisfaction level?
Q3) Any other comments/suggestions

The results are shown in figure 4a and 4b.

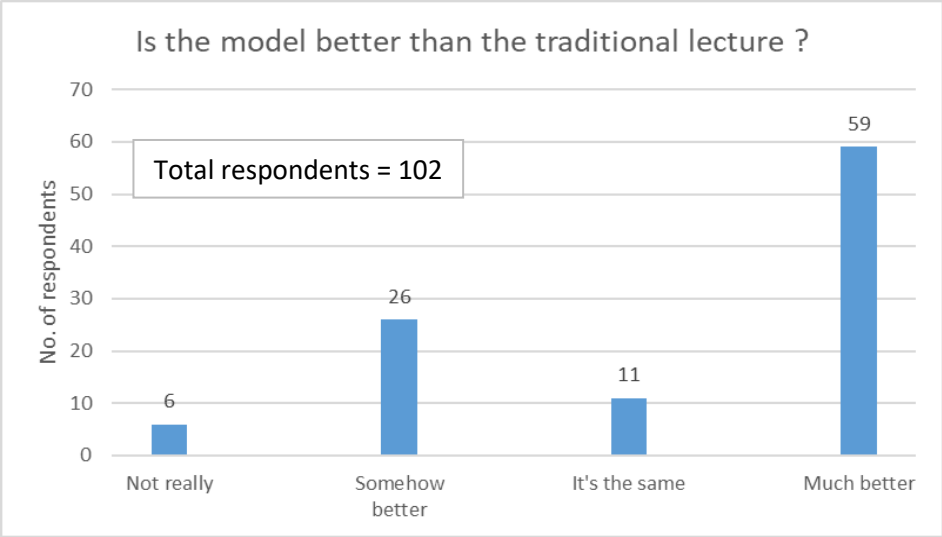


Figure 4a and 4b: Responses of the questionnaire for in-class flip experiment

On the first question; 59 students responded in favour of this model while 26 students feel that this was better up to some extent. On the second question; 58 students felt that their satisfaction level has gone up to some extent while 42 students reported that this model has greatly increases their satisfaction level. Some interesting open-ended answers were noted and while we cannot share all responses due to limited space, we share snapshots of few critical ones in figure 5.

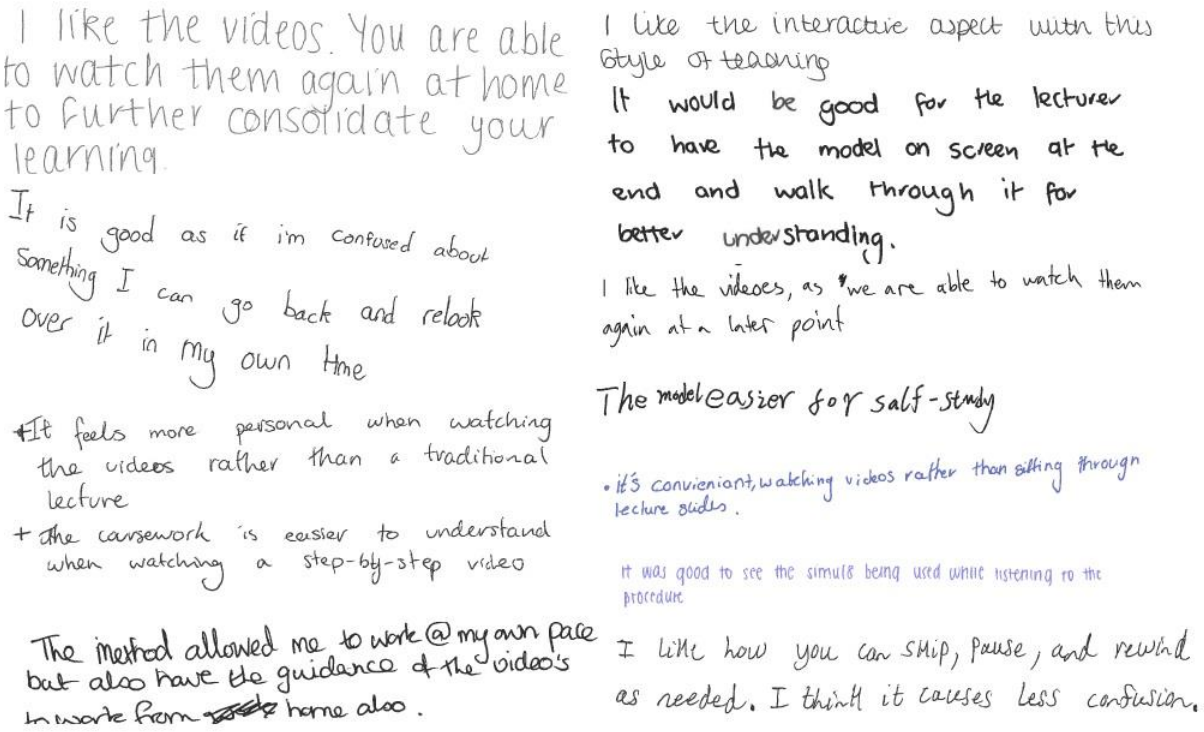


Figure 5: sample of open-ended responses for in-class flip experiment

In total around 85% students reported that the model is better **(**significantly or to some extent) from the traditional lecture style teaching. Similarly, almost all student think that their satisfaction level has raised **(**either marginally or significantly) when learning through this model in the class.

6 CONCLUSION AND FUTURE WORK

In this study, we present an adaptation of in-class flip model to be used for teaching a simulation course at HEI. The initial results seems promising but there is a need of more detailed experiment on even larger classes **(**with different teaching facilities, various room layout, lecture rooms, available IT support etc.) to generalise the results. While there is clear appreciation of the model from majority of the students, it is hard to identify any relationship between acceptance level and satisfaction level of this model as these two responses might not be related strongly. For e.g., majority of the students see this model as "much better" but also think that their satisfaction level has only increased marginally. It would be interesting to investigate as what factors make this model appealing and what factors make this model more satisfying. Nevertheless, there is clear evidence of high inclination of students towards this model as witnessed by the results.

An additional advantage of this approach is shown to be an effective utilisation of the teaching resource where a lecturer can cover much broader area during the teaching **(**in some cases, two separate rooms) which would otherwise not be possible without either duplicating the class delivery or having another lecturer for the same teaching content.

There are some obvious limitations of this model which needs to be addressed. Firstly, this model may not be suitable for teaching non-simulation type courses outside the lab environment where all students would not have access to PCs. There is a real challenge to ask students to bring their own laptops, smart phones and tablets within lecture halls. However, with the growing embedded technology within VLEs (like planetstream etc.) and high speed internet, students seem very comfortable in watching videos as it does not require heavy resources to download and play them. It would be interesting to see the results of such an experiment which is planned as a future work. Accessibility issues might need addressing (in few cases) by providing a transcript in advance to the special need students, so they are not left behind during the activities. Video captions and subtitles can also be used but all of this needs to be planned within the available time and resources. There is always a risk of technology going wrong (headphones or internet not working, students not registered on VLE etc.) and back up must be sorted out. Finally, to avoid students watching the video passively, more interactive element needs to be developed while they watch the videos using technologies like H5P, online multiple choice questions etc. and should be embedded within VLE.

REFERENCES

Amresh A., Carberry A R, & Femiani J (2013, 2013). *Evaluating the effectiveness of flipped classrooms for teaching CS1*.

Barnes M D, & Gonzalez J (2015). *Hacking education: 10 quick fixes for every school*: Times 10 Publications.

Bergfjord O J, & Heggernes T (2016). Evaluation of a “flipped classroom” approach in management education. *Journal of University Teaching & Learning Practice, 13*(5), 17.

Bergmann J, & Sams A (2012). *Flip your classroom: Reach every student in every class every day*: International society for technology in education.

Chen F, Lui A M, & Martinelli S M (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical education, 51*(6), 585-597.

Crouch C H, & Mazur E (2001). Peer instruction: Ten years of experience and results. *American journal of physics, 69*(9), 970-977.

DeLozier S J, & Rhodes M G (2017). Flipped classrooms: a review of key ideas and recommendations for practice. *Educational psychology review, 29*(1), 141-151.

Edutopia (2019) https://www.edutopia.org/topic/flipped-classroom accessed 10th February 2020

Flippedlearning (2019) : https://flippedlearning.org/who-we-are/ accessed 21st November 2019

Forsey M, Low M, & Glance D (2013). Flipping the sociology classroom: Towards a practice of online pedagogy. *Journal of Sociology, 49*(4), 471-485.

Gonzalez J (2014). Modifying the flipped classroom: The" in-class" version. In: Edutopia. Retrieved from https://www.edutopia.org/blog/flipped-classroom-in-class-version-jennifer-gonzalez accessed 10 February 2020.

Karabulut-Ilgu A, Jaramillo Cherrez N, & Jahren C T (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology, 49*(3), 398-411. doi:10.1111/bjet.12548

Khan (2019) (https://www.khanacademy.org/ accessed 21st November 2019)

Khan, S. (Producer). (2011). *Let’s Use Video to Reinvent Education [Internet]*

Lage M J, Platt G J, & Treglia M (2000). Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *The Journal of Economic Education, 31*(1), 30-43. doi:10.1080/00220480009596759

Mayer R E, & Gallini J K (1990). When is an illustration worth ten thousand words? *Journal of educational psychology, 82*(4), 715.

Mazur E (1997). Peer Instruction: A User’s Manual. 1997 Upper Saddle River. *NJ Instruction: A User’s Manual Series in Educational Innovation*.

Ramirez M (2017). What's an in-class flip? Retrieved from http://martharamirez.com.co/blog/whats-an-in-class-flip/ accessed 10th February 2020.

Ramirez M (2018). In-Class Flip: Flipping a Literature Class for Student-Centered Learning. In *Innovations in Flipping the Language Classroom* (pp. 93-103): Springer.

Ramirez M **(**2019). What's an in-class flip ? **(**revisited). Retrieved from https://flippedlearning.org/in-flip/whats-an-in-class-flip-revisited/ accessed 10th February 2020.

Schmidt S. M. P, & Ralph D L **(**2016). The flipped classroom: A twist on teaching. *Contemporary Issues in Education Research (CIER)*, *9***(**1), 1-6.

Tucker C **(**2016). The in-Class Flip : The Flipped Classroom meets the station rotation model Retrieved from https://catlintucker.com/2016/01/inclassflip/ accessed 10th February 2020.

AUTHOR BIOGRAPHIES

FAHIM AHMED is a teaching fellow at the department of Management Science, University of Strathclyde. His areas of expertise are mainly in Discrete Event Simulation and more broadly in the use of simulation methodologies within service operations. He attained his doctorate from Loughborough University, UK, in 2015. His PhD thesis was focused on the adaptation of computer science methodologies in simulation conceptual modelling. His current research includes pedagogical innovations within university teaching for OR and Management science courses.
https://www.strath.ac.uk/staff/ahmedfahimdr/

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.017

WHEN SIMULATION BECOMES HUMAN CENTRIC ANALYTICS

Dr Christina J. Phillips

Liverpool John Moores University
Liverpool Business School, Redmonds Building, Brownlow Hill, Liverpool L3 5UG
c.j.phillips@ljmu.ac.uk

ABSTRACT

In the following we describe a recent development in the integration of analytics into human processes called Human Centric Analytics (HCA), and take a look at how simulation fits into the design process that lies at the heart of HCA. The article will outline what HCA is, and how it works, and how simulation can facilitate its development. Referring to literature from OR and Design we show how simulation can become a mutually designed controller, forming part of a cybernetic evolutionary process toward improved ends and increased knowledge. Simulation is capable of facilitating more than an end artefact for experimental design and analyses, as it can provide a focus for group agency and an opportunity to see into processes and problems in novel and creative ways.

Keywords: Analytics, Human Centred Design, Participation, Cybernetics

1 INTRODUCTION

Analytics is described as a dianoetic management system that enables management decisions via a fusion of technology, quantitative methods and decision science (Mortenson, Doherty and Robinson, 2015). On the analytics stack of descriptive, predictive and prescriptive methods (Davenport and Harris, 2007) simulation is used mainly in predictive tasks but can be useful across the stack. In the era of big data and digital twins, simulation has become an important part of a firms analytics capabilities but there are still many challenges to overcome (Fowler and Rose, 2004).

In the realm of healthcare, participative simulation has been used effectively to solve problems, implement lean, create knowledge and increase awareness (Robinson *et al.*, 2012; Kotiadis, Tako and Vasilakis, 2013; Pessôa *et al.*, 2015; Baril *et al.*, 2016; Lamé, Jouini and Stal-Le Cardinal, 2019), to a lesser extent the same thing has been observed in industry (Holweg and Bicheno, 2002; Abdulmalek and Rajgopal, 2007; Pool, Wijngaard and Van der Zee, 2011; Phillips and Nikolopoulos, 2019). Many researchers have noted that the process of performing the participative simulation can be as informative as the end artefact (Robinson, 2001; Holweg and Bicheno, 2002; Pool, Wijngaard and Van der Zee, 2011). This is where simulation becomes human centric and cybernetic as knowledge is created in the moment across a surface of becoming (Pickering, 2009). The process then becomes as important as the end product, and this means we need tools with which to study and perform it well (Mingers and Brocklesby, 1997; Ormerod, 1997; Yearworth and White, 2014).

2 HUMAN CENTRIC ANALYTICS

HCA is a design paradigm for analytics that involves the human users and producers of data within the design process as completely as possible (Phillips, 2019). It requires a cross disciplinary stance, which mixes methodologies and positivist and interpretive paradigms, to create a technique that bridges the technical and social domains. It is not proscriptive in the methods used since it is a paradigmatic stance that requires high human involvement, iterations of design, effective communication, and an understanding of context and multiple viewpoints. The simplicity of the idea belies the complexity of working with human actors, technologies and systems to create analytical artefacts, including appropriate data pipelines. The technique necessarily mixes methodologies and in this respect needs to maintain, what Lane and Oliva (1994) call ‘dynamic coherence’ as stakeholders are alternately exposed

to suggested changes that instigate instinctive reactions, or behaviour induced by possible causal structure, via the modelling process.

3 THE CYBERNETIC SYSTEM

The cybernetic view of design as modelling is explored by Maier et al. (2014) and Maier et al. (2012), viewing designing as a cybernetic system regulated by methods and process models. Cybernetics is taken from the Greek Kybernetes and means helmsman or steersman (Ashby, 1956). It is a paradigm that views systems as self-regulating via controllers, iterations and information flows. It has an almost hylozoist view of human-machine systems, ignoring boundaries and viewing the socio-technical in a holistic way. Studies in cybernetics have shown how one can get exceedingly complex and unpredictable behaviour from iterative interactions of relatively simple inputs (Beer, 1966; Pickering, 2009).

As simulators we know that to try to model every detail is, in general, both futile and unnecessary (Robinson, 2014). Frequently clients do not recognise this, and there is also a modelling process of problem simplification which to us is intuitive but which can provide unexpected insight for stakeholders into previously messy and intractable problem domains (Phillips and Nikolopoulos, 2019). Simulation models and the accompanying analyses that help to determine data and parameters provide a means of group derived agency which can communicate with others, be that the modeller/designer, the system itself or different functions and hierarchies within an organisation (Eckert, Maier and McMahon, 2005). From a cybernetic viewpoint when we model in a participative and iterative way we are adapting to create a future collaboratively at each step. The models have to be simple to achieve a general applicability, to achieve a complete system model we would have to copy the requisite variety of the system (Ashby, 1956) which would be overly detailed and too specific (Maier et al., 2012) to be useful.

Providing views, which work to gain insight, requires methods and models that are not overly prescriptive so that they are open to multiple interpretation, both as a group and individually, and the ambiguity of situations with high human involvement, and multiple viewpoints, requires simple and parsimonious simulation models (Robinson et al., 2014). As a cybernetic system, simulation modelling needs the ability to be reactive and reflexive in the face of change that comes about as part of the design process. HCA must be reflexive and dynamic or it cannot incorporate the creativity of the stakeholders as part of the design, since they are the subject matter experts it is their creativity and free moves in overcoming resistance to change that enable analytics which augment their work (Pickering, 1995; Phillips, 2019a). These moments are the bridging of existing human practice with novel technologies and data views, which in turn facilitate an evolution of knowledge and design. The design process brings about a certain level of unpredictability, which in a very cybernetic way should be embraced and accepted (Pickering, 2009).

4 MIXING METHODS

PartiSim (Tako and Kotiadis, 2015) fosters HCA as it uses the broad bounded, yet complete methodology of SSM combined with DES. Due to its multiple iterations of design with high human involvement, in as much of the process as possible, it is a natural vehicle for HCA. Phillips and Nikolopoulos (2019), used PartiSim in a manufacturing environment which had unexpected consequences due to the simple contextualised models which were needed to facilitate participation throughout. They provided visualisations and simple analysis to allow stakeholders to choose model parameters in an informed and inclusive way. This prompted improvements in forecasting as well as to the planning and scheduling process under study. Their end model was used by senior management to make major strategic decisions regarding the factory, something which has been lacking in OR applications (Lane, 2010), and the forecasting improvements helped the company make large inventory savings. In this case the participative simulation development had multiple consequences and these were enabled by the simple yet well-structured constitutive rules of PartiSim, which were robust in the face of change and easily flexed as necessary.

SSM has been used by many simulation modellers and others in mixed methods research that bridges the social and the technical (Lane and Oliva, 1994; Tako and Kotiadis, 2015; Small and

Wainwright, 2018; Lamé, Jouini and Stal-Le Cardinal, 2019). The SSM methodologies have a strong underlying philosophy and lend themselves to being decomposed into separate parts, which can be cherry picked for situational suitability (Mingers and Brocklesby, 1997). This also allows the methods to be kept simple yet flexible as the cybernetic view would suggest they need to be. Lamé, Jouini and Stal-Le Cardinal (2019) combine DES and SSM, but also found that ethnography provided a way to both study the system under consideration and to provide an objective perspective to the stakeholders involved in the study. They used Analyses I, II, and III, and root definitions from SSM (Checkland, 1999), Phillips (2019b) had root definitions predefined and found the CATWOE (Customers, Actors, Transformation process, Worldviews, Owners, Environment) to be most useful. Pessôa et al. (2015) mixed methods using cognitive mapping and DES to involve experts in simulation experiments, leading to problem solution and improvement. They did not use any of the rest of the SODA method usually associated with cognitive mapping (Eden, 1988).

5 CONCLUSION

When simulation modellers handle the social and political elements of participative simulation exercises, via mixed methodologies, there are often insights that go beyond the initial intention of the exercise. The simulation activity becomes a performative process and the models become facilitators of thought, knowledge creation, and creativity, encouraging ideation in a group setting. Until the end simulation is settled upon there is a continuous design process. This can foster lean and continuous improvement (Holweg and Bicheno, 2002; Abdulmalek and Rajgopal, 2007; Van der Zee, Pool and Wijngaard, 2008; Pool, Wijngaard and Van der Zee, 2011; Phillips and Nikolopoulos, 2019) by providing a focus of ideation, and a possibility to create an artefact derived via group agency.

Phillips (2019), used PartiSim to improve testing and scheduling in a pharmaceutical factory, but also used action research (Eden and Ackermann, 2018) and intervention theory (Argyris, 1970), to take the simulation from working with shop floor stakeholders to providing a decision tool for senior managers. These additional intervention techniques maintained a research framework to provide structural and ethical/social/political guidance as well as recoverable information from the many human centric interactions.

Techniques such as ethnomethodology (Franco and Greiffenhagen, 2018), ethnography (Lamé, Jouini and Stal-Le Cardinal, 2019), symbolic interaction theory (Gallant and Kleinman, 1983) and the mangle of practice (Pickering, 1995; Ormerod, 2014) can help us to see into the practice of simulation with high human involvement, moving forward theory around participative simulation and HCA.

Perhaps one way to overcome Fowler and Rose's (2004) last and most difficult ‘grand challenge’; that of persuading managers to engage more with simulation, is to help them see it as a vital part of the analytics stack. In particular, as a means to foster HCA that can not only increase knowledge and provide useful artefacts, but that can also shift a culture toward being more proactive and data curious (Phillips, 2019).

REFERENCES

Abdulmalek, F. A. and Rajgopal, J. (2007) ‘Analyzing the benefits of lean manufacturing and value stream mapping via simulation: A process sector case study’, *International Journal of Production Economics*, 107(1), pp. 223–236. doi: 10.1016/j.ijpe.2006.09.009.

Baril, C. *et al.* (2016) ‘Use of a discrete-event simulation in a Kaizen event: A case study in healthcare’, *European Journal of Operational Research*. Elsevier Ltd., 249(1), pp. 327–339. doi: 10.1016/j.ejor.2015.08.036.

Davenport, T. H. and Harris, J. G. (2007) *Competing on Analytics: The New Science of Winning*. 1st edn. Boston, MA: Harvard Business School Press. Available at: http://dl.acm.org/citation.cfm?id=1526366 (Accessed: 13 March 2017).

Fowler, J. W. and Rose, O. (2004) ‘Grand Challenges in Modeling and Simulation of Complex Manufacturing Systems’, *Simulation*, 80(9), pp. 469–476. doi: 10.1177/0037549704044324.

Franco, L. A. and Greiffenhagen, C. (2018) ‘Making OR practice visible: Using ethnomethodology to analyse facilitated modelling workshops’, *European Journal of Operational Research*. North-Holland, 265(2), pp. 673–684. doi: 10.1016/J.EJOR.2017.08.016.

Gallant, M. J. and Kleinman, S. (1983) 'Symbolic Interactionism vs. Ethnomethodology', *Symbolic Interaction*, 6(1), pp. 1–18.

Holweg, M. and Bicheno, J. (2002) 'Supply chain simulation - A tool for education, enhancement and endeavour', *International Journal of Production Economics*, 78(2), pp. 163–175. doi: 10.1016/S0925-5273(00)00171-7.

Kotiadis, K., Tako, A. A. and Vasilakis, C. (2013) 'A participative and facilitative conceptual modelling framework for discrete event simulation studies in healthcare', *Journal of the Operational Research Society*, 65(2), pp. 197–213. doi: 10.1057/jors.2012.176.

Lamé, G., Jouini, O. and Stal-Le Cardinal, J. (2019) 'Combining Soft Systems Methodology, ethnographic observation, and discrete-event simulation: A case study in cancer care', *Journal of the Operational Research Society*. Taylor & Francis, 0(0), pp. 1–18. doi: 10.1080/01605682.2019.1610339.

Lane, D. C. and Oliva, R. (1994) 'The Greater Whole: Towards a synthesis of SD and SSM', *1994 International System Dynamics Conference*, p. 134.

Mingers, J. and Brocklesby, J. (1997) 'Multimethodology: Towards a framework for mixing methodologies', *Omega*, 25(5), pp. 489–509. doi: 10.1016/S0305-0483(97)00018-2.

Mortenson, M. J., Doherty, N. F. and Robinson, S. (2015) 'Operational research from Taylorism to Terabytes: A research agenda for the analytics age', *European Journal of Operational Research*. Elsevier Ltd., 241(3), pp. 583–595. doi: 10.1016/j.ejor.2014.08.029.

Ormerod, R. J. (1997) 'The role of OR in shaping the future: smart bits, helpful ways and things that matter', *Journal of the Operational Research Society*, 48(11), pp. 1045–1056. doi: 10.1057/palgrave.jors.2600465.

Pessôa, L. A. M. *et al.* (2015) 'Integrating soft and hard operational research to improve surgical centre management at a university hospital', *European Journal of Operational Research*, 245(3), pp. 851–861. doi: 10.1016/j.ejor.2015.04.007.

Phillips, C. J. (2019) *Human Centric Analytics: Creating a data driven culture to manage uncertainty (under revision EJOR)*.

Phillips, C. J. and Nikolopoulos, K. (2019) 'Forecast quality improvement with Action Research: A success story at PharmaCo', *International Journal of Forecasting*. Elsevier B.V., 35(1), pp. 129–143. doi: 10.1016/j.ijforecast.2018.02.005.

Pickering, A. (2009) 'Beyond design: Cybernetics, biological computers and hylozoism', *Synthese*, 168(3), pp. 469–491. doi: 10.1007/s11229-008-9446-z.

Pool, A., Wijngaard, J. and Van der Zee, D.-J. (2011) 'Lean planning in the semi-process industry, a case study', *International Journal of Production Economics*. Elsevier, 131(1), pp. 194–203. doi: 10.1016/j.ijpe.2010.04.040.

Robinson, S. (2001) 'Soft with a hard centre: Discrete-event simulation in facilitation', *Journal of the Operational Research Society*, 52(8), pp. 905–915. doi: 10.1057/palgrave.jors.2601158.

Robinson, S. *et al.* (2012) 'SimLean: Utilising simulation in the implementation of lean in healthcare', *European Journal of Operational Research*. Elsevier B.V., 219(1), pp. 188–197. doi: 10.1016/j.ejor.2011.12.029.

Yearworth, M. and White, L. (2014) 'The non-codified use of problem structuring methods and the need for a generic constitutive definition', *European Journal of Operational Research*. Elsevier B.V., 237(3), pp. 932–945. doi: 10.1016/j.ejor.2014.02.015.

AUTHOR BIOGRAPHY

CHRISTINA J. PHILLIPS is a Senior Lecturer in Business Analytics at Liverpool Business School and a Visiting Researcher at both Leeds and Bangor Universities. She gained the Siemens & CBLESS studentship for an industrial research placement which gave her first-hand experience of participative modelling in a real world setting. Her previous experience, first as a manager, then as a designer and artist came into use more than expected, culminating in observations on the design paradigm of Human Centric Analytics. https://www.linkedin.com/in/christinajphillips/

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.018

AN EXPLORATORY STUDY ON THE USES OF "WRONG" DISCRETE EVENT SIMULATION MODELS IN PRACTICE

Mr. Naoum Tsioptsias

Loughborough University
Epinal Way, Loughborough LE11 3TU
n.tsioptsias@lboro.ac.uk

Dr. Antuela Tako

Loughborough University
Epinal Way, Loughborough LE11 3TU
a.takou@lboro.ac.uk

Professor Stewart Robinson

Loughborough University
Epinal Way, Loughborough LE11 3TU
s.l.robinson@lboro.co.uk

ABSTRACT

It is generally accepted that no model is perfect. Literature has focused mainly on presenting successful simulation studies. This paper examines the extent to which "wrong" models are used in practice and how modellers and clients deal with them. We interview 22 discrete-event simulation modellers and analyse 54 real-life stories to identify possible uses of "wrong" models. We also focus our analysis on the model's level of detail as a possible reason for considering a model wrong, and analyse how this can affect acceptance and use of a "wrong" model. Results indicate that "wrong" models are indeed utilised in practice even if they are not recognised so by all parties involved. Our results show that usefulness and model acceptance do not necessarily concur. The outcomes of this investigation can be further utilised in future studies to show how modellers and clients may interact when encountering a "wrong" model.

Keywords: Wrong models, usefulness, discrete-event simulation, uses, model acceptance

1 INTRODUCTION

Simulation modelling is central to the practice of Operational Research (OR). Yet, when creating a model, various issues may be encountered. These could lead to an outcome that may not be considered as adequate by clients or even modellers. Instead, the model may be viewed as "wrong". Should such a model be discarded or can we still use it somehow?

Literature has suggested that "wrong" models may still have certain uses for their users. Papers prefer to focus on successful outcomes (Eskinasi and Fokkema, 2006) not usually including feedback from preceding failed attempts. As a result, we know little about the usefulness of "wrong" models because they are not discussed in the existing published work. The need for this sort of research becomes apparent when dealing with cases of "wrong" models in real life. This paper explores the use of "wrong" models based on suggestions found in the literature. We aim to provide a better understanding of "wrong" models by exploring the extent to which they are used in practice. The main contribution of our work lies in that it provides evidence based on real life stories as described by simulation analysts. We also carry out a focused analysis on one specific reason of wrongness, namely the model's level of detail, and consider how it affects acceptance and use. Modellers may utilise ideas from our empirical analysis presented here and find arguments that justify to their clients that it is possible to take advantage even from a model that they may deem "wrong".

Concerning our approach to the definition of "wrong" models and to their uses, a very important note should be made that applies throughout the context. When discussing "wrong" models and their possible acceptance or use here, we refer to any model element that may be deemed as improper by clients or modellers (or both). We do not equate acceptance or use of a "wrong" model to a modeller

willingly developing a bad model for their clients. For instance modellers may have to accept a "wrong" model due to their clients' preference, or on the other hand, a perfectly acceptable model by the modeller, may be rejected by the client. Thus wrongness is subjective and "wrong" models may be present even if they are not considered or actually be wrong. To demonstrate that this is our definition to wrongness, we use quotes when referring to opinions on a model considered as "wrong".

To achieve our aims, we interview discrete-event simulation modellers on their experience of using "wrong" models and carry out qualitative text analysis. We focus on how modellers deal with "wrong" models and identify the reasons for using them or not by presenting actual cases of "wrong" models from practice. We first review the literature by exploring model usefulness and possible uses of "wrong" models, also focusing on the model's level of detail and its effect on credibility. Then, the methodology and objectives of the study are provided. Results follow the interview analysis with some first deductions. A summary and discussion of proposed future work conclude the paper.

2 "WRONG" MODELS IN LITERATURE

In this section we discuss existing views on "wrong" models in literature in terms of their usefulness and possible uses. Some additional concepts relevant to the needs of this work are also analysed.

Considering how "*all models are wrong*" (Box and Draper, 1987), the concept of "wrongness" is highly diverse in literature. Tsioptsias et al. (2018) reference such examples from various works where the terms include: "*bad*" or "*inadequate*" (Hodges, 1991), "*unvalidated*" or "*unvalidatable*" (Hodges and Dewar, 1992), "*false*" or "*incorrect*" (Bankes, 1998), or simply "*wrong*" (e.g. Bankes, 1993). Regarding reasons of wrongness, various ideas are found on why models may be considered "wrong", like limitations in time or funding (Balci et al., 2002), decisions (Vennix et al., 1999), etc.

Accordingly, different factors during the modelling study may affect model development. These include the gathering of data and creation of a conceptual model, coding, experimentation, and, implementation of the model (Robinson, 2014). The experimentation or testing phase also contains the Validation and Verification (V&V) process of the model (Robinson, 2014). V&V has long occupied literature in terms of different approaches and tests to craft a "proper" model. Their analysis is not within the scope of this paper, yet this has been summarised in Tsioptsias et al. (2016). Lastly, validation relies on "trade-offs" within different validity types and cannot be universal or absolute (Groesser & Schwaninger, 2012; Sargent, 2012). Thus, we may consider that instead of searching for a "perfect" model, we should be selecting among the available alternatives (Brooks and Tobias, 1996). Considering the above, we explore wrongness in this paper as a matter of perspective.

Despite the previous unconsolidated opinions, it is suggested that even a "wrong" model may entail some usefulness. Castaño (1999) mentions that a model is useful if it addresses the problems it is expected to, while Jessop (2002) explains that a model is useful if it provides the groundwork for taking decisions. For Mens and Van Gorp (2006) usefulness resorts to helping system understanding and proper decision-making. Despite the various definitions on usefulness, possible uses of "wrong" models have not been explored empirically within OR, and studies on the matter are scarce. Very few exceptions are encountered. Hodges (1991), and, Hodges and Dewar (1992) reference some relevant possibilities: use to promote or communicate a selling idea, use for training support, use for storing information, exploitation of model for creating new knowledge where precision is not required, etc. Similarly, Bankes (1993; 1998) explains in what ways exploratory and weakly predictive models can assist in decision making. We notice a gap in empirically exploring the aforementioned possible uses.

Due to the broadness of the topic, and besides the general need to address the gap on usefulness, a more focused analysis should also be considered on "wrong" models as it is lacking from literature as well. This would allow a better understanding on the ideas of accepting and using "wrong" models. In order to further focus our analysis, a specific reason of wrongness and a distinct viewpoint of examination are required. We notice that an important factor concerns a model's level of detail. Level of detail refers to a model's description of its objectives and leads to simplification as the methodological approach of reaching or applying that proper level (Innis and Rexstad, 1983). If done in extreme or if it is lacking, then we may end up with a "wrong" model (e.g. Goldberg et al., 1990). Regarding the viewpoint of examination, in OR clients/users are central to its scope, which means that acceptance of a model is very important. Since we are also investigating model usefulness, we need to consider the two notions together. We find that Landry et al. (1983) mention that there is a relation

between model validation and usefulness but still keep the notions distinct. Indeed, usefulness is usually referenced as a separate notion from validity - and consequently credibility which is another concept related to model validation (as explained below). For example Lewandowski (1982) states that "*although generally a valid model is useful this may not always be the case*". In other words, accepting a model for a purpose, and, finding it useful are not necessarily one and the same. Their possible combinations can be demonstrated in Table 1:

Table 1 *Juxtaposition of model acceptance and usefulness*

		Model usefulness (by clients)	
		Useful model	**Not-useful model**
Model acceptance (by clients)	**Yes**	-Accepted by clients -Model identified as useful by clients	-Accepted by clients -Model not identified as useful by clients after all
	No	-Rejected by clients -Model still identified as useful by clients after all	-Rejected by clients -Model not identified as useful by clients

Model acceptance here refers to clients accepting a model, initially, for their purposes. Even an accepted model may be considered "wrong" by any side due to some reason.

Table 1 shows the possible combinations between accepting a model and finding it useful. Interpreting this for "wrong" models, there are two "obvious" cases: an accepted model also considered useful, and, a rejected model not considered useful. Also, models may be rejected initially, but they may still entail some usefulness. A final - slightly contradicting but still plausible - case regards an accepted model that is not considered useful after all. We have thus expressed that usefulness may exist for a model even if it is not accepted. Its importance as a distinct idea when dealing with "wrong" models is becoming obvious. In OR, model acceptance can further be considered as a decision based on what the involved parties think of a model, that is a subjective decision following their denoted credibility. Credibility is one of the various concepts related to model validation (Tsioptsias et al., 2016). While model validation is "*the process of ensuring that the model is sufficiently accurate for the purpose at hand*" (Robinson, 2014), credibility is termed as the clients' belief that a model has credential value (Gass, 1983) and matches the aforementioned consideration of clients being central to the scope of OR. In other words, if a model is credible then we consider that it may be accepted. To summarise, the model level of detail is considered for exploration as a reason of wrongness under the scope of credibility to determine how acceptance and usefulness are affected.

Having elaborated on how literature views usefulness and uses of "wrong" models, as well as focusing on a specific reason of wrongness and how acceptance and usefulness are juxtaposed, we next set objectives over these topics and presents the employed methodology to address them.

3 OBJECTIVES AND METHODOLOGY

The above analysis suggests a gap in literature in terms of how "wrong" models can be used in practice regarding their usefulness. This section presents our approach by setting and elaborating on the objectives and our expectations, and, by introducing the research design.

3.1 Objectives and explanations

We consider the following. The term "wrong" is subjective and we use it for a model with at least one of its elements not considered good enough by either clients or modellers. Also, as stated in Section 1, using a "wrong" model does not equate to a modeller willingly developing or adapting a bad model.

This paper aims to examine the extent to which "wrong" models are used in practice and how modellers and clients deal with them. Based on the identified gap, we set two objectives. The first objective is to explore possible uses of "wrong" models in practice (O1). The second objective is to perform a more in-depth, focused analysis by investigating how the level of detail as a specific reason of wrongness affects model acceptance and model use (O2). Cases of "wrong" discrete-event simulation (DES) models are investigated to identify potential uses as well as to focus the analysis.

O1 is a direct consequence of the gap found in literature in empirically exploring uses for "wrong" models (Section 2) and addresses the question "*are 'wrong' models used in practice and how?*". We would expect that if "wrong" models are used in practice, then the uses of such models could be examined, compared to literature, and, be extrapolated for use in other cases.

For O2, we specialise the investigation for level of detail as a reason of wrongness and its effect on model use and acceptance. Model acceptance is viewed as a decision following clients' perception of a model (credibility). It can result from a model's level of detail by considering a range of acceptance for the client depending on their perception as shown in Figure 1 (adapted from Robinson, 2014). Since increasing the level of detail (complexity) leads to an increase in accuracy only up to a point, we consider two points "x" and "y" which deem the model as too simple ("oversimplified") or too complex ("overcomplicated"). The combination of perceived complexity and accuracy puts model acceptance "on the map" for modellers and clients. Perceived complexity is considered the subjective complexity based on someone's perception of a model. Perceived accuracy refers to the subjectively denoted accuracy leading to credibility and acceptance. The main curve shows the modeller's perception of a model, while the green and red dotted curves create a perception range for the clients (blue arrows). If their perception is "over" the modeller's (green curve) then a model deemed as oversimplified or overcomplicated may be accepted. If their perception is "below" the modeller's (red curve) then a model deemed as adequate may be rejected. The objective thus addresses the question "*does the level of detail affect model acceptance and model use for 'wrong' models?*". We consider that level of detail, acceptance and usefulness are interrelated and may affect each other.

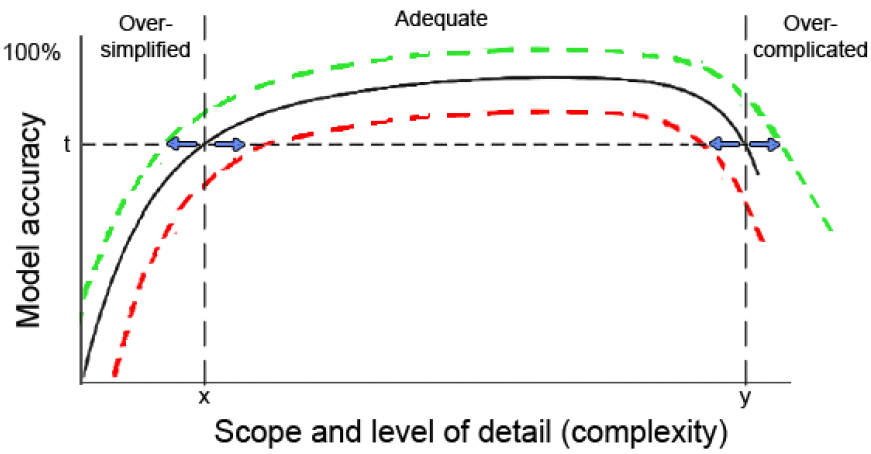


Figure 1 *Adapted from Robinson (2014). Points "x" and "y" separate level of detail (modeller's viewpoint). The main curve shows the modeller's perception of a model. The dotted curves create a perception range for the clients leading to acceptance or rejection of models.*

3.2 Methodology and process of interviews

To accomplish the two objectives, semi-structured interviews with DES modellers are conducted. Interviews appear as the most straightforward method in order to collect information from real-life cases compared to other qualitative methods. Their results can be categorised and compared (Gogi, 2016) in order to lead to specific deductions for the objectives and contrary to case studies and ethnography, interviews require less time and can take place remotely if required. Following, we present considerations regarding the interviews as well as information on the process followed.

The interviews required modellers to recall stories relevant to "wrong" models. Time was given in advance to the participants to recall details based on their memory. This would allow free and detailed narrative which is a requirement when conducting interviews (Brinkmann and Kvale, 2015). Another concern was the number of interviews to have enough material. It has been suggested that saturation tends to happen at around 12 interviews (Guest et al., 2006). Our case study consists of 22 interviews. Based on the amount of material collected and the outcomes (Section 4), we consider the number as adequate for our cause. Ethical guidelines are also important (Brinkmann and Kvale, 2015) and thus an ethics clearance was performed including a consent form informing every interviewee in advance.

The process for each interview is as follow. Initially, DES modellers from different companies are contacted to invite them to participate in the study. More details are provided upon agreement, without referring to the concept of "wrong" models to avoid biasing the modellers' own understanding of the term. Each interview takes between 40 minutes and 1 hour and is done in person unless not feasible, in which case the interview is to be carried out over the phone or Skype. Once collected, all results are anonymised for confidentiality purposes. The questions were based on the objectives formulated for the interviews. We asked the interviewees to comment on each story they detailed, on problems they encountered, the clients' responses per case, and how they dealt with these situations.

Section 4 presents information on the collected material and the results deriving from the analysis.

4 RESULTS

The interviews took place between November 2018 and January 2019 with the modellers. We present some information on the interviews, followed by the results per objective.

4.1 Information on sample

A total of 22 interviews was conducted with simulation analysts that use DES as part of their job. The participating companies are from the private sector, the majority being in consulting with the exception of one, from the manufacturing sector. Modellers of different seniority contributed a total of 54 stories. Out of the 22 participants, half of them were considered experienced as 6 interviewees had relevant experience of approximately two decades within simulation and 5 more interviewees had experience of well over 20 years. A number of 6 interviewees were classified as novice, being quite new to the field (between 1 and 3 years), while the rest were classified as confident users and spanned between 4 and 8 years of work. One interviewee contributed only general input on model wrongness. All interviews were made face-to-face with the exception of two being over the phone, and one over Skype. All stories were considered relevant to "wrong" models as they presented issues related to the model or process followed, or, there was a view related to model wrongness from an involved party.

4.2 Objective 1: Are "wrong" models used in practice and how?

The extent to which "wrong" models are used in practice, and if so how, is tackled first.

Since all stories contained a "wrong" model, it is considered that real-life cases have to deal with such issues. It was first examined if "wrong" models were used in any manner. The interview material was analysed iteratively on whether a story's model was used by the client regardless of credibility or acceptance and how, either as direct answers to interview questions on what happened with each model or as logical deductions from each narrative. Table 2 summarises the stories as follow: A model is considered to have been "used" if there was a reference of a possible utilisation of that model in its corresponding story despite being considered "wrong". A model is considered to not have been "used" if there was a reference of no further use in its story after being considered "wrong". Any story not delivering either information for a model was separately categorised as unclear.

Table 2 *Categorisation of whether "wrong" models were used*

Category	Number of models
Used	25 (46.3%)
Not used	24 (44.4%)
Unclear	5 (9.3%)
Total	**54**

About half of the cases were attributed to the category of using a "wrong" model in at least some manner - regardless as stated who considered it "wrong". Unclear cases regarded confidentiality issues from the modeller's clients or lack of knowledge in view of what happened next.

If a "wrong" model was thus used, then the reasons for using it and its uses were considered. Iteratively revisiting the content of the discussions, both reasons for use and specific uses of those models were grouped and categorised. Table 3 presents first the reasons for using "wrong" models.

Table 3 *Reasons for using "wrong" models*

Reasons	**Number of models**
Further investigation	8 (32%)
No alternative	7 (28%)
Client happy	7 (28%)
Third party	3 (12%)
Total	**25**

Each model is attributed to one main category of reason for being used

Results suggest that in most stories a "wrong" model may have been used due to further exploration of the issues taking place with the modeller showing to clients that the model may still be useful regardless of wrongness. Also, a "wrong" model may have been used because there were no other present options or because changes to that model had no effect on its wrongness. A similar number of cases presents utilisation of a "wrong" model due to clients considering the model adequate for their purposes as for example being aware that the model has wrong elements but still chose to implement, or, not finding anything problematic with the model from the start. Lastly, the involvement of possible third parties had an effect on 3 cases, where a "wrong" model was inclined to be further utilised. Third parties here may refer to anyone interested in the project (e.g. a director or a client of a client) that could have affected the acceptance of a model without necessarily being involved in its development.

Next, the uses of those models are provided in Table 4. These derive from categorising the way that the models were used in the 25 cases identified to have a "wrong" model used (Table 2). The categories were created by revisiting iteratively the stories and checking for commonalities.

Table 4 *Uses of "wrong" models*

Reasons	**Number of models**
Decisions or actions	17 (68%)
Hypothesis or scenario testing	8 (32%)
Surface issues	7 (28%)
Better understanding or thinking development	3 (12%)
Promote or communicate selling	2 (8%)
Store or monitoring	1 (4%)

A model may be attributed to more than one categories of uses

The categories suggest that a "wrong" model may have been used most often for helping in decision-making, exploring options on actions. Training support is included in this category. The second most often use refers to helping the clients experiment on parameters when outcomes are not necessarily required to be numerically perfect, followed by models used to highlight problems that involve the clients or their collaborators. Next, a model may have been used because it allowed insights to the simulated reality. Lastly, some cases were found where a "wrong" model was used to either promote a selling idea or for storing information. As a final note, if a model was not used, two main reasons were found: the models were either not used at all (6 cases - 25%), or, they were changed into an "adequate" model and then used thus not being considered "wrong" any more (18 cases - 75%).

Concluding O1, we have found that "wrong" models are indeed encountered and also used in practice quite often. Reasons and uses were identified and categorised. Next, we focus our analysis.

4.3 Objective 2: Does the level of detail affect model acceptance and use?

Having explored whether "wrong" models are used in practice as well as their uses, we now focus on level of detail and its effect on model acceptance and use. This specification will help us understand better how model acceptance and use may interact based on an exemplified reason of wrongness.

A total of 13 out of the 54 stories highlighted issues related to level of detail. These could have derived from modellers, clients or both. Models considered simpler than expected or to be missing elements were denoted as "oversimplified" (OsM), and, models considered too complex or having

more elements than required were denoted as "overcomplicated" (OcM). Table 5 presents the number of models viewed as "wrong" by modellers and clients due to their level of detail.

Table 5 *Number of "wrong" models due to level of detail*

	Modellers	Clients	Unique "wrong" models
OsM	4	4	7
OcM	8	4	8
"Wrong" models per side	**12**	**8**	**13**

OsM = Oversimplified models, OcM = Overcomplicated models

In our sample, the modellers tend to denote wrongness due to level of detail more often than the clients, with 12 cases compared to 8. Additionally, the modellers here reference more often issues with OsM than OcM, with 8 cases compared to 4. The clients' concerns in 8 stories are equally split between OsM and OcM. In 5 out of these 8 stories the opinions coincide and in 2 stories the opinions are opposite to the modellers while in 1 story only the clients denoted issues. In total, 7 stories referenced an OsM and 8 an OcM, repeating counting for the 2 models with opposite views.

Comparing the above with model acceptance, we get the following correspondence in Table 6.

Table 6 *Number of "wrong" models and model acceptance*

		Modellers	Clients	Unique models
Accepted by clients	**OsM**	4	2	5
	OcM	2	1	2
	Total	**6**	**3**	**7**
Not accepted by clients	**OsM**	0	2	2
	OcM	6	3	6
	Total	**6**	**5**	**6**

Two models are viewed differently by modellers/clients and are repeated in counting (in red).
OsM = Oversimplified models, OcM = Overcomplicated models

Model acceptance was deducted based on the interviewees' views of their clients' credibility. Out of the 13 models, 7 were accepted and 6 rejected. From Tables 5 and 6, we notice that 6 out of the 12 models with issues denoted by the modellers were accepted. On the contrary, only 3 out of the 8 models with issues denoted by the clients were accepted. This is expected, as the clients may have a stronger opinion on wrongness and acceptance. We highlight that 5 out of the 7 OsM were accepted, contrary to only 2 out of the 8 OcM. This could mean that OsM can be more easily apprehended.

Regardless if a model was initially accepted, we also check whether that model was further used. Though we do not refer here to the reasons and ways of using or not using a model, we consider that an accepted model may have ended up not being used, while a rejected model may have still been found useful (see Table 1). By iterating the material, in view of usefulness we notice that: Out of the 7 accepted "wrong" models, 5 were actually used which means that 2 models despite initial acceptance were not further utilised. Regarding these 2 models not used, 1 of them was considered as oversimplified by the modeller but adequate by the clients while the other one was considered adequate by the modeller but overcomplicated by the clients. Out of the 6 rejected "wrong" models, 5 were not used which means that 1 model despite initial rejection was further utilised. That model was considered overcomplicated by the modeller and the clients but they still found some use in it. It is noted that the 2 models with contrasting opinions on their complexity were neither accepted nor used.

Table 7 *Comparing acceptance and use of "wrong" models due to level of detail*

	Used	Not used	Total ("wrong" models)
Accepted by clients	5	2	7
Not accepted by clients	1	5	6
Total ("wrong" models)	**6**	**7**	**13**

Table 7 summarises the discussed combinatory outcomes of acceptance and usefulness. Each of these 13 models could be depicted on Figure 1 based on level of detail and acceptance.

Consolidating Tables 5-7, we notice that a model denoted by the clients as "wrong" due to its level of detail might be accepted but not necessarily used and vice versa. For example, an accepted OsM denoted as such by the clients is not used after all, while 1 of the 2 used models that was denoted as OcM by the clients was not initially accepted. Furthermore, there is an equal split for the models where the modeller considered level of detail as an issue in view of their acceptance as well as use.

Table 7 is an empirical evaluation of Table 1 and corroborates the claims made in view of acceptance and usefulness. Our expectation that level of detail, acceptance and usefulness affect each other cannot be supported or rejected quantitatively as the sample is very small, but from a qualitative point of view we could state the following: an accepted model has a higher chance of being used, and, the level of detail as an issue for a model does not seem to affect acceptance as often when deriving from the modellers' side but it is taken into better consideration when deriving from the clients' side. These deductions suggest that models may be denoted as "wrong" by a side but still be accepted and/or used while their acceptance may not necessarily coincide with the usefulness they may provide to their clients after all. This focused analysis could be repeated for other reasons of wrongness.

To summarise, our in-depth investigation on one specific reason of wrongness and its effect on acceptance and usefulness offered some interesting inputs. We found that opinions on level of detail do not always coincide and may even differ, with modellers addressing level of detail more often as a problem for models. Also, accepting a model seems more related to cases where the clients did not have a negative opinion of the model since clients may have a stronger opinion on model wrongness. Yet, this may still change and they may utilise an initially rejected model. Lastly, oversimplified models may be accepted more frequently than overcomplicated ones. The next section summarises and discusses the work of the paper, alongside limitations and future expansions.

5 DISCUSSION, LIMITATIONS AND FUTURE WORK

This paper is the first to address in practice the possible usefulness that "wrong" models may still entail within OR. After presenting how literature regards model usefulness and acceptance, a research gap was encountered. Two objectives were set to address the gap: whether and how "wrong" models are used when encountered in practice, and, a focused investigation combining a specific reason of wrongness and its effect on model acceptance and model use. Interviews with DES modellers to explore the objectives were conducted, discussing DES modelling stories. Our aim was to provide a better understanding on "wrong" models by exploring the extent and way to which they are used in practice. The most relevant findings are discussed here followed by study limitations and future work.

The possible usefulness of "wrong" models is lacking in literature. We showed how model acceptance can be juxtaposed with usefulness, suggesting that the two may not coincide. Still, and despite the plethora of reasons for model wrongness, the possible uses of models considered "wrong" had not been addressed in practice.

The examination of the first objective showed that "wrong" models are indeed encountered and utilised in practice in at least some manner for almost half of the sample's examined cases. Reasons for using "wrong" models were identified and categorised. The most often cases included models used for helping with decision support or hypothesis testing, and the reasons for using them were most often due to lack of alternatives or because clients were happy with a specific model regardless if they found it "wrong", or, due to models being proven to have some value by the modellers to the clients. It is thus easily suggested that views on wrongness may differ or even collide. A "wrong" model for a modeller may seem perfectly fine for a client. Or, vice versa, a client may dismiss a model that is found to be adequate by its modeller. This finding, alongside the possible uses of "wrong" models, can be utilised when the two sides - modellers and clients - derive to unconsolidated opinions on a model, i.e. even if the model is deemed as "wrong" there may be usefulness and specific uses to it. The uses of "wrong" models explain as well the fact that the subjectively denoted idea of a created model being unfit for a cause can still find fruitful derivatives for a client.

The analysis on the stories for issues on their models' level of detail for O2, offered some further understanding on how that specific reason of wrongness affects model acceptance as a decision

following credibility, and, model use. It was suggested from the sample that modellers denote wrongness due to level of detail more often than clients but opinions do not always coincide and may even be opposite. It was inclined that clients considering a model "wrong" may still accept it but not necessarily use it and vice versa. Additionally, clients have a stronger effect on acceptance or rejection of a project than modellers (as expected), especially for oversimplified compared to overcomplicated models. Yet, their opinions on using or not a model may be altered by the modellers. These deductions lead to the idea that accepting and using a model does not necessarily coincide when level of detail is the main attributed reason of wrongness, nor does the opinions between the modeller and the clients in view of the assigned level of detail of a model (too much or not enough).

The contributions of our work can be used as empirical evidence and ideas to help support the interaction between clients and modellers in case of disagreement on model complexity/simplicity as well as the utilisation of such a model regardless of the initial acceptance or rejection. Indeed, a model may be rejected by a client but if the modeller can prove its further usefulness, it may still apply to some extent as an additional tool for decision-making or some other function.

A number of limitations apply to this study. The stories were reviewed from the modellers' perspectives, thus an extension to clients or validation of results with their help would allow more concrete deductions. The analysis considered only one reason of wrongness under the scope of credibility, while the sample was very limited with only 13 models referencing this reason. Possible interrelated reasons between issues leading to model wrongness and their effect on accepting or using a model could apply, which has not been tackled here. The interaction with the clients' and modellers' decisions could be further explored. Lastly, all of the above analysis is subject to interpretation bias.

The above pave the way towards a further examination of how "wrong" models may be of use to both clients and modellers through the idea of learning, since using a "wrong" model may offer specific benefits. In future work, the authors aim to further analyse the interviews, under the scope of learning, and to identify how different stakeholders interact with "wrong" models when encountering them. As a result we aim to develop a framework of interaction between the stakeholders involved in a simulation study to suggest possible courses of action when "wrong" models are encountered.

ACKNOWLEDGMENTS

The authors acknowledge the support of SIMUL8 Corporation for offering a travel grant to support research activities of the PhD project that this paper is part of. The authors would also like to thank all the simulation modellers and their companies for taking part in the interviews presented in this paper.

REFERENCES

Balci, O., Nance, R. E., Arthur, J. D. and Ormsby, W. F. (2002). Expanding our horizons in verification, validation, and accreditation research and practice. *Proceedings of the Winter Simulation Conference*, 1(2000), pp 653–663.

Bankes, S. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, 41(3), 435–449.

Bankes, S. (1998). Policy Analysis for Complex and Uncertain Systems Through Computational Experiments. In *IEEE Aerospace Conference*, pp 9–21, vol.1.

Brooks, R.J. and Tobias, A.M. (1996). Choosing the best model: Level of detail, complexity, and model performance. *Mathematical and Computer Modelling*, 24(4), pp 1–14.

Brinkmann, S. and Kvale, S. (2015) *Interviews: Learning the Craft of Qualitative Research Interviewing*. 3rd Edition, Sage Publications, Thousand Oaks, CA.

Box, G. E. P. and Draper, N. R., (1987) *Empirical Model-Building and Response Surfaces*, John Wiley & Sons.

Castaño, Y. Á. (1999). Dynamic Behavior of NPD. *Proceedings of the 17th International Conference of the System Dynamics Society and 5th Australian & New Zealand Systems Conference*, 10.

Eskinasi, M. and Fokkema, E. (2006). Lessons learned from unsuccessful modelling interventions. *Systems Research and Behavioral Science*, 23, pp 483–492.

Gass, S. I. (1983). Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Operations Research*, 31(4), pp 603–631.

Gogi, A. (2016). *Insight generation in simulation studies : an empirical exploration*.
Groesser, S. and Schwaninger, M. (2012). Contributions to model validation: hierarchy, process, and cessation. *System Dynamics Review*, 28(2), pp 157–181.
Guest, G., Bunce, A. and Johnson, L. (2006). How Many Interviews Are Enough? *Field Methods*, 18(1), pp 59–82.
Hodges, J. and Dewar, J. (1992). Is it you or your model talking? A framework for model validation. *Retrieved from http://www.rand.org/content/dam/rand/pubs/reports/2006/R4114.pdf*
Hodges, J. S. (1991). Six (Or So) Things You Can Do with a Bad Model. *Operations Research*, 39(3), pp 355–365.
Innis, G. and Rexstad, E. (1983). Simulation model simplification techniques. *Simulation*, *41*, 7–15.
Jessop, A. (2002). Exploring structure: A blockmodel approach. *Civil Engineering and Environmental Systems*, 19(4), pp 263–284.
Landry, M., Malouin, J.-L. and Oral, M. (1983). Model validation in operations research. *European Journal of Operational Research*, 14(3), pp 207–220.
Lewandowski, A. (1982). Issues in model validation. *International Institute for Applied Systems Analysis*, RR-82-37(March), pp 1–26.
Mens, T. and Van Gorp, P. (2006). A taxonomy of model transformation. *Electronic Notes in Theoretical Computer Science*, 152(1–2), pp 125–142.
Robinson, S. (2014). *Simulation - The Practice of Model Development and Use*, 2nd edition. Macmillan education, Palgrave.
Sargent, R.G. (2012). Verification and validation of simulation models. *Journal of Simulation*, 7(1), pp 12–24.
Tsioptsias, N., Tako, A. and Robinson, S. (2016). Model Validation and Testing in Simulation: a Literature Review. *5th Student Conference on Operational Research*, 6:1-6:11.
Tsioptsias, N., Tako, A. and Robinson, S. (2018). Can we learn from wrong simulation models? A preliminary experimental study on user learning. *Proceedings of the Operational Research Society Simulation Workshop 2018 (SW18)*.
Vennix, J. A. M. (1999). Group model-building: tackling messy problems. *System Dynamics Review*, 15(4), pp 379–401.

AUTHOR BIOGRAPHIES

NAOUM TSIOPTSIAS is a Doctoral Researcher at Loughborough University in Operational Research. He received a Bachelor on Finance & Accounting from the University of Macedonia (Thessaloniki) and an MSc on Informatics & Management from the Aristotle University of Thessaloniki. His current research concerns "wrong" DES models uses, while previous interests regarded data mining on financial crises. Ha has also worked as an analyst for an HR company.

ANTUELA TAKO is a Reader in Operational Research at the School of Business and Economics, Loughborough University. She holds a PhD in Simulation and an MSc in Management Science and Operational Research from the University of Warwick. Her research interests include the comparison of simulation approaches, facilitated and participative simulation modelling, conceptual modelling and health care modelling. She is an Associate Fellow of the Operational Research Society (AFORS), Associate Editor of the Journal of the Operational Research Society and the Journal of Simulation.

STEWART ROBINSON is Dean and Professor of Management Science at Loughborough University, School of Business and Economics. Previously employed in simulation consultancy, he supported the use of simulation in companies throughout Europe and the rest of the world. He is author/co-author of six books on simulation. His research focuses on the practice of simulation model development and use. Key areas of interest are conceptual modelling, model validation, output analysis and alternative simulation methods (discrete-event, system dynamics and agent based). Professor Robinson is co-founder of the Journal of Simulation and the UK Simulation Workshop conference series. He was President of the Operational Research Society (2014-2015). Home page: www.stewartrobinson.co.uk.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.019

FULL-DAY SIMULATION OF FIELD ENGINEERING OPERATIONS

Nicola Wong, Mathias Kern, Anne Liret, Raphael Dorne

Applied Research, BT
Adastral Park, Martlesham, Ipswich IP53RE
nicola.wong@bt.com, mathias.kern@bt.com, anne.liret@bt.com, raphael.dorne@bt.com

ABSTRACT

This paper describes an on-going project on a full-day simulation of field engineering operations, enabling analysis of a company's capability to cope with changes in work demand and labour supply. This case study incorporates the operational dynamics in the scheduling process, illustrating a schedule of work for service delivery in BT Group plc (BT) over the course of an operational day. These jobs range from repairing existing network cables; installing new infrastructure; connecting cities to the internet and beyond. In this project, we model our operations using a discrete event simulation approach, adding a number of key in-day disturbances which are considered turbulent to BT's service delivery. Furthermore, we discuss the current development state and future directions for this project.

Keywords: Discrete-Event Simulation, Resource Management, Scheduling, Operational Dynamics

1 INTRODUCTION

In many service organisations such as communication or utility companies, teams of field engineers deliver services to customers. For these field teams, it's a daily reality that things do not go as planned, e.g. work could take longer than expected or might not be completed due to certain constraints. Therefore, schedule planners need to frequently review and update work schedules for their engineering resources in order to continuously balance work demand with supply of labour. This enables teams to meet the often challenging delivery targets towards the end of the day.

When creating a simulation for service operations, it is important to model in-day dynamics to ensure a close alignment between simulated results and real life outcomes. This not only has the potential to facilitate the verification and improvement of existing scheduling setups, but also allows the assessment of any impact to the operations when introducing new products or service offerings.

In this paper, we present a model that simulates both the start-of-day and in-day dynamics for a team of field engineers in the context of service delivery by BT.

2 DYNAMIC NATURE OF FIELD OPERATIONS

Service companies often utilise highly sophisticated scheduling systems for the allocation of work to their field resources. In BT, FieldSchedule (Liret et al, 2007 ; Owusu et al, 2017) is used to organise the large engineering teams in Technology's and Enterprise's field operations. A typical scheduler setup, also known as "predictive-reactive scheduling"(Chin S C, Appa I S and Robert G, 2003) includes a start-of-day scheduling engine run which produces an initial schedule for a full working day based on the work and resource information available very early in the morning. The system then updates this schedule with periodic "in-day" scheduling runs applying any latest changes in the field teams. This information can include the time taken to progress or complete a task, updates on travel duration; unplanned work being added and planned jobs being removed. Moreover, a day's plan can often go pear-shaped when job priorities are modified or availability of labour changes in the course of an operational day. In short, field operations can be very volatile, and a good scheduling system needs to manage these dynamics.

In order to address to different service domains and their requirements, scheduling systems can be customised to a comprehensive set of configurations, by using what can be hundreds of parameters. The setup and optimisation of scheduling configurations for specific scenarios is a non-trivial challenge that can have a profound impact on the service and productivity delivered. As live trials are often time-intensive and expensive, simulation plays an important role in evaluating different scheduling configurations. While traditionally, the focus has been on improving an individual schedule engine run at start-of-day, the aforementioned dynamic nature of field operations will require a simulation for an entire operational day – predictive schedules at start-of-day and actual outcomes at end-of-day are often very different. In the next section, we outline our approach to a full-day simulation of field operations.

3 SIMULATION APPROACH

We have developed a simulation system that not only simulates individual start-of-day scheduling scenarios but that also incorporates the key dynamics of operations with in-day scheduling. The end-of-day results produced by this full-day simulation provide more realistic business insight than a start-of-day simulation alone.

We have chosen a discrete event simulation approach (Liret A, 2009) in a combination with an Object Oriented (OO) program to model the change of states in a team of field resources and tasks on an operational day. The system's state is made up of a list of task instances, a list of resource instances and a 'current schedule' instance, maintained at all times. The state changes over a notion of simulated time according to the execution of a dynamically ordered list of events. An initial event list is created with events of a fixed time, such as the engineer signing on, the start-of-day schedule run event and the periodic in-day scheduling run events. Other events, such as engineer travel events, task begin/complete events and resource absences events, are dynamically added to the list. The event list always holds the events in a chronological order, and the simulation steps through the events in the same manner.

Figure 1 illustrates a sequence of key events which update the system states. Dark solid arrows represent the creation of an event. An event is executed when the simulation timeline arrives at the event start time. The execution of an event can create further events. Key information including start time of an event is often determined by its 'parent' event. Estimation functions may be applied to model these features of a generated event. For example, a 'task complete' event is generated following a 'task begins' event, and the time difference, i.e. the simulated task duration, is stochastically determined. The event list is maintained chronologically and kept up-to-date when an event inserts one or more offsprings at particular time points after its execution.

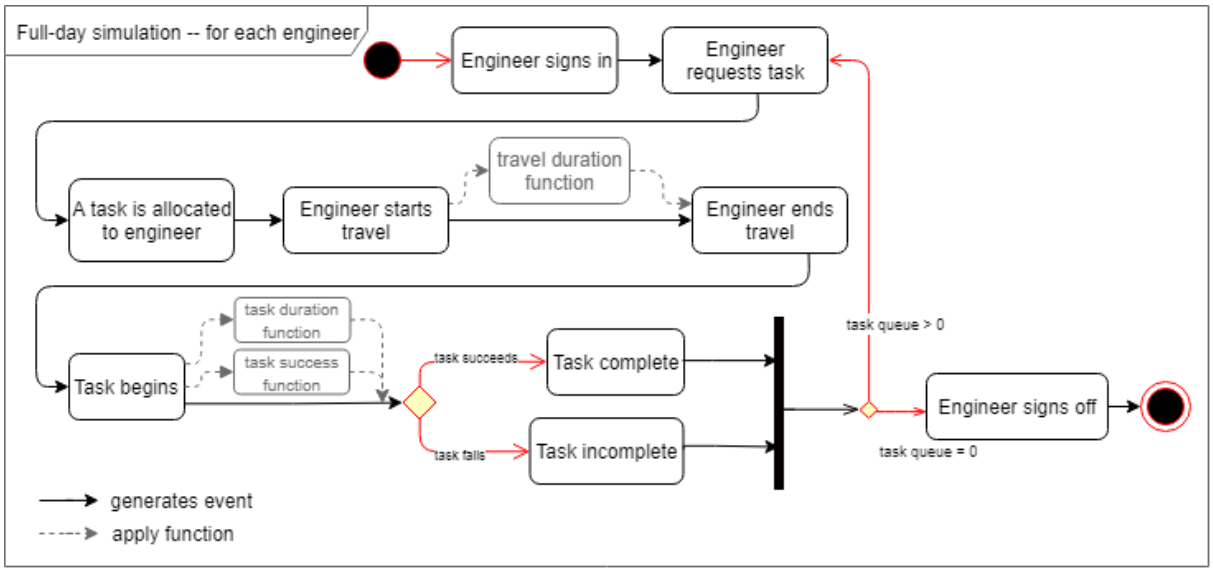


Figure 1 *State diagram of event sequence upon each engineer's sign-on*

As outlined earlier, numerous events can impact the delivery of service by field engineering. Based on feedback from BT's operational teams, our main stakeholder in this use case, these turbulences are categorised in the following 6 types of unexpected events in our model:

i. Variation in actual task times
ii. Variation in actual travel times
iii. Success rates for the completion of work
iv. Unplanned absences of resources
v. Arrival of additional work
vi. Cancellations of existing work

One of the biggest challenges to model these stochastic disturbances is to determine the estimation function for their occurrences. As a case in point, The solid line on the graph shown in Figure 2 illustrates the distribution of historically observed task times for two typical types of work. We can see that the actual task time varies a lot. When a Gamma distribution is fitted to approximate the distribution of varying task times, i.e. dotted lines in Figure 2, we can see our estimation produces a close match to our historical data. So while a traditional single start-of-day schedule organises jobs bases on the an average of task times, the full-day simulation can produce a more realistic model by incorporating the varying task times as updates when frequently rescheduling the tasks.

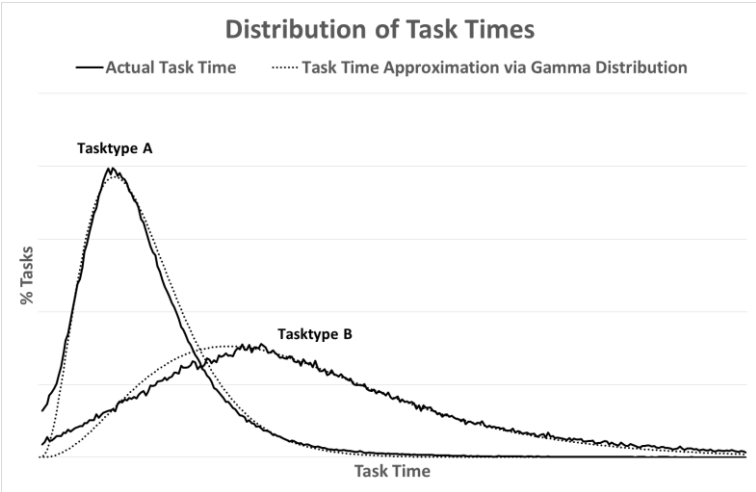


Figure 2 *Distribution of actual task times and approximation using Gamma distributions*

Likewise for the other the key disturbances listed above. A probability function is applied to their occurrences. These probabilism are determined by our analysis similar to this example with task times, based on the data contributed by our operational teams.

In this project, a simulated operational day starts at 00:00:00 and ends at 23:59:59 of the day. A start-of-day schedule is produced at 00:00:00. From that point on, a configurable number of periodic in-day schedules are generated throughout the simulated day, e.g. every 30 or 60 simulated minutes from 06:00:00 onwards. This means at start-of-day, the event list contains the different schedule run events, engineer sign-on events, new task arrival events that are generated stochastically, and more. The simulation retrieves the next event from the list, executes this event which induces an update in the state of resources, tasks and schedules, and inserts any new 'child' events to the event list. The simulation then moves on to the next chronological event. The simulation terminates once all events from the event list are executed for the simulated day. The simulation collects key statistics throughout its execution, building metrics such as average travel times, task completion rates, on-time service, and more.

This collection of data forms a central output of the simulation to assess the task coverage in a particular scheduling or service setup. In the occasion of testing a new work prioritisation business rule, the start-of-day schedule might not reflect the impact of such service setup. However, cases of increased travel or lower task completion could be modelled by a full-day simulation. Frankly, a better insight is presented.

4 CURRENT STATE AND FUTURE WORK

Our work so far has focused on developing a system that is capable of simulating an entire operational day of a field engineering team. We have achieved this aim and are able to simulate both start-of-day and in-day events including key dynamics and disturbances such as variation in task times and task completion rates. We have been able to run initial tests which underline that outcomes envisaged by start-of-day schedules and outcomes delivered at the end-of-day can differ significantly.

The current technical implementation in Java is built in a modular infrastructure where the application is divided into four components: a data retrieval process, an API that queries the scheduling algorithm, a type-agnostic main simulation algorithm and a user-interface for displaying results. By having a dedicated simulation process that communicates with the scheduling engine via a generic API, we can plug different scheduling engines into the simulation and compare them. At the moment, we have employed FieldSchedule, an engine developed and maintained by our team, into the full-day simulation.

Going forward, our research and development work will focus on the following key areas:

1. Ability calibrate the simulation system, i.e. ensure that end-of-day results produced by our full-day simulation match observations from real life operations closely.
2. Provide meaningful suggestions to the business stakeholders on adjusting our work-force in BT utilizing simulation results
3. Enhance what-if and how-to simulation capabilities of the system
4. Enhance user interface to make the tool available to operational units
5. Build connections to plug-in other / third-party schedulers

REFERENCES

Chin S C, Appa I S and Robert G (2003) Dynamic scheduling I: simulation-based scheduling for dynamic discrete manufacturing. *Conference Paper in Proceedings* – Winter Simulation conference January 2004 DOI:10,1109/WSC.2003.1261590

Liret, A., Lesaint, D., Dorne, R., Voudouris, C. (2007). iSchedule, an optimisation toolkit for complex scheduling problems. *Multidisciplinary International Scheduling Conference:Theory and Applications (MISTA)*, Paris, France, August 2007.

Owusu G. , Shakya S., Liret A., Mccormick A. (2017) Leveraging OR Techniques for Smarter Field Operations. Impact journal, vol 3(1), pp.18-21, March 2017. DOI: 10.1080/2058802X.2017.11964018

Liret, A., Workforce Dynamics Simulator in Service Operations Scheduling Systems, International Conference on Enterprise Information Systems CENTERIS 2009, Ofir, Portugal, October 2009

AUTHOR BIOGRAPHIES

NICOLA WONG is currently a research specialist in the Operational Transformation research theme at BT Applied Research. She received a BSc (Hons) Computer Science in 2016.

MATHIAS KERN is a senior research manager in BT's Applied Research team, and his research interests focus on the application of AI, optimisation and simulation techniques to real-life resource management problems. He gained his PhD in Computer Science from the University of Essex researching population-based optimisation methods, and has been with BT since 2004.

ANNE LIRET is currently research manager at BT's Applied Research. Her research includes Operational Transformation, AI and Optimisation problem modelling and solving, with a focus on interactive explainable resource scheduling, planning and simulation. She has a PhD in AI and formal computation, and joined BT in 2000.

RAPHAEL DORNE works as a Principal Consultant in BT Applied Research. His focus is on applying advanced AI techniques to areas such as Resource Scheduling, Asset Management and Network Design.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.020

STRATEGIC DECISION MAKING FOR THE NORTH WEST AIR AMBULANCE CHARITY USING DISCRETE EVENT SIMULATION

Dr Lucy E. Morgan

Lancaster University
Lancaster University Management School
l.e.morgan@lancaster.ac.uk

Dr. Roger J. Brooks

Lancaster University
Lancaster University Management School
roger.brooks@lancaster.ac.uk

ABSTRACT

This paper is a preliminary investigation of the provision of enhanced emergency health care at night by North West Air Ambulance (NWAA). A discrete event simulation model of the current NWAA operational system is presented. This model is used to investigate the provision of NWAA services by an air ambulance vs. a rapid response vehicle. More widely, this model will be used within the scope of investigating what NWAA describe to be their 'optimal service hours'.

Keywords: Discrete Event Simulation, Air Ambulance, Helicopter Emergency Medical Service.

1 INTRODUCTION

North West Air Ambulance (NWAA) are a charity funded organisation that provide enhanced healthcare interventions throughout the Northwest in the areas of Greater Manchester, Lancashire and Cumbria. NWAA provide advanced medical expertise and equipment that can be utilised at the scene of an emergency. They can also facilitate conveyance of patients to hospital when required. At present the charity has six vehicle assets, three air ambulances (helicopters) and three rapid response vehicles (RRVs), and attend over 2,000 missions each year. NWAA have three advanced healthcare teams and two base sites; two of the healthcare teams are based at Barton near Manchester and one at Blackpool. All team members have received advanced training for the types of missions they are likely to attend with NWAA. The teams are made up of specialist paramedics and/ or a highly trained doctor (consultant or anaesthetist). Each team can be assigned to a single helicopter or RRV; therefore there is a maximum of three NWAA assets in use at any one time. See Table 1 for a comprehensive list of the NWAA assets, including the asset call signs and air ambulance and RRV pairings.

In this paper we present the initial results of investigating the strategic working hours of NWAA. The problems faced by NWAA concern what they describe as 'optimal working hours' i.e. what resources to provide at which times of day. In this paper we describe a simulation model developed in WITNESS to

Table 1: A breakdown of the NWAA assets, their base locations and healthcare teams.

Asset (call sign)	Base location	Healthcare team
Helicopter (H72) RRV (HX01)	Barton	Consultant + Paramedic
Helicopter (H75) RRV (HX03)	Barton	2 Paramedics
Helicopter (H08) RRV (HX02)	Blackpool	2 Paramedics

evaluate a range of service hour options and then describe its use to consider one particular issue, providing single resources outside of daylight hours. Currently NWAA complete missions during daylight hours only, but have an interest in extending operational hours beyond this if there is evidence of a need for enhanced healthcare provision at night. Note that the operation of air ambulance missions at night comes with increased risk and cost; investment in a new helicopter would be required before NWAA could carry out any air ambulance missions at night. Our interest therefore lies firstly in whether there is the need for NWAA to provide a night time service, and secondly in whether an air ambulance is the most appropriate asset to do this. Could a RRV could provide a comparative service to an air ambulance through the night? No additional investment, other than crew costs, would be required to use a RRV at night.

We state clearly that the work presented in this paper is a work in progress; all results presented are preliminary and may be subject to further investigation.

In Section 2 we briefly discuss the background of air ambulance provision at night and the current literature in the area. In Section 3 we discuss model development for the NWAA operational system including the sources of data used to validate the baseline model and the assumptions/ estimates made during the experiments. In Section 4 we present the results of some preliminary experiments, and in Section 5 we conclude and discuss future experiments.

2 BACKGROUND AND LITERATURE

The use of helicopters to provide emergency medical services at night has been a topic of interest for some time in the UK. In North America and continental Europe night flying air ambulances are already common McQueen et al. (2015), but such missions come with additional costs and risk. All night time flights in the UK must adhere to the Civil Aviation Safety Directive (CAA 2019).

To date retrospective studies have been popular for identifying the number of incidents an air ambulance may have attended during the night over a given period. Lyon et al. (2015) use a retrospective analysis of trauma patients as part of their investigation into the need for a UK helicopter emergency medical service by night in Kent, Surrey and Sussex (KSS). Their analysis included five independent Helicopter Emergency Medical Service (HEMS) clinicians who were asked to identify which patients met HEMS activation criteria. Within this investigation they also completed a prospective study in which a HEMS dispatch paramedic was present in the ambulance dispatch centre during the night to identify HEMS activation cases. This study was used to motivate the need for a night time air ambulance service in KSS. Curtis et al. (2017) followed up on this work by looking at the implementation of an emergency medical night time service in KSS over a two year period. This paper sought to compare the actual need for the HEMS night time service to the previously estimated need. A number of interesting results arose from this study including the increased number of conveyances of patients to hospital, the greater severity of injury of patients and the introduction of a mission planning phase into mission cycle time during the night. Note that this practical study also reported that significant delays were encountered when the enhanced care team had to respond via RRV during nights when weather meant the air ambulance could not fly.

McQueen et al. (2015) also use a retrospective study to investigate whether helicopters are the answer to responding to major trauma incidents in the West Midlands (WM). Cases likely to meet helicopter activation criterion were identified from historical data by considering injury severity, enhanced care team activations and location. Unlike in KSS the WM already provided emergency care response at night using a fast response vehicle; during the day this team utilised an air ambulance. In the WM a number of voluntary care teams also exist throughout the region to provide emergency health provisions, and before the study the WM were already said to provide good emergency healthcare provision during the night. This may be a result of the good road networks in the region and the volunteer services. Using evidence from the retrospective study McQueen et al. (2015) found that there was little evidence to suggest the need for an air ambulance service in the WM during night.

These sources highlight the importance of emergency health care services at night whether that provision is delivered by an air ambulance or not. It is not clear which characteristics lead to the need for night

Table 2: A description of the possible NWAA mission results.

Job Result	Description
SD	NWAA have accepted the job, but are stood down at take off, en route or at the scene.
TR	NWAA treat a patient/ patients at scene.
CO	NWAA convey a patient or follow the ground escort to hospital.
Missed	The NWAA asset required for mission is unavailable, and the job is missed.

time air ambulance services; this may depend on the characteristics of the region including the current healthcare provisions. Another point of interest is how both studies only considered trauma patients to motivate the need for air ambulance services at night. Whilst trauma patients make up a large part of the number of patients NWAA attend to, their dispatch criteria is not exclusive to trauma patients.

3 THE MODEL

Before considering any changes to NWAA operations we first required an understanding of the current operational system and a model that accurately reflected it. Access was provided to the NWAA owned database, HEMSBase, by NWAA. The HEMSBase provided a great amount of detail on each mission carried out by NWAA including exact times for take off, mission cycle time and the mission result. To gain understanding of which entries were of greatest importance we liaised with managers and medical staff from NWAA. After these discussions and some initial data exploration we developed a simple conceptual model of the job process within the NWAA system, see Figure 1. The NWAA job process comprises of: a job arriving, when occurs when an emergency call has been flagged as requiring NWAA assistance; a job being allocated to an appropriate asset; mission time which can include travel time and treatment and/or conveyance if required, and the job leaving the NWAA system; which can describe NWAAs part in the mission being complete or NWAA rejecting/ missing a job.

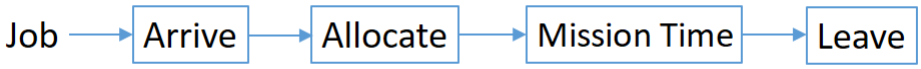


Figure 1: A conceptual model of the job process.

On further consideration of the HEMSBase we were able to pick out the input models we would use to drive our simulation of the NWAA system. We first considered the distribution of the jobs over different days and found that whilst the behaviour on weekdays was reasonably constant the difference between weekdays and weekends was significant. For the remainder of this paper we shall focus on weekdays alone. Further investigation is required to draw conclusions about weekends although many of the decisions we made about the input models still hold.

For weekdays simple analysis confirmed that whilst the arrival profile was not significantly different, both the the proportion of arrivals and the mission cycle times were significantly different for each asset. We therefore estimated a single arrival profile for all six assets and estimated the proportion of jobs to assign to each asset using the number of jobs completed per asset during periods when all assets were on duty. The arrival profile was approximated by a Poisson process with a piecewise constant arrival rate over 30 minute intervals. The mission cycle time distributions for each vehicle were broken down further as analysis of the HEMSBase showed a significant difference between the mission cycle times from missions with different results. The job results recorded in the HEMSBase were: stand down (SD), treatment (TR), conveyance (CO) and missed, see Table 2 for a key to the meaning of the job results. The probability of the job result for each vehicle was inferred directly from the proportion of the results for each asset in the HEMSBase. The job results were found to be independent of vehicle, day of the week and time of day.

When fitting the input models we considered entries from the HEMS database between 28th March 2018 and 30th July 2019. In this period NWAA saw 3,426 total jobs and treated 1,796 patients.

3.1 Inferring First Preferences

Although the information in the HEMSBase was very detailed it did not list the first preference vehicle, air ambulance or RRV, for each mission. Whilst the database did record which vehicle actually completed the mission we had no data on instances when the preferred vehicle was busy and another vehicle was sent in its place. We therefore had to infer the preference for each vehicle and the second (and sometimes third) preference if that asset was unavailable so we could mimic the correct job allocation behaviour within our DES model. Note that we assume that missions with first preference of a helicopter will only be reallocated to other helicopters and not RRVs and vice versa. To estimate first preferences we used two distinct but complementary approaches. The first approach was data driven; first preferences were inferred by considering the proportion of times each vehicle was allocated to a mission at times when all three healthcare teams were free. Intuitively if all assets are free then the chosen asset will have been first preference. The arrival rate when everything is available will also be a "true" arrival rate with no rejected jobs (this deals with the general modelling problem of estimating an arrival rate when some customers do not enter the system when it is busy). Second preferences were calculated similarly from the missions observed when a single asset, say the doctor helicopter (H72), was busy. At times when H72 is busy it cannot be allocated to missions; some of the missions that arrive during the times H72 is busy will have had first preference H72, but not all. To estimate the second preference proportions for H72, say the probability H75 was the second preference, we first use the estimated first preference proportions for all the other assets and remove these jobs and then look at the proportion of remaining jobs that were done by H75. This process was repeated for all assets. Third preferences were estimated similarly, but we noticed that in the HEMSBase when two assets were busy a number of jobs were rejected. We therefore also allowed for a small probability of rejecting a mission when two of the three assets were busy.

Our second approach to inferring the first choice preferences considered the utilisation rate, U_i, of each asset. The basis of this method is to recognise that the rate, J_i, at which an asset i undertakes a job is a linear function of the total arrival rate of jobs λ, asset utilisation U_i and the asset preference probabilities. Let π_i denote the first preference probability for vehicle i and π_{ki} denote the probability asset i is the second preference when asset k was the first, then

$$J_i = \lambda \pi_i (1 - U_i) + \sum_{k \neq i} \lambda \pi_k U_k \pi_{ki}.$$

Therefore, since the utilisation rate, U_i, is just the rate of jobs undertaken, J_i, multiplied by the mean service time of the asset, denoted τ_i, for each asset we have

$$U_i = J_i \tau_i = \left(\lambda \pi_i (1 - U_i) + \sum_{k \neq i} \lambda \pi_k U_k \pi_{ki} \right) \tau_i.$$

This can be expressed in matrix format for the three helicopters as follows

$$\lambda \begin{pmatrix} (1-U_{H72})\tau_{H72} & U_{H75}\,\pi_{H75H72}\,\tau_{H72} & U_{H08}\,\pi_{H08H72}\,\tau_{H72} \\ U_{H72}\,\pi_{H72H75}\,\tau_{H75} & (1-U_{H75})\tau_{H75} & U_{H08}\,\pi_{H08H75}\,\tau_{H75} \\ U_{H72}\,\pi_{H72H08}\,\tau_{H08} & U_{H75}\,\pi_{H75H08}\,\tau_{H08} & (1-U_{H08})\tau_{H08} \end{pmatrix} \begin{bmatrix} \pi_{H72} \\ \pi_{H75} \\ \pi_{H08} \end{bmatrix} = \begin{bmatrix} U_{H72} \\ U_{H75} \\ U_{H08} \end{bmatrix}. \quad (1)$$

Taking the inverse of this matrix gives values for π_{H72}, π_{H75} and π_{H08}, but notice that these values are a function of the second preference probabilities, π_{ki}. In practice the final first preference values were not sensitive to the second preference values, and good estimate values for the first preferences were possible using crude values of the second preferences. This secondary approach gave us the opportunity to double

check the first preferences estimated using the data driven method described above by passing the data driven estimates of the second preferences through the inverse of Equation 1.

Once we had the inferred vehicle preferences for each job this allowed us to correct the percentage of arrivals allocated to each vehicle within the DES model and to build in logic to reroute jobs to other assets if the first preference was found to be busy.

3.2 Validation

Before experimenting with our DES model we needed to check that it reflected the real-world behaviour in the current NWAA operational system well; two types of validation exercise were completed to check this. We first preformed a white box validation exercise by discussing the choices behind the model input parameters, model logic and model outputs with representatives from NWAA. After a positive result from this exercise we went on to perform a black box validation exercise by comparing key performance indicators (KPIs) calculated from a single long run of the simulation model to the same KPIs calculated from the missions observed in the HEMBase.

One difference between the simulation model and the true NWAA system is that the simulation model assumes that all assets are available for the same time period each day. In reality, although NWAA planned to start missions at 7am and end by 7pm every day, when we looked into the HEMSBase we found that the number of helicopters on duty often fluctuated in the first and last few hours of the day. For validation purposes we therefore focused on the hours where NWAA offer 'full capacity' coverage i.e. all three healthcare teams are on duty at once. This narrowed our validation interval to 11am-5pm. A warm up period from 9-11am was used within the simulation model as in reality the system is unlikely to start from empty at 11am. Considering the KPIs of accepted jobs, stand downs and jobs accepted by a helicopter per week Table 3 describes the output of the simulation model compared to the observed missions from the HEMS database. It is clear that although there are differences between the KPIs from the real-world data and the DES model the results are close.

Table 3: Blackbox validation of the DES model for 'full capacity' NWAA system between 11am-5pm.

KPI	HEMS database	Model output (CI)
Accepted jobs	32.89	36.46 (36.37, 36.55)
Stand downs	14.81	17.66 (17.60, 17.73)
Completed by helicopter	31.43	27.91 (27.85, 27.97)

Another validation measure was to consider the utilisation of each vehicle, comparing the utilisation profile observed in the simulation to the actual utilisation seen in the HEMS data base. Figure 2 shows the utilisation profile for the three helicopters by call sign (H72 = Barton doctor helicopter, H75 = Barton paramedic and H08 = Blackpool paramedic) and the busiest of the RRVs (HX01 = Barton doctor RRV). It is clear that the utilisation profiles for each of the vehicles shown in Figure 2 match the utilisation observed in the HEMS database very well.

We are satisfied that we have a valid model of the current operation system for NWAA. We shall now use the DES model to provide some insight into using an air ambulance compared to a RRV at night.

4 PRELIMINARY EXPERIMENTS

Recall that our aim is to investigate which assets NWAA should provide at which times of the day. Towards this we will now present our preliminary results for comparing the use of an air ambulance or a RRV for night-time missions. We again state that the results presented here are preliminary and part of a wider project that is a work in progress. We start by comparing KPIs when both vehicles are assumed to perform the same at night as they do during the day (no change to the mission cycle time distributions). We then consider adding a constant to each mission cycle time to account for the possible increase in mission length

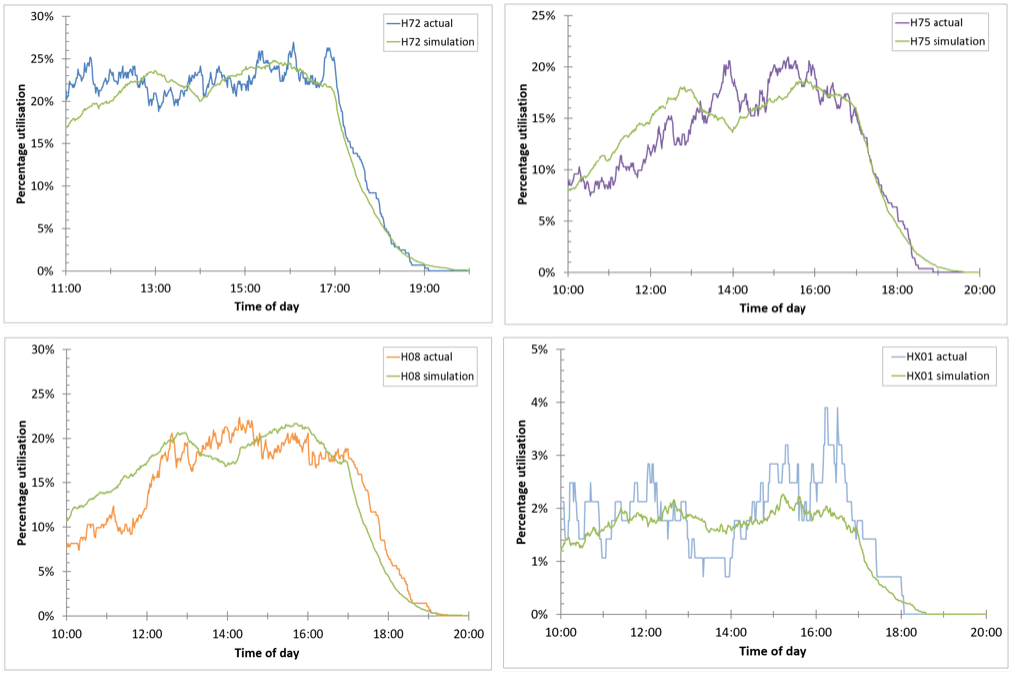


Figure 2: The utilisation profile for the three helicopters (H72 = Barton doctor helicopter, H75 = Barton paramedic and H08 = Blackpool paramedic) and HX01 the Barton based RRV with the doctor).

for both vehicles at night. Note that to use an air ambulance during the hours of darkness would require NWAA to invest in a new helicopter with night time flying capabilities; using a RRV would not incur the cost of a new vehicle.

Discussions with NWAA highlighted two possible shift patterns of interest that would incorporate night time missions. The first involved two 10 hour shifts per day (7am-5pm and 5pm-3am) and the second involved two 12 hour shifts per day (7am-7pm and 7pm-7am). In this experiment our interest lies in the operational performance of NWAA during the night shift we therefore only report our KPIs for missions undertaken during the night. For the purpose of our experiments in the second shift of the day, we assume that a single asset is on duty; this could be either an air ambulance or RRV. We also assume that there is always an asset available at the start of the night shift i.e. the night shift starts from empty with no warm up required. This is a reasonable assumption as we expect the utilisation of the assets to be low at the end of the day, and the health care team would need to swap over between the day and night shift.

Since there were no observations in the HEMSBase on arrivals to the NWAA system at night we used a second database to infer the number of missions that NWAA would be called out to. This database was provided by the North West Ambulance Service (NWAS). To infer the night time arrival profile we calculated the proportion of jobs completed by NWAA within the NWAS database within the 'full capacity' hours (11am-5pm); it turned out that approximately 5% of NWAS jobs were completed by NWAA during that period. We therefore assumed the same proportion of the NWAS jobs would be completed by NWAA at night which allowed us to create arrival profiles for the 10 hour shift (5pm-3am) and the 12 hour shift (7pm-7am).

After speaking to experts from NWAA about this assumption it was suggested that the proportion of NWAS jobs completed by NWAA would change during the night due to more serious incidents often occurring in the early morning hours, particularly in densely populated (inner city) areas. We therefore adjusted the estimate arrival profile to take into account the proportion of serious incidents that occurred in the NWAS database at night. This resulted in an arrival profile with an average of 5.5 emergency calls per

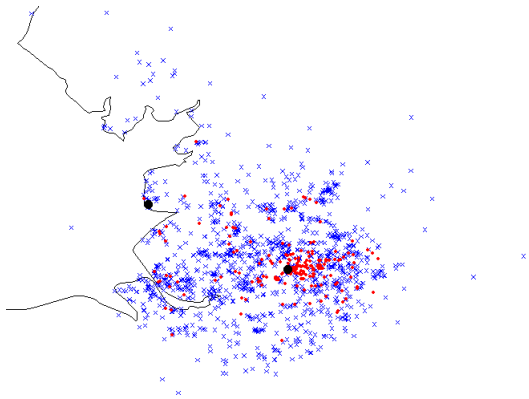


Figure 3: A map of mission locations for the H72 air ambulance (blue) and HX01 RRV (red) during the day. The Barton and Blackpool bases are also marked (black).

night for the 10 hour night shift and 4.8 emergency calls per night for the 12 hour night shift. We assume that as there is only one asset available in the evening all arrivals would be allocated to this resource during the night shift. If that resource is busy then the job is missed. We also assume that the healthcare team available to the asset throughout the night is the highly skilled doctor/consultant and paramedic paring. The NWAA asset on the night shift is therefore equivalent to either the Barton doctor helicopter (call sign: H72) or the Barton doctor RRV (call sign: HX01).

Estimation of the effect of night-time flights on the mission cycle-time distributions used within our model were harder to infer. Discussions with NWAA highlighted that helicopter flights at night would take longer to prepare for take off due to additional checks/ planning being required, and would also take longer to land at scene due to the larger landing zone required in the hours of darkness. McQueen et al. (2015) also discuss the problem of many hospital helipads in the UK lacking the infrastructure for night time landings which again would add additional time on to any conveyance mission. For these reasons we suspected that night time mission cycle times would need to be increased, but we had no data to infer by how much we should increase them. Using the current mission cycle time distributions would clearly give an optimistic view of the air ambulances performance as these distributions were fitted from missions that occurred during the day.

We also had little information on the expected mission cycle time of a RRV during the night. NWAA performed a short experiment in January 2019 that saw a their fast response land vehicle respond to incidents during the night, but this experiment coincided with filming for a BBC series and only few missions were accepted. We therefore did not use the data from this experiment to fit our mission cycle time distributions. Inferring the mission cycle time distributions having no data is a difficult problem. Assuming both the air ambulances and RRVs would be based at the Barton site at night we started by comparing the geographical locations of the missions that H72 and HX01, the vehicles with doctors, completed during the day, see the red points in Figure 3. It is clear from Figure 3 that the RRV tends to attend missions close to the Barton base. Using the current mission cycle time distribution for HX01 should therefore be thought of as being overly optimistic as the travel time is likely to be higher at night when the RRV has to cover a wider area.

Despite both asset mission cycle time distributions being overly optimistic the natural preliminary experiment to perform was a comparison of the two assets at night assuming they could perform as they did during the day, see Table 4. Note that for this experiment we set the proportion of stand down, treatment and convey results to be the same for each asset as the job result is likely to be the property of a mission not the type of vehicle; both assets therefore saw the same proportion of job results. The job result proportions were inferred from the day time missions for the H72 helicopter. This preliminary experiment was replicated $n = 50$ times. In Table 4 we report the mean result of the replications along with a 95% confidence interval.

Table 4: Comparing KPIs for night time missions undertaken by either a single helicopter or a single RRV assuming night and day time mission cycle time distributions are the same. The KPIs reported are the average of 50 simulation replications; they are reported with 95% confidence intervals.

	10 hour shift		12 hour shift	
	Helicopter	RRV	Helicopter	RRV
Total jobs / wk	38.6 (38.44, 38.75)	38.0 (37.90, 38.11)	34.0 (33.87,34.17)	33.2 (33.15,33.24)
Missed jobs %	30.5	28.0	25.1	22.5
Missed jobs / wk	11.8 (11.67, 11.84)	10.6 (10.54, 10.65)	8.5 (8.46, 8.62)	7.46 (7.40, 7.51)
Accepted jobs %	69.5	72.0	7.49	77.5
Accepted jobs / wk	26.8 (26.75, 26.93)	27.4 (27.34, 27.48)	25.5 (25.38, 25.59)	25.7 (25.68, 25.80)
Stand down %	48.1	48.2	48.1	47.6
Stand down / wk	12.9 (12.84, 12.97)	12.9 (12.87, 12.98)	12.25 (12.19, 12.32)	12.26 (12.20, 12.32)

Table 4 illustrates that if both assets were to perform as they do during the day shift then they would have very similar performance during the night. Note that during the night shift we attempt to allocate all of the missions to a single asset. The number of missed jobs is therefore likely to be higher than in the day when three assets are available. This is reflected in the results, even with both assets performing as they do in the day, which is optimistic as discussed, we see the percentage of missed jobs varies between 22-30%. This indicates that another resource might be useful during the night shift. The experiment also highlights that the chosen shift pattern matters. In Table 4 we see that the 12 hour shift has fewer jobs per week on average compared to the 10 hour shift. This occurs because 5-7pm is a busy period for arrivals and only the 10 hour shift encompasses this period; the 12 hour shift starts at 7pm at the end of the busy period.

Our second experiment investigates increasing all mission cycle times by a constant to account for the possibility that both vehicles are likely to need take longer to complete a mission during the night. The choice of parameters for the mission cycle time distributions for the air ambulance and RRV for our second experiment stem from our conversations with NWAA. During our meetings NWAA representatives expressed the view that the travel time to an incident for the air ambulance and RRV may be roughly equivalent during the hours of darkness due to the additional time required to plan the mission, for take off and to land the helicopter at scene. Clearly if the mission cycle times were equivalent for the air ambulance and RRV there would be little need to invest in the new helicopter.

Since we cannot quantify how the mission cycle time distributions would change during the night for either vehicle, we instead consider incremental increases in all mission cycle times. This aims to give NWAA an idea of how the assets might perform if they were a certain amount slower on each mission than the H72 helicopter during the day. Let us denote the mean mission cycle time of the air ambulance during the day by μ (for H72 $\mu \approx 49$ mins). Our next experiment considers our KPIs when all mission cycle times of H72 are increased by a constant of either $10\%\mu$, $50\%\mu$ or $100\%\mu$ i.e. all missions were either 4.9, 24.5 or 49 minutes longer.

We chose to increase all mission times by a constant rather than a percentage as the type of vehicle and time of day are only likely to affect the travel time to scene and not other components of the mission cycle time; the addition of a constant amount onto all mission cycle times therefore seemed more appropriate. This experiment will allow us see how sensitive our KPIs are to the mission cycle time distributions. It will also allow us to judge whether the air/ RRV provide a reasonable service even if they have much longer mission cycle times. Future work is needed to better estimate the mission cycle times of both assets.

The results of the experiments for the two night time shift patterns are displayed in Tables 5 and 6. All experiments were replicated $n = 50$ times, the results in table 5 and 6 are the average of these results reported with 95% confidence intervals. Common random numbers were used across all experiments.

Table 5: KPIs for the 10 hour night shift missions undertaken by a single RRV assuming all the night mission cycle times are 10%, 50% and 100% greater than the current mission cycle times of the helicopter. The KPIs reported are the average of 50 simulation replications; they are reported with 95% confidence intervals.

	Helicopter +10%μ	Helicopter +50%μ	Helicopter +100%μ
Total jobs / wk	38.74 (38.57, 38.91)	38.74 (38.57, 38.91)	38.74 (38.57, 38.91)
Missed jobs %	32.5	39.6	46.4
Missed jobs / wk	12.59 (12.50, 12.69)	15.33 (15.22, 15.44)	17.97 (17.85, 18.09)
Accepted jobs %	67.5	60.4	53.6
Accepted jobs / wk	26.15 (26.06, 26.23)	23.41 (23.33, 23.49)	20.77 (20.71, 20.83)
Stand down %	48.2	48.1	48.1
Stand down / wk	12.60 (12.54, 12.66)	11.27 (11.21, 11.32)	9.99 (9.94, 10.05)

Table 6: KPIs for the 12 hour night shift missions undertaken by a single RRV assuming the night mission cycle times are 10%, 50% and 100% greater than the current mission cycle times of the helicopter. The KPIs reported are the average of 50 simulation replications; they are reported with 95% confidence intervals.

	Helicopter +10%μ	Helicopter +50%μ	Helicopter +100%μ
Total jobs / wk	34.02 (33.87, 34.17)	34.02 (33.87, 34.17)	34.02 (33.87, 34.17)
Missed jobs %	26.8	33.3	39.7
Missed jobs / wk	9.13 (9.05, 9.21)	11.33 (11.24, 11.42)	13.51 (13.41, 13.62)
Accepted jobs %	73.2	66.7	60.3
Accepted jobs / wk	24.81 (24.72, 24.90)	22.69 (22.59, 22.78)	20.51 (20.43, 20.59)
Stand down %	48.1	48.1	48.0
Stand down / wk	11.92 (11.87, 11.98)	10.91 (10.85, 10.97)	9.84 (9.78, 9.90)

In Table 5 we report the results of our experiment for the 10 hour night shift. We see that as the mission cycle time increases the percentage of missed jobs increases; this was expected as the single asset has higher utilisation so there is more chance of arrivals finding it busy on entry. We also see that although the number of stand downs is decreasing the percentage stays constant as the mission cycle time increases. In Table 6 we see similar behaviour to Table 5. In general notice that if either asset were to act similarly to H72 in the day, then the average number of patients assisted i.e. the average number of acceptances minus the average number of stand downs in Table 4, equates to over 13 missions per week. Comparing this to the worst case we investigated where all missions were 49 minutes slower for either asset, we see that NWAA would still assist over 10 people per week on average. This is a positive result. It indicates that the number of people on average that are helped by NWAA per week during the night is not particularly sensitive to the mission cycle time. In truth we do not believe the use of either vehicle would make missions 49 minutes longer. The lack of sensitivity of the KPIs to the mission cycle time indicate that the RRV may be the better choice of vehicle due to the additional cost and risk involved in night time flights. From Tables 5 and 6 it appears that the number of missed jobs is more of an issue than the mission cycle times themselves. This indicates that, if the arrival profile is accurate, a second asset might be useful at night.

Whether an air ambulance or a RRV is the asset of choice, by looking at the arrival profile information alone this investigation has told us that NWAA are likely to see over 30 calls during the night per week. This number is high and we believe motivates the need for some form of emergency health care provision through the night.

5 CONCLUSIONS AND FURTHER WORK

This paper considers the need for an air ambulance vs. a RRV to provide emergency health care through the night in the Northwest region. In conclusion to our preliminary investigation we believe that some form of night time emergency health care provision is required due to the large number of jobs estimated to fall during the night that would meet the HEMS dispatch criteria. We also believe that the more appropriate asset is the RRV at present due to the lack of sensitivity of our KPIs to mission cycle time and the additional risk and cost associated with night time flights. Although at the time of writing the paper the project was ongoing, and further analysis/ experimentation is required to investigate when and where would be most appropriate to provide this service.

Going forward, we also intend to consider the addition of a HEMS RRV to be based roadside within the region during the night shift, as the number of missed jobs was reasonably high even when we were optimistic about the mission cycle time distributions. Consideration of the time of year is also of interest and how might this might effect which asset is of most use as it is likely that time of year effects the night shift more than the day shift. Finally, consideration of how the job result distributions may change during the night should be investigated. Lyon et al. (2015) suggest the patient conveyance needs may be higher at night due to the increased injury severity.

On a different note, research into how the HEMS dispatch criteria might be modified at night, and the knock on effect of this may have on the NWAA provision should be investigated. Both (Lyon et al. 2015) and (McQueen et al. 2015) only consider missions involving trauma patients during the night whereas within our model we took into account all incidents that meet NWAA dispatch criteria during the day.

ACKNOWLEDGMENTS

We gratefully acknowledge Lanner for allowing us to use the simulation software WITNESS.

REFERENCES

CAA 2019. "CAP1535: The Skyway Code". https://publicapps.caa.co.uk/docs/33/CAP1535_Skyway_Code_V2_PRINT.pdf. Online; accessed 28 October 2019.

Curtis, L., M. Salmon, and R. M. Lyon. 2017. "The impact of helicopter emergency medical service night operations in south East England". *Air medical journal* 36 (6): 307–310.

Lyon, R. M., J. Vernon, M. Nelson, N. Durge, M. Tunnicliff, L. Curtis, and M. Q. Russell. 2015. "The need for a UK helicopter emergency medical service by night: a prospective, simulation study". *Air medical journal* 34 (4): 195–198.

McQueen, C., T. Nutbeam, N. Crombie, F. Lecky, T. Lawrence, K. Hathaway, and S. Wheaton. 2015. "Enhanced care team response to incidents involving major trauma at night: Are helicopters the answer?". *Injury* 46 (7): 1262–1269.

AUTHOR BIOGRAPHIES

LUCY E. MORGAN is a Development Lecturer in Simulation and Stochastic Modelling in the Department of Management Science at Lancaster University. Her research interests are input uncertainty in simulation models and arrival process modelling. Her e-mail address is l.e.morgan@lancaster.ac.uk.

ROGER J. BROOKS is a lecturer in the Department of Management Science at Lancaster University. His research areas of interest include conceptual modelling, agent-based simulation, applications of simulation in health, and statistical analysis of sports data and health data.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.021

DEVELOPING A DISCRETE EVENT SIMULATION MODEL USING QUALITATIVE AND QUANTITATIVE DATA SOURCES

Heather Callaghan

Loughborough University
University of Nottingham Ningbo (UNNC)
Heather.Callaghan@nottingham.edu.cn

Dr. Antuela Tako

Loughborough University
Epinal Way, Loughborough, LE11 3TU
A.Takou@lboro.ac.uk

Professor Lisa Jackson

Loughborough University
Epinal Way, Loughborough, LE11 3TU
L.M.Jackson@lboro.ac.uk

Dr. Sarah Dunnett

Loughborough University
Epinal Way, Loughborough, LE11 3TU
S.J.Dunnett@lboro.ac.uk

ABSTRACT

This paper will discuss the development process of discrete event simulation models with regards to using data from multiple sources, that may be gathered both quantitatively and qualitatively, before being incorporated into a single simulation model. The aim of this paper is to more formalise the less discussed qualitative practices in simulation modelling. There will be a brief overview of where this has been touched upon previously in literature, before moving on to the most commonly occurring research methods and types of data gathered; and how these can be incorporated into a model. This paper will go on to consider potential benefits and drawbacks to this approach before presenting an application of how this thinking was applied to a research project.

Keywords: Discrete Event Simulation, Mixed Methods, Police Custody

1 INTRODUCTION

Computer based simulation models are developed using a variety of information from different sources. In many published papers that review the development of an operational research-based simulation model for an application, the quantitative dataset is often the only data source discussed. There is little mention of the other data sources that contribute to the development and testing of discrete event simulation models. However, in order to develop an accurate imitation of a system, the modeller must first have some understanding of this system, such as the inputs, outputs, order of processes, etc. This information cannot usually be gleaned from quantitative data alone, so it stands that the modeller must have learned about the system from an alternate source. This could be from conversations with managers/operators within the system, from spending time observing the system, from previous literature, etc. Experienced simulation modellers often incorporate this information into the simulation modelling process intuitively and with little discussion. Consequently, there is little guidance on how this can be achieved, or how the information can be extracted for novice modellers. This paper will attempt to will describe the data collection needs and to define the process for modellers to follow, so as to further understand how, as modellers, we develop our simulations. This paper will specifically look at discrete event simulation modelling and provide an example of an application of the process.

2 LITERATURE REVIEW

A simulation is a computer-based model that aims to replicate a real-life situation, where the modelled can experiment with different parameters to better understand how the system may behave under different circumstances (Law 2007; Robinson 2004; Pidd 2004). In order to build such a model, the relevant system is broken down into its components; including but not limited to, the service receiver, the resources available within the system and the tasks/stages that make up the system as a whole. There are various simulation modelling approaches including agent-based modelling, continuous modelling, and discrete event simulation. A system dynamics (or continuous modelling) simulation model represents environments through stocks and flows and is applicable to a wide range of situations from engineering to socioeconomics. Agent based modelling works by modelling the 'agents' within a system and their behaviours and interactions and is commonly used in modelling pharmacological systems. Monte Carlo simulation focuses more on the outcomes of scenario and is most often used to model risk. However, this paper is going to focus specifically on discrete event simulation modelling (DES). A DES model 'models the operation of a system as a discrete sequence of events in time' (Sharma, 2015). Once an event or activity ends, the simulation moves onto the next activity, and this repeats until the service receiver has reached the end of the activities relevant to their path and they exit the system. The completion time for each stage is commonly based on statistical distributions, derived from a quantitative dataset, so as to accurately imitate the length of time the service receiver spends in the system.

Pidd (2004) describes the 3 types of data collection for the purpose of modelling – contextual data, model realisation and model validation. Contextual data is information used to understand the system being modelled, model realisation data is required to develop the model and model validation data used to check the model is fit for purpose. Contextual data is primarily what this paper is discussing for the use of qualitative data, however depending on the situation being simulated, this discussion may also apply to the model realisation data and model validation.

Mixed methods research has multiple definitions but for the purpose of this paper will be defined as the type of research in which a researcher, or a team of researchers, integrates qualitative and quantitative approaches within a single study or a set of closely related studies (Creswell and Plano-Clark, 2007; Johnson et al, 2007) Research methods can usually be classified into either quantitative or qualitative quite clearly. Quantitative methods are generally considered to be empirical studies where variables can be reliably measured; whereas qualitative methods are variables that are better described rather than measured (Newman et al., 1998). The difference is that with simulation modelling, whilst usually considered to be quantitative, there are the multiple stages of data collection, as defined above, that contribute to the overall simulation model, and these are not necessarily all quantitative.

There is some literature regarding how qualitative data can be included in a discrete event simulation model. One example of this is Partisim (Kotiadis and Tako, 2015), a framework for conducting facilitated workshops with the various stakeholders within a system and how the information gleaned from these sessions can be incorporated into a simulation model. On the subject of problem definition, the first stage of developing a simulation model, Kotiadis (2007) discussed how Soft Systems Methodology can be used to define study objectives. The ideas in these papers are quite specific, whereas this paper aims to give a more general approach to qualitative and quantitative data in a simulation model.

There are seven generally accepted stages of discrete event simulation modelling – problem definition, conceptual modelling, data collection and analysis, model development, validation and verification, experimentation, and implementation (Tako, 2011). Each one of these stages requires data to progress and move on to the consequent stage of the process. In the next sections, these stages will be broken down as to the type of data that is required and the possible sources to obtain that data from.

3 DATA SOURCES

Incorporating the qualitative research into a simulation framework is something that has been touched upon in literature and specific frameworks developed (Partisim, etc.) but there is minimal general guidance on how to blend qualitative research with a quantitative simulation approach for a novice modeller. With the data collected through the various methods as outlined above, the findings can be

applied at the different stages of modelling, to fill in or expand on the knowledge required for a more accurate simulation model. It is worth noting that due to the iterative nature of simulation modelling, data collection may take place at any point in time during the model development process, even this data could be used in the initial modelling stages such as problem definition.

The most commonly discussed data used to develop a simulation model is the quantitative type of data often used for the model realisation and model validation. For a discrete event simulation model, this can be a series of times activities take place, a timetable/count of available resources, etc. However, contextual data required for understanding the situation and for developing a simulation model, normally part of the earlier modelling stages, may come from alternative sources as it is difficult to gain a sufficient understanding of the system solely by viewing the quantitative dataset. For this at least one further source of data is required.

Data sources can be either qualitative or quantitative or both and can be collected formally, by following the traditional data collection techniques, e.g., ethnographic observations, structured/semi-structured interviews, scientific experiments, questionnaires, etc., or informally, for example through general meetings, conversations, etc. Examples of data sources for simulation could be a meeting with a stakeholder/manager where the system and the aims for the simulation model are discussed, ethnographic observations conducted by the researcher, discussions with various staff within the system, etc. While the data the researcher takes away from these actions may not be as obviously impactful as a large dataset, it can still provide context, and result in amendments made to the model throughout the process, thus contributing to the simulation process.

4 MODELLING STAGES

As discussed in the previous sections, the modelling process can be broken down into seven stages and the possible incorporation of data at each stage will be discussed below, but it should be noted that this list is not exclusive. Depending on the system being modelled and the aims of the research, there may be further possibilities to incorporate data from alternate sources.

Problem definition is the first stage in developing a simulation model, where the aims of the project and the system being modelled are defined. In order to narrow down the problem, the modeller must acquire some relevant knowledge of the system being modelled and the situation within the system that requires attention. This can come often from a stakeholder as an issue that they are having in the form of a conversation – whether by email, in person, etc. In this manner a modeller is receiving data relevant to their model that will assist them in eventually developing their simulation. This data can often initially be considered qualitative – close to the form of an informal interview or survey. A modeller then may perform some simple analysis to identify themes within the responses and further define the problem and narrow down their focus and aims for the model.

Conceptual modelling can be defined as ‘a non-software specific description of the computer simulation model (that will be, is or has been developed), describing the objectives, inputs, outputs, content, assumptions and simplifications of the model’ (Robinson, 2008). The data to develop a conceptual model can be gathered from a multitude of sources, for example, through observations. The quantitative dataset can be used to deduce the stages of a system and its inputs and outputs, but qualitative data may give the modeller a better overview of the different stages of a system and an idea of what data is recorded throughout the process and how the entities flow through the process and the different paths they may take.

The next stage is data collection and analysis. This is required so as the conceptual model can be actualised into the computer model. Clearly the quantitative dataset is key at this stage and will principally be featured in the analysis. However, the qualitative, background data can be used to decide what quantitative data is available or useful.

Developing the computer simulation model is a combination of the conceptual modelling stage and the quantitative data analysis, so the data that was used in these stages is indirectly used and combined for this stage.

Validation and verification are necessary to ensure the model is fit for purpose. There are various ways to validate and verify a simulation model, as documented in literature. Verification in this paper will be defined as “ensuring that the computer program of the computerized model and its implementation are correct” (Sargent, 2013) and can be assessed through comparison with analytic

results, input-output combinations, etc. Qualitative approaches to validating the computer model is correct can include having it checked by shareholders through focus groups, interviews and other research methods featuring interactions between modellers and participants and can be defined as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model" (Thacker et al., 2004).

Experimentation is the testing of various scenarios using the simulation model to achieve the objectives defined for the study. Deciding the appropriate scenarios to test is one of the key challenges for a modeller, as the results from these experiments determine the outcomes and conclusions drawn that inform the stakeholders. The scenario testing needs to be robust enough for the modeller to have confidence in their results, but specific enough that they are still of use to the shareholder. Determining these scenarios can be deduced from a quantitative dataset, however in order to do this the modeller would need to have a clear understanding of the system, which comes from the contextual data. Observations of the system can give the modeller an idea of what scenarios may need to be tested, and discussions with people within the system (through interviews, meetings, focus groups, etc.) can provide the modeller with direction and/or confirmation of which scenarios to test and on what scale.

While implementation of the results ordinarily falls to the stakeholders of the project, the modeller has responsibility for the recommendations provided based on their simulation model. Having a good understanding of the system and its limitations, through contextual data, can aid the modeller in providing more realistic recommendations to managers, etc. of the system. This ultimately makes the model more useful and valuable to stakeholders.

5 BENEFITS AND DRAWBACKS OF USING QUALITATIVE DATA

The benefits to using a mixed method approach to research has been widely discussed (Amalki, 2016; Mayoh and Onwuegbuzie, 2013, etc.), so this paper will discuss the benefits of incorporating qualitative data into the simulation modelling process, as discussed above. One of the clear advantages is that the modeller would have more detailed knowledge of the system being modelled and a better understanding of the problem being explored. This can allow the modeller to determine what is more useful for the problem being modelled and exclude aspects deemed irrelevant leading to a, potentially, more relevant simulation. However, the possible danger in doing this, is that if irrelevant information is not removed, the simulation model may become overcomplicated with too much detail making the results harder to analyse for purpose and the model slower to run on the computer (Robinson, 2004).

Another benefit to this approach is that using multiple types of data/data sources can offset the weaknesses of a single source or data type. Quantitative data sets alone may not give a modeller the full picture, and qualitative data can be used to fill in the gaps. Multiple data sources may also aid to limit the bias in the data and any bias from the researcher. The obvious drawback to collecting data from multiple sources, is that it can be time consuming. There may also be data gathered that is a repeat of previous data, so the modeller may be duplicating work. Alternately data from two sources may produce conflicting data, in which case the modeller must decide how to mediate this.

One of the difficulties in this method of working is that it may be difficult for the modeller to gain access to certain data, or it may take more resources to gather. This may be particularly true for qualitative data when using a method such as interviews or focus groups, which can be hard to arrange or incentivize. However, it does offer the opportunity to include the stakeholders and people who work within the system to have more involvement in the modelling process. This can help stakeholders to remain interested in the research, and to be more motivated to implement results.

6 APPLICATION EXAMPLE

We will now present a case study of how this thinking was applied in practice to develop a police custody simulation model. With budget cuts and resource reduction in UK police forces, it is necessary for the forces to manage their resources effectively, so as to meet demand. Police custody is just one of the aspects of policing that has faced shortages due to the budget cuts. A discrete event simulation model of a police custody suite was developed following the approach discussed previously for the purpose of resource optimization. Previous discrete event simulation models were developed of police custody (Greasley, 1998, 2000 and 2001), one of which addressed resource allocation. However, these

were before the current issues, and the system of police custody has faced multiple changes since these models were developed. This research was conducted to update and provide a deeper understanding of this system. The system was broken down into the service receiver (detainees), resources (cells and staff) and the tasks that occur (e.g., booking-in, interviewing etc,). The primary objectives were to model the system and its resources to see if there were any bottlenecks, and if the resources could be more effectively used to meet demand.

6.1 Data Sources

Due to the limited literature available regarding the stages and resources in police custody, combined with the modellers lack of knowledge in this area, it was clear that substantial contextual data was required. Having viewed the quantitative data that was available for the research, it was considered sufficient for model realization data, but it did not provide enough context, so alternative data sources were sought, and quantitative research conducted. The stages of simulation followed are outlined in Figure 1 below, along with the data sources used.

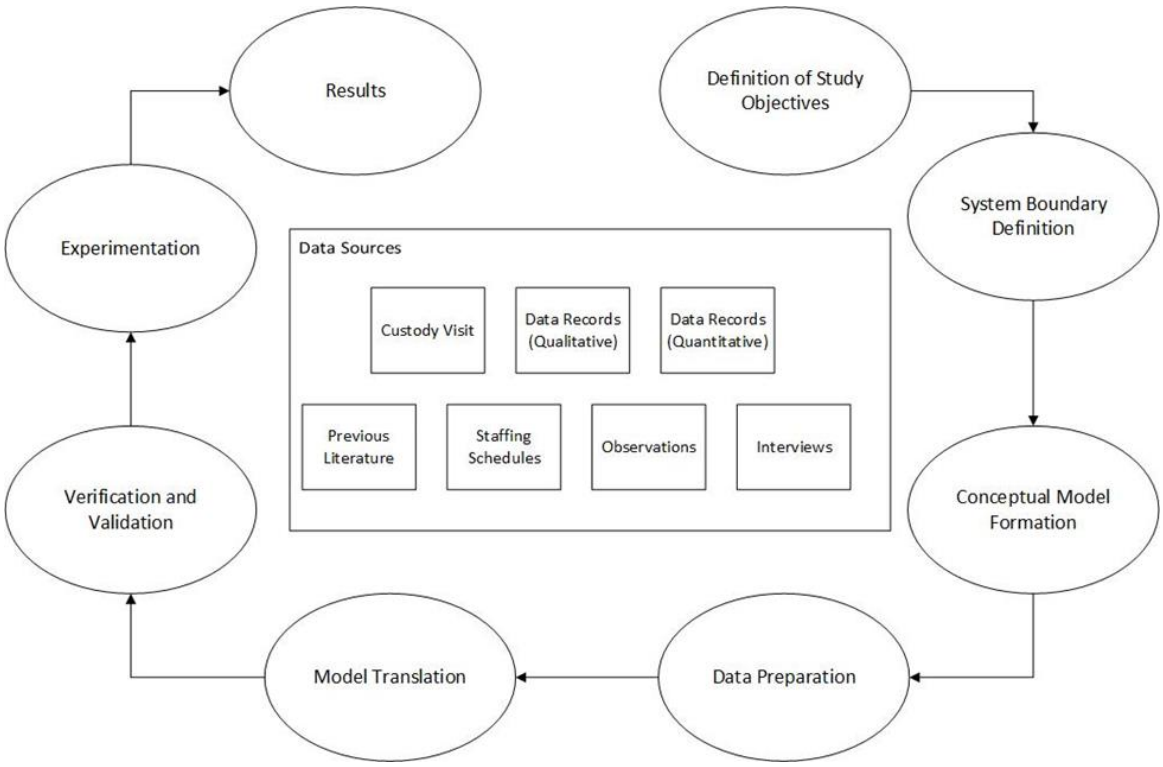


Figure 1. Data Sources and Simulation Process.

The data sources, as specified above, were a custody visit, the data records – both qualitative and quantitative information, previous literature, staffing schedules, observations, and interviews. When the project was first initiated, a visit and tour of a custody suite was organized first. During this, officers explained the processes and stages of police custody and answered any questions about the system. The data records provided by the involved police forces contained quantitative timings for each of the stages being modelled, as well as some qualitative data in the form of written comments attached to the relevant stage. This dataset did not provide data on the types of resources available in the system, or which resources were required at each stage of the process. Staffing schedules were also provided. The previous literature (Greasley, 1998, 2000 and 2001), provided some contextual data of the system but due to time of publishing of these papers, this data was considered informative but outdated. Observations conducted in police custody took place in multiple suites, over a period of months on varying days and times. This was to ensure the modellers view of the system was not biased by a

particular suite or time of day, so the modelling was more robust. Interviews took place with staff members in every role within a custody suite, in different custody suites, again to reduce bias.

6.2 Modelling Process

There were a few aspects to the problem definition stage of modelling – defining the system to be modelled, the issue to be explored and the objectives of the study. In this study the data sources used to define these were the custody visit, the observations, the interviews, and the previous literature. The previous literature (including the custody inspection reports) were used to give a general view of the situation within custody and how issues had been tackled previously. The custody visits and observations allowed for the system to be viewed and observed to help define the system from the researchers' point of view, whereas the interviews took into account the problems as considered by the experts. This stage of the modelling process was conducted entirely through qualitative data; it would have been very difficult to complete it without this.

The next stage was the conceptual modelling stage. The data sources used for the conceptual model of police custody were previous literature, custody visit, the qualitative aspect of the data records and the observations. The previous literature gave a base model to build on with further data gleaned about the system from the additional data sources. The qualitative data gleaned for the data set and the custody visit helped to update this initial model. During the observations attention was paid to the paths detainees took through the system and the resources that were involved at each stage, which aided in fleshing out the conceptual model. This stage, again, was completed almost entirely from qualitative data.

Data collection and analysis were conducted next. The quantitative data records are clearly an integral part of this stage; however, the previous literature and custody visit were used to help define what data specifically needed to be collected. Based on the previous literature, particularly Greasley (1998) where resource allocation was discussed, a general idea could be gathered of the stages of custody, as they had used data to model these stages. The custody visit gave a tour of custody and the shareholders explained what data was collected at each stage, making it easier to choose what was relevant and available.

Developing the computer simulation model was a combination of the conceptual modelling stage and the quantitative data analysis, so the data that was used at this stage is indirectly used for this one too. The staffing schedules were also incorporated into the model when considering resource levels and availability.

Validation and verification are necessary to ensure the model is fit for purpose. In this instance, a second smaller quantitative dataset was analysed and measured against the simulation model for validation. Through observations, the modeller was able to verify the model and in interviews, the model was presented to the staff in the various roles within custody, for them to verify as well. The quantitative aspect to this stage made the verification stage more robust, in that it was verified by multiple people within depth knowledge of the system being modelled.

The experiments, the testing of various scenarios using the simulation model, were run with the model developed in the previous stages. The scenarios that were tested were deduced using data obtained through interviews and observations. The input of people working within this system was invaluable at this stage, as it provided guidance and much more insight the modeller could have gained alone.

This research is still ongoing, but in the final stage of implementation, it is believed that the contextual data gathered from the custody visit, observations, and interviews, will give the modeller a better sense of judgement as to what recommendations were realistic, and more likely to be implemented in practice.

7 CONCLUSION

In conclusion, discrete event simulation modelling is generally considered to be a quantitative research method, but this paper has discussed how the data used in developing a simulation model for an application usually comes from multiple sources, particularly the contextual data, and tried to further the discussion in how this can more formalized. These data sources could be either quantitative or

qualitative and may contribute to the simulation model at different stages. Whilst there are both advantages and disadvantages to using multiple data sources or research methods, there are situations where it can be of benefit, such as the example explained, where there was a lack of contextual data available. The next stage in this research would be to develop a framework on how quantitative and qualitative data can be clearly incorporated into a single process, to develop a simulation model.

REFERENCES

Almalki, S., 2016. Integrating Quantitative and Qualitative Data in Mixed Methods Research--Challenges and Benefits. *Journal of education and learning*, *5*(3), pp.288-296.

Creswell, J.W. and Plano Clark, V.L.2007. *Designing and Conducting Mixed Methods Research.* Sage, Thousand Oaks, CA, USA.

Greasley, A. and S. Barlow. 1998. "Using Simulation Modelling for BPR: Resource Allocation in a Police Custody Process". *International Journal of Operations & Production Management*, 18(9/10):978-988.

Greasley, A. 2000. "A Simulation Analysis of Arrest Costs". *Journal of the Operational Research Society*, 51(2):162-167.

Greasley, A. 2001. "Costing Police Custody Operations". *Policing: An International Journal of Police Strategies & Management*, 24(2):216-227.

Johnson, R.B., Onwuegbuzie, A. J. and Turner, L.A. 2007. Towards a Definition of Mixed methods research. *Journal of Mixed Methods Research.* 1(1) 112-33.

Kotiadis, K. (2007). Using Soft Systems Methodology to determine the simulation study objectives. *Journal of Simulation*, 1:3, 215-222, DOI: 10.1057/palgrave.jos.4250025.

Law, A. M. 2009. "How to build Valid and Credible Simulation Models". In *Proceedings of the 2009 Winter Simulation Conference,* edited by M.D. Rossetti et al., 24 – 33.

Lehaney, B. and R.J. Paul. 1996. The Use of Soft Systems Methodology in the Development of a Simulation of Out-Patient Services at Watford General Hospital. *Journal of the Operational Research Society*, 47(7):864-870.

Mayoh, J. and Onwuegbuzie, A.J., 2015. Toward a conceptualization of mixed methods phenomenological research. *Journal of mixed methods research*, *9*(1), pp.91-107.

Newman, I., Benz, C.R. and Ridenour, C.S., 1998. *Qualitative-quantitative research methodology: Exploring the interactive continuum*. SIU Press.

Pidd, M. 2004. *Systems Modelling: Theory and Practice*. 1st Edition. Chichester: Wiley.

Robinson, S. 2004. *Simulation: The practice of Model Development and Use.* Wiley, Chicchester. UK

Robinson, S. 2008. Conceptual modelling for simulation Part I: definition and requirements. *Journal of the operational research society*, 59(3), 278-290.

Sargent, R. G. 2013. Verification and validation of simulation models. *Journal of simulation*, 7(1), 12-24.

Sharma, P., 2015. Discrete-event simulation. *International journal of scientific & technology research*, *4*(4), pp.136-140.

Tako, A. A. and K. Kotiadis. 2015. PartiSim: A Multi-Methodology Framework to Support Facilitated Simulation Modelling in Healthcare. *European Journal of Operational Research*, 244(2):555-564.

Tako, A. 2011. Model Development in Discrete-Event Simulation: Insights from Six Expert Modelers. In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain et al, 3928-3939. IEEE.

Taylor, S. J., Eldabi, T., Riley, G., Paul, R.J., and Pidd, M. 2009. Simulation modelling is 50! Do we need a reality check?. *Journal of the Operational Research Society*, 60(1), S69-S82.

Thacker, B.H., Doebling, S.W., Hemez, F.M., Anderson, M.C., Pepin, J.E. and Rodriguez, E.A., 2004. Concepts of model verification and validation.

AUTHOR BIOGRAPHIES

HEATHER CALLAGHAN is currently a Mathematics Tutor at the University of Nottingham Ningbo China (UNNC) and completing a PhD in Simulation Modelling at Loughborough University. Her email address is Heather.Callaghan@nottingham.edu.cn.

PROFESSOR LISA JACKSON is the Professor of Risk and Reliability in the Department of Aeronautical and Automotive Engineering at Loughborough University. Her main interests lie in reliability of engineering systems, system safety and enhanced service provision. Her email address is L.M.Jackson@lboro.ac.uk.

DR SARAH DUNNETT is a Senior Lecturer in Control and Reliability in the Department of Aeronautical and Automotive Engineering at Loughborough University. Her research interests include reliability of engineering systems and aerosol science. Her email address in S.J.Dunnett@lboro.ac.uk.

DR ANTUELA TAKO is a Reader in Operational Research at the School of Business and Economics, Loughborough University. She holds a PhD in Simulation and an MSc in Management Science and Operational Research from the University of Warwick. Her research interests include the comparison of simulation approaches (discrete-event simulation and system dynamics), facilitated and participative simulation modelling, conceptual modelling, and health care modelling. She is Associate Editor of the Journal of the Operational Research Society. Her email address is a.a.takou@lboro.ac.uk and her webpage is: http://www.lboro.ac.uk/departments/sbe/staff/anthi-antuela-tako/.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.022

GENERATING INSIGHT IN DISCRETE EVENT SIMULATION AND AGENT-BASED MODELLING: EXPERIMENTAL EVIDENCE

Hilya Mudrika Arini

Universitas Gadjah Mada
Yogyakarta, Indonesia
hilya.mudrika@ugm.ac.id

Nurul Lathifah

Universitas Gadjah Mada
Yogyakarta, Indonesia
nunurullathifah@gmail.com

Nur Aini Masruroh

Universitas Gadjah Mada
Yogyakarta, Indonesia
aini@ugm.ac.id

Laela Dewi Maharani

Universitas Gadjah Mada
Yogyakarta, Indonesia
laela.dmaharani@gmail.com

ABSTRACT

This study aims to understand the difference between discrete-event simulation (DES) and agent-based modelling (ABM) user and non-user in generating insight. In total, 41 undergraduate students who were DES and ABM users and non-users participated in the experiment and given task of solving problem using simulation models. The insight generation is measured by four variables: (1) task performance, (2) problem understanding, (3) discontinuity in thinking, and (4) change in problem understanding. This study concludes that there are no significant differences between DES and ABM users and non-users in generating insight particularly for simulation with low complexity as utilised in this study. However, both ABM users and non-users had higher solution rates, indicating higher insight occurrence, than DES users and non-users. Since this study limited to simulation with low complexity, there is a room of improvement for further study to investigate different complexity level of the system simulated in generating insight.

Keywords: Behavioural Operational Research, Generating Insight, Discrete Event Simulation, Agent-Based Modelling

1 INTRODUCTION

Simulation has been used in many service and manufacturing sector projects. It is considered a cost-effective method for improving, investigating, and evaluating the performance of resource allocation and alternative operating policies (Chung, 2004). In simulation, to test and enhance system performance, a complex system is simplified using a computer program. This condition helps by creating opportunities to test various concepts and ideas, which have been called scenarios. Due to the advantages it offers, many simulation techniques have been developed.

One of the simulation techniques that is widely used is discrete event simulation (DES). DES is a simulation technique that can help in decision-making and problem-solving. In DES, the real world is represented by simulating its dynamics on an event-by-event basis (Babulak & Wang, 2010). However, DES helps in modelling the system by focusing on its events or processes only. In addition to DES, agent-based modeling (ABM) is another simulation technique that is popular for modelling people's behaviour. ABM consists of agents that can interact with and influence each other to create an emergence of behaviour. In addition, these agents can learn from their experiences and adjust their actions to better suit their environment (Macal & North, 2010). According to Pegden et al. (1995), the

purpose of ABM is to gain insight into the operation of the system. Aside from Pegden et al. (1995), Leemis and Park (2004) have also stated that the purpose of DES is insight, which means getting a better understanding of how the system operates and responds to changes.

Both DES and ABM appear to be simulation approaches that could model and simulate human behaviour (Majid et al., 2016). As Dubiel and Tsimhoni (2005) identify, there is some research that integrates DES and ABM. Some of this previous research has sought to compare the implementation of DES and ABM in different fields, including operations research and implementation of new policies (Zankoul et al., 2015). Thus, it does not rule out the possibility that DES and ABM can be compared in terms of insight generation.

Insight is defined as novel ideas that provide a better understanding of a phenomenon. Gogi et al. (2016) explained that, in simulation, insight occurs when people who have used a simulation model discover how to improve a system's performance by testing a 'what-if' scenario. In previous research by Gogi et al. (2016), the analysis of insight occurrence in DES has been performed. According to Gogi et al. (2016), simulation can help in generating insight even though statistically there is no significant difference between the use of animated displays and the statistical results from simulation model. This study also explained that the result obtained depends on the case simulated and the type of simulation, where different simulation cases and different simulation types might produce different results.

Accordingly, this study attempts to develop the research by Gogi et al. (2016) by utilising different cases and comparing the insight occurrence in DES and in ABM. Both DES and ABM can provide an animated display that is easier to understand, as well as the statistical results of the model that has been run to aid comprehension and assist in further data processing. As both DES and ABM are intended to provide insight, in this paper, empirical evidence of the difference in insight generation through simulation, in particular using DES and ABM, is provided by comparing the users and non-users of each simulation technique.

2 LITERATURE REVIEW

Simulation is usually performed using a computer program or simulation software that models the behaviour of real systems, and is used to analyse the system's behaviour and then formulate a policy decision (Chung, 2004). Most of the recent research on simulation discusses the implementation, development, and improvement of simulation. Some scholars have also conducted research to improve the verification and validation process of simulation, to ensure the accuracy of the modelling.

Bannet et al. (2013) explains that the usefulness of a model is not only about the accuracy of the model, but it can also be evaluated by the user. However, studies evaluating simulation models are limited, particularly studies that take into account the insight generation in simulation (Gogi et al., 2016). However, Aalst and Voorhoeve (2000) state that, through creating a simulation model, the insight generated from simulating existing or proposed future situations can be useful.

Gogi et al. (2016) explain that insight occurrence can be measured based on four variables: (1) task performance, (2) problem understanding, (3) discontinuity in thinking, and (4) change in problem understanding. Task performance is defined as the ability of the problem-solver to achieve the goals of the task, whilst problem understanding can be understood as the ability of the problem-solver to solve the problem by using a simulation model. Discontinuity in thinking occurs when the problem-solver follows a procedure that leads to multiple possible solutions for the problem. A change in understanding is defined as the ability of the problem-solver to gain a better understanding, and can be measured by comparing the understanding before and after using the simulation model.

According to Gogi et al. (2016), simulation plays an important role in the process of generating insight. Thus, this study will expand on Gogi et al.'s (2016) research to provide empirical evidence related to insight generation in simulation, especially DES and ABM. In this study, the behaviours of users and non-users of simulation will be analysed. Two different cases will be used order to represent DES and ABM.

3 METHODOLOGY

This section will provide more detail about the experiment will be described. This includes the design of experiment, details about participants, the materials used for the experiment, procedures carried out, as well as the details of and results from the pilot study.

3.1 Experiment Design

In this study, two simulations were used, DES and ABM. Each experiment of simulation model consisted of users and non-users of the simulation who were involved separately. The experiments were divided into three sections: the pre-test, treatment, and post-test.

In order to avoid discrepancies in the information given, the data collection process was carried out using a pre-prepared script. For the treatment section, an initial simulation model was provided and participants were required to reuse the model and develop the scenario until the optimal solution had been identified and the goal that had been defined by the researcher was achieved. The use of simulation model used only depends on the model given and had nothing to do with the real conditions of the system. Thus, the only factors considered were experience of using simulations not the knowledge of real systems. Finally, the analysis was carried out by comparing the results of the pre-test and post-test as well as the results from each scenario that was built by the participants.

3.2 Participants

This study included a total of 41 undergraduate students in Indonesia. The participants consisted of 11 DES users, 10 DES non-users, 10 ABM users, and 10 ABM non-users. DES users were students who had studied DES, whilst DES non-users were students who had not studied DES. Similarly, the ABM users were students who had studied ABM, whilst ABM non-users were students who had not studied ABM. The participants were tasked with solving a problem using simulation, either with DES or ABM.

In order to ensure that the participants' performances could be compared, this experiment used homogeneous participants, where the participants of this experiment had an equal standard of their knowledge in simulation. To achieve this, certain requirements needed to be met to become a participant. For user participants, these were: the participant must be an undergraduate student that have a "good" academic score, demonstrated by a minimum score of B for courses to the case study (operational research, system modelling, and simulation) and a minimum GPA of 3.00 out of 4.00. Meanwhile, for non-user participant, the different requirement were located on the absent on taking system modelling and simulation course.

3.3 Materials

This section will describe the materials used during the experiment. This includes the experimental protocol, an explanation of the case study used, and an explanation of the simulation model used in the experiment.

3.3.1 Experiment Protocol

The protocol used in this study is an adaptation of the experiment protocol used previously by Gogi (2016). The protocol set out the steps that had to be followed by participants according to the condition they were assigned to, whether DES or ABM. In the protocol, general instructions were given, as well as an explanation of the case study and instructions for using the simulation software. In addition, it contained a pre-test questionnaire and a post-test questionnaire.

3.3.2 Case Study

Two case studies were used in this experiment. For the DES condition, the case study used was about the use of a student corner (SC) facility in the Mechanical and Industrial Engineering Department of Universitas Gadjah Mada (UGM). The SC is one of the facilities provided by UGM to support the students in having discussions and carrying out other learning activities on campus. Currently, there are eight round tables with capacity of five people per table and three rectangular tables with capacity of twelve people per table. The total number of students in the Mechanical and Industrial Engineering Department at UGM is approximately 280 students. The process flow of this case study is that, when a student who wants to use the SC enters the campus, they will go to the SC area where they will see the SC condition, whether there is an empty chair or not. When there is a chair available, the student will sit and use the SC. However, when all the seats in the SC are full, it is assumed that the student will go somewhere else, namely the reference room, which has 50 seats. The students will wait in the reference room and then, when there is a chair available, return to the SC.

The DES model uses three entities: students, the SC area, and the reference room. The model has five main activities: the students come to the SC; students look for an empty chair in the SC on the floor; students look for an empty chair in the SC on the second floor; students use the SC; students use the reference room. The activity cycle diagram (ACD) for this model can be seen in Figure 1.

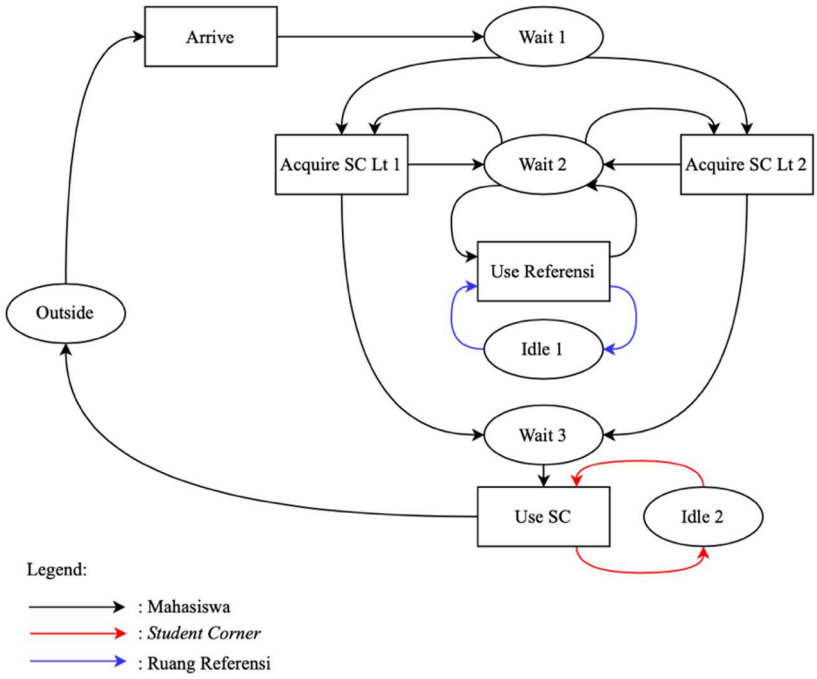


Figure 1 *Activity Cycle Diagram for DES Model*

For ABM, the case study used was about the HIV virus, and can be found in the Netlogo library. HIV can spread in various ways, such as by needle-sharing among injecting drug users, through blood transfusions, from HIV-infected women to their babies, and from sexual contact. The models examine the emergent effects of four aspects of sexual behaviour, namely: (1) average coupling tendency, (2) average commitment or amount of time that couple in the population stays together, (3) average condom use or the tendency of the population to use a condom; and (4) average test frequency or tendency to get tested for HIV. In this model, three colours are used: green, which represents uninfected individuals; blue, which represents infected individuals whose infection is unknown; and red, which represents infected individuals whose infection is known.

The ABM model consist of two entities: people, and the environment in which the people move. The state variables for the people are divided into: people who are uninfected, infected-unknown, and infected-known. At the initiation of the model, there are around 50–500 people who can be set in the model. In this model, people in the population will interact and form a 'couple'; those people who are not in couples will wander around the environment until they find a coupling. Considering the number of entities, state variables, purpose, and processes in the model, it can be said that the model complexity is simple.

From the descriptions above, it is clear that both the DES and ABM models used in this experiment are simple in terms of model complexity. The DES model consists of three entities with a simple flow and activity, whilst the ABM model consists of two entities with the purpose of examining the spread of a virus in a small isolated human population. Thus, due to having the same level of model complexity, the two simulation models can be compared.

3.3.3 Simulation Model

The SC simulation model was developed by Hersetiawan et al. (2019) using Flexsim software. The simulation model presents the process of SC daily use based on the number of uses for each table and the different usage times. This case focuses only on the use of SC facilities, whilst the use of the reference room, which also exists in the model, was ignored because it is considered a waiting area. The

SC simulation model in Flexsim can be seen in Figure 2. In the model, the animated display and statistical results generated from the simulation run can be analysed by participants.

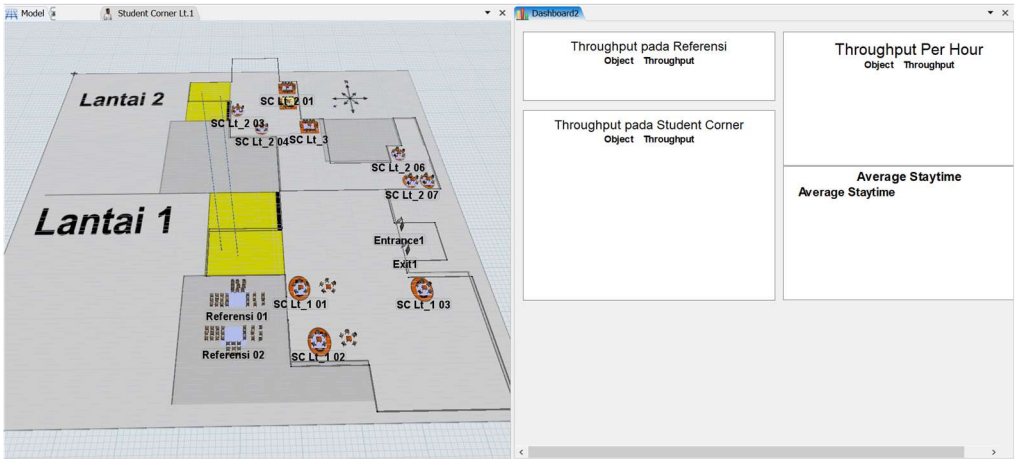


Figure 2 *Interface of the Student Corner Model in Flexsim*

The simulation model used for ABM was the HIV model that exists in the Netlogo software library, and can be seen in Figure 3. This HIV model was developed by Wilensky (1997). The model simulates the spread of the HIV virus through sexual transmission in a small, isolated human population. Thus, the model illustrates the effects of certain sexual practices across a population. In this model, several variables are used, which are: initial person, average coupling tendency, average commitment, average condom use, and average test frequency. The output is the rate of HIV infection in the population (as a percentage) over a certain period of time.

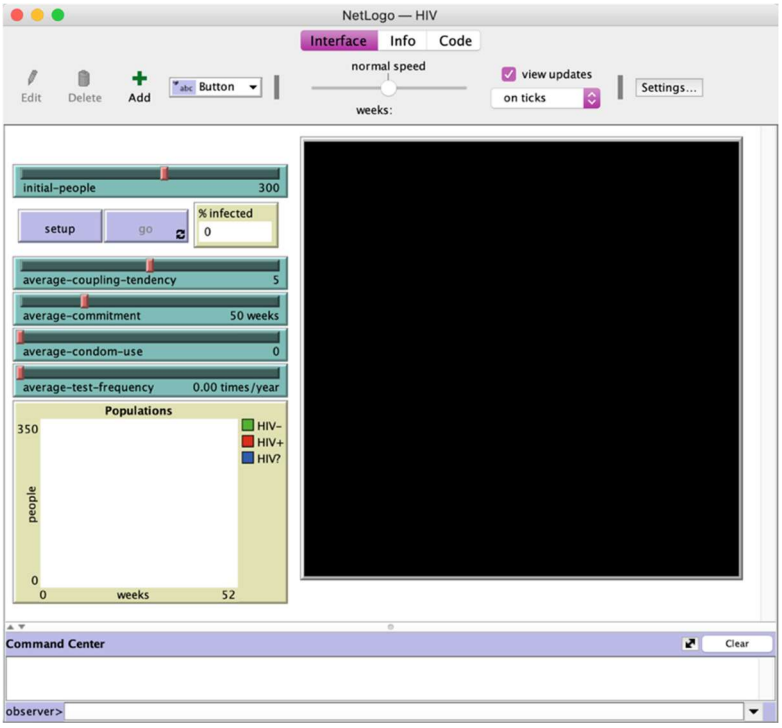


Figure 3 *Interface of the HIV Model in Netlogo*

3.4 Experiment Procedure

The experiments in both DES and ABM were conducted using the same procedure: they were conducted online using web and video conferencing tools. In the experiment, each participant had a private session with the authors. At the beginning of the session, the experiment to be carried out was explained to participants. Then, the participants were asked to read general instructions and complete the consent

form. Next, the participant was asked to read the case study and complete a pre-test questionnaire that asked their opinion about the problem of the case study and its causes. This pre-test questionnaire was used to analyse the participants' problem understanding before using simulation.

Next, the participants were given a period of time to read the instructions for how to use the simulation software. Then, they were asked to run and work on the SC simulation that had been modelled in Flexsim or the HIV model in the Netlogo software. The participants were asked to analyse the model by creating scenarios in order to improve the system or solve the problem in the model. In Flexsim, the scenarios were created by changing the number of tables from the previous model. In Netlogo, the scenarios were created by changing the four parameters that were available in the model – average coupling tendency, average commitment, average condom use, and average test frequency. This process was called the solving session and had to be completed in 30 minutes. After the participants had completed the solving session, they were asked to complete the post-test questionnaire, which contained some questions about the scenarios that could provide the best solutions for the problem. This post-test questionnaire was used to assess participants' understanding of the problem after using the simulation.

4 RESULTS AND DISCUSSION

In this study, four variables were analysed in order to assess the insight occurrence (Gogi et al., 2016). The variables are: (1) task performance, (2) problem understanding, (3) discontinuity in thinking, and (4) change in problem understanding. These variables were used to analyse the differences in insight-generating processes among users and non-users of DES and ABM.

4.1 Task Performance

The task performance was measured based on whether or not the participant achieved the goal of the task, as determined by the participant's answers in the post-test questionnaire. Depending on their answers, the participants were classified into a 'solver' group and a 'non-solver' group. The solver group consisted of participants who were able to submit the optimal solution for the problem. The participants who did not submit the optimal solution, or submitted a solution that did not solve the problem, were classified into the non-solver group.

In DES, participants who submitted scenarios that involved increasing the capacity of the SC were considered solvers. The participants who reduced the capacity of the existing SC were categorised as non-solvers, because they had not overcome the existing problem. In ABM, participants who could achieve the goal set at the outset, which was to have a maximum of 2.84% of the population infected in week 156, were considered solvers. The difference in the number of solvers among users and non-users of DES and ABM can be seen in Table 1 below.

Table 1 *Proportion of Solvers Among Users and Non-users of DES and ABM*

Solvers	DES		ABM		Total	
	n	%	*n*	%	*n*	%
Users	9	53%	10	50%	19	51%
Non-users	8	47%	10	50%	18	49%
Total	17	100%	20	100%	37	100%

In order to test the difference of task performance that the problem-solving process for users and non-users of DES and ABM, the chi-square test was adopted to compare the solution rate for both conditions, DES and ABM. The result of the statistical test showed a p-value of 0.8584. From this result, with a significance level of 0.05, it can be concluded that there is no significant difference between the problem-solving process for users and non-users of DES and ABM.

4.2 Problem Understanding

Participants' problem understanding was measured in order to determine whether the participants were solving the problem by generating insight or by intuition (Gogi et al., 2016). If the participants were

able to justify their solution at the end of the session after running the simulation model, the participants were identified to have solved the problem by generating insight. This problem understanding was measured using two open-ended questions that asked about the reasons why the chosen scenario could solve the problem and what the user of the system should do in order to achieve the target. The participants' answers were grouped into three categories: inaccurate, incomplete, and complete.

The inaccurate category includes incorrect or unclear actions and inaccurate justifications of the reasons the proposed scenario can solve the problem. The incomplete category includes the answers of participants who did not directly solve the problem. The complete category includes the answers of participants who directly solved and fully recognised the cause of the problem and the action that must be taken to solve the problem.

Table 2 *Problem Understanding After Running Simulation Model*

Post-test Problem Understanding	DES (n = 17)		ABM (n = 20)		Total
	Users	Non-users	Users	Non-users	
Inaccurate	0	1	0	0	1
Incomplete	1	0	0	1	2
Complete	8	7	10	9	34
Total	9	8	10	10	37

From Table 2, it can be seen that three participants could not justify their answer or did not fully recognise the cause of problem at the end of the session. This condition indicates that these participants tried to solve the problem without generating insight but by using intuition. This finding will be taken into account when analysing the solution rate, which indicates the insight occurrence, in the next section. The only participant who had inaccurate problem understanding was in the DES group. On the other hand, the two participants that had incomplete problem understanding were in. the ABM and DES groups. Most of participants had complete problem understanding and were able to provide solutions by understanding the existing problems. These participants were considered to have solved the problem by generating insight.

4.3 Discontinuity in Thinking

Discontinuity in thinking was observed based on the number of scenarios built by the participants. According to Gogi et al. (2016), if the participants can solve the problem correctly on the first attempt, they have not shown discontinuity in thinking. This implies that the participant tried to solve the problem without generating insight because they knew how to solve the problem, not through the use of simulation. In this study, the number of scenarios that were built by the participants can be seen in Table 3.

Table 3 *Number of Scenarios Built*

Number of Scenario	DES		ABM	
	Median	(lower-upper quartiles)	Median	(lower-upper quartiles)
User	3	(1–5)	5	(1–13)
Non-user	2	(1–4)	6	(1–13)

The experiment results showed that among both users and non-users in DES and ABM, there were participants that only built one scenario and submitted that scenario as the optimal solution. This indicates that there were participants who solved the problem on their first attempt without requiring the help of the simulation model. These participants were considered not to have discontinuity in thinking because they knew how to solve the problem after reading the case study. This finding will be taken into account when analysing the solution rate, which indicates insight occurrence, in the next section.

In addition, statistical analysis using the chi-square test was also conducted in order to determine the difference in the number of scenarios built by users and non-users of DES and ABM. The results showed a p-value of 0.5896. By using the significance level of 0.05, it can be concluded that there was no significant difference between the number of scenarios built by users and non-users of DES and ABM.

4.4 Change in Problem Understanding

Change in participants' problem understanding was measured based on a self-assessment completed by participants. In the post-test questionnaire, the participants were asked about their change in understanding after attempting to solve the problem using the simulation model. The question used a five-point Likert scale with 1 indicating "A lot worse" and 5 indicating "A lot better". The results for participants' change in problem understanding are presented in Table 4.

Table 4 *Participants' Self-assessment of Change in Understanding*

Self-Assessment	DES (n = 17)		ABM (n = 20)		Total
	Users	Non-users	Users	Non-users	
A lot better	3	2	1	1	7
Better	5	4	8	9	26
Similar	1	2	1	0	4
Worse	0	0	0	0	0
A lot worse	0	0	0	0	0
Total	9	8	10	10	37

Most of the participants reported that they had better problem understanding after attempting to solve the problem using the simulation model. However, from Table 4 it can be seen that four participants claimed that their understanding remained the same before and after attempting to solve the problem using simulation. This could be because participants believed that they knew how to solve the problem after reading the case study. It could also be because the participants tried to solve the problem using intuition. The participants who claimed to have similar problem understanding were considered to have attempted to solve the problem without generating insight. This finding will be taken into account when analysing the solution rate in the next section.

After analysing the four variables, the rates of insight occurrence were calculated. This was done by combining the insight occurrence results for the four variables as presented in the previous section. The result of task performance analysis showed that there are 90% DES users, 73% DES non-users, all ABM users, and all ABM non-users who can solve the problem by generating insight. Furthermore, the results of the change in problem understanding show that 11% DES users, 12.5% DES non-users, and 10% ABM non-users attempt to solve the problem without generating insight.

After analysing the discontinuity in thinking of participants, it could be concluded that there were some participants who solved the problem without generating insight. Based on the participants' change in problem understanding, it was established that 11% DES users, 25% DES non-users, and 10% ABM users solved the problem without generating insight. To summarise these results, the solution rates indicating insight occurrence for users and non-users of DES and ABM are presented in Table 5.

Table 5 *Solution Rates Indicating Insight Occurrence*

Solution rates, indicating insight occurrence	DES (n = 17)	ABM (n = 20)	Significance of Difference
Users	50% (5/10)	60% (6/10)	No (p-value = 0.7215)
Non-users	55% (6/11)	60% (6/10)	

The results above show that 50% of DES users and 60% of ABM users solved the problem by generating insight. Even though the non-users of DES and ABM had not studied the simulation techniques, 55% of DES non-users and 60% of ABM non-users could solve the problem by generating insight. However, users and non-users of ABM had higher solution rates, indicating greater occurrence of insight than users and non-users of DES. This could be because ABM has a simpler interface; the scenario changes can also be made easily, simply by changing the parameters that are shown in the model interface.

A statistical test was conducted in order to calculate the differences between users and non-users of DES and ABM. Using the Fisher exact test, a p-value of 0.7215 was determined. With a significance level of 0.05, it was concluded that there was no significant difference between users and non-users of DES and ABM in solving the given problem by generating insight.

5 CONCLUSION

From the discussion in previous sections, it can be concluded that, statistically, there is no significant difference in insight-generation by users and non-users of DES and ABM. However, users and non-users of ABM showed higher solution rates, indicating the occurrence of insight, than users and non-users of DES. However, this study applied reused models, meaning the participants were not required to develop their own model but to solve the problem using the given model. Additionally, the systems simulated both in DES and ABM in this study have low complexity only consisted of two to three parameters. Therefore, further research could analyse the difference in the model-building process between users and non-users of DES and ABM with the different complexity level.

ACKNOWLEDGEMENTS

The authors acknowledge the contribution of Hersetiawan et al. (2019) in regard to the student corner model in Flexsim software, and Wilensky (1997) in regard to the HIV model in the Netlogo library. Sections of this paper are based on Gogi, Tako and Robinson's (2016) experimental investigation into the role of simulation models in generating insights, published in *European Journal of Operational Research*, (pp. 931–944).

REFERENCES

Aalst, W. M. Van Der & Voorhoeve, M. (2000). *Simulation Handbook*. Eidhoven: Technical University Eindhoven, Eindhoven.

Babulak, E., & Wang, M. (2010). *Discrete Event Simulation: State of the Art, Discrete Event Simulations*. Aitor Goti (Ed.), InTech.

Bennett, N. D., Croke, B. F., Guariso, G., Guillaume, J. H., Hamilton, S. H., Jakeman, A. J., & Pierce, S. A. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, *40*, 1–20.

Chung, C. A. (2004). *Simulation Modeling Handbook: A Practical Approach*. CRC Press.

Dubiel, B., & Tsimhoni, O. (2005). Integrating agent-based modeling into a discrete event simulation. *Proceedings of the 2005 Winter Simulation Conference*, 1029-1037

Gogi, A. (2016). *Insight Generation in Simulation Studies: An Empirical Exploration*. Loughborough University.

Gogi, A., Tako, A. A., & Robinson, S. (2016). An experimental investigation into the role of simulation models in generating insights. *European Journal of Operational Research*, 249(3), 931–944.

Hersetiawan, A., Sunggoro, A. M., Alsana, S., Puspita, S., Pramudita, S. I., & Gani, Y. A. B. (2019). *Simulasi Student Corner UGM menggunakan Software Flexsim*.

Leemis, L., & Park, S. (2004). *Discrete-Event Simulation: A First Course*. Prentice Hall. 4–17.

Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162.

Majid, M. A., Fakhreldin, M., & Zuhairi, K. Z. (2016). Comparing Discrete Event and Agent-Based Simulation in Modelling Human Behaviour at Airport Check-in Counters. *International Conference on Human-Computer Interaction*, 510–522.

Pegden, C. D., Shannon, R. E., and R. P. Sadowski. 1990. *Introduction to Simulation Using SIMAN.* McGraw-Hill, Inc., New York.

Wilensky, U. (1997). Netlogo HIV model. http://ccl.nortwestern.edu/netlogo/models/HIV. Centre for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Zankoul, E., Khoury, H., & Awwad, R. (2015). Evaluation of Agent-Based and Discrete-Event Simulation for Modelling Construction Earthmoving Operations. *32nd International Symposium on Automation and Robotics in Construction and Mining*,

AUTHOR BIOGRAPHIES

HILYA MUDRIKA ARINI holds a PhD in management science from the University of Strathclyde, Glasgow. She is an industrial engineering faculty member at Universitas Gadjah Mada, Indonesia. Her research interests focus on behavioural operations research and decision analysis.

NUR AINI MASRUROH is an associate professor in the Industrial Engineering Programme, Universitas Gadjah Mada, Indonesia. Her primary research interest is in operations research, decision making under uncertainties, healthcare system simulation and optimisation, production planning, inventory and scheduling, and revenue management.

NURUL LATHIFAH received her bachelor's degree in industrial engineering from Universitas Islam Indonesia in 2017 and an M.Sc degree in industrial engineering from Universitas Gadjah Mada in 2020. Her master's degree research focused on the implementation of behavioural operational research concepts in simulation. Currently, she is a research assistant focusing on research on routing problems at Universitas Gadjah Mada. Her main research interests include behavioural operational research, simulation, supply chain, and vehicle routing problems.

LAELA DEWI MAHARANI has completed her studies in industrial engineering at Universitas Gadjah Mada. She was an assistant lecturer in optimisation mathematics and is interested in conducting behavioural operations research.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.023

MODELLING METHODS - REPURPOSING A DISCRETE EVENT SIMULATION MODEL TO INCLUDE COVID 19 WORKFLOW CHANGES IN A COMPUTED TOMOGRAPHY DEPARTMENT

Mary Conlon

National University of Ireland, Galway
Health Service Executive
m.conlon17@nuigalway.ie

Dr. Owen Molloy

National University of Ireland, Galway

owen.molloy@nuigalway.ie

Ms Avril Hanly

Health Service Executive
Portiuncula University Hospital
avrilm.hanly@hse.ie

ABSTRACT

The importance of infection control has taken centre stage with the arrival of Covid-19, and radiology staff must take precautions to limit contamination. This paper identified the changes required to an existing discrete event simulation model of a CT service to repurpose it for post COVID-19. Methods: Radiology workflow was mapped using Microsoft Visio to capture changes to roles, tasks and communications. Task and delay times were observed. Interviews with cleaning and clinical staff verified observational findings. Rich picture diagramming was used to include staff perceptions. In partnership with decision makers a culturally desirable and feasible scenario was identified and the increase in the consumed staff time post COVID-19 demonstrated. Conclusion: While CT throughput has decreased, the individual inpatient workload in terms of staff resource utilisation has increased. Separate inpatient and outpatient services are recommended to increase throughput and efficiency.

Keywords: Computed tomography, discrete event modelling, COVID-19, Rich Picture

1 INTRODUCTION

We are passengers on an aircraft which we endeavour to fly and redesign in mid-flight (Sterman, 2001). With crowded waiting rooms a thing of the past, COVID-19 has rendered many models of radiology service delivery obsolete. Responding to a call to arms, this empirical work examines how an existing model was repurposed with COVID-19 parameters (Currie et al., 2020).

Demand for CT is partially driven by population growth and age profile as well as increased incidences of chronic diseases (Adam, 2006; Central Statistics Office., 2015; The Royal College of Radiologists, 2020). Other factors affecting demand include increased screening and new clinical guidelines which incorporate CT in the clinical pathways. Demand for CTs has increased by 10% and demand for MRIs by 8% over the past year (The Royal College of Radiologists, 2020).

The importance of infection control has taken centre stage with the arrival of COVID-19, and radiology staff must take precautions to limit contamination (Zanardo et al., 2020). Twenty minutes to one hour of downtime is necessary where scanners have been used for suspected or confirmed cases of the virus and where the patient requires aerosol generating procedures. The allocation of dedicated COVID-19 CT scanners in departments has been recommended but may not always be feasible in single scanner departments (Mossa-Basha et al., 2020; Orsi et al., 2020).

In "A guide for building hospital simulation models" three simulation methods were evaluated: discrete event simulation (DES), system dynamics (SD), and agent-based simulation (Gunal, 2012). While SD considers entire cohorts and populations, the other methods consider the individual. Where SD uses rates to pass a population through its model, DES uses process blocks and the individual agents/patients moves through as a complete entity. DES models are stochastic in nature and can take account of variability in the time taken to carry out activities and the times between arrivals into the system, as well as the utilisation of resources (Currie et al., 2020). DES has been used extensively in radiology and has touched on service improvement, staff burnout and fatigue, pathway redesign (Booker et al., 2016; Oh et al., 2011; Rachuba et al., 2018; Reinus et al., 2000; Van Lent et al., 2012).

This case study elicited knowledge from radiology staff using RP diagramming, workflow analysis and observation to create a shared understanding and identified factors affecting CT service delivery during a pandemic. The intertwined elements of a CT service including the motivations and priorities of those involved in the service are examined using tools from soft systems methodology (SSM) (Crowe et al., 2017). Opportunities for service improvement were identified and tested in the resultant discrete event simulation model of the service to determine their impact on the CT waiting list and staff workload. Finally, feasible and culturally desirable targeted simulations were identified for testing in the DES model which are applicable in the context of service provision during a pandemic. This paper sets out to:

1. Provide input parameters for a simulation model of a CT service specific to the handling of COVID-19 cases.
2. Determine how staff workflow and workload in terms of consumed staff time has changed because of COVID-19.
3. Make desirable and feasible recommendations for future service delivery.

2 METHODOLOGY

The case study hospital provides a 24/7 acute surgical, medical and critical care service with emergency and maternity services with approximately 100 inpatient beds. In radiology, a single CT scanner provides a scheduled service from 8.30am to 5pm with a 24-hour emergency service for inpatients and accident and emergency patients. Approval to conduct the study was obtained from the hospital management team. The researcher was employed as a radiographer for four years in the CT department prior to commencing the research work. The identity of the radiographer as a researcher and the purpose of the research were disclosed to the staff before interview. Anonymized data was used and stored in line with local data protection guidelines. A framework was developed for a simulation-based decision support system for the CT Department, Figure 1. The iterative approach taken proved capable of accommodating the changed circumstances resulting from the COVID-19 epidemic, allowing repurposing of the model.

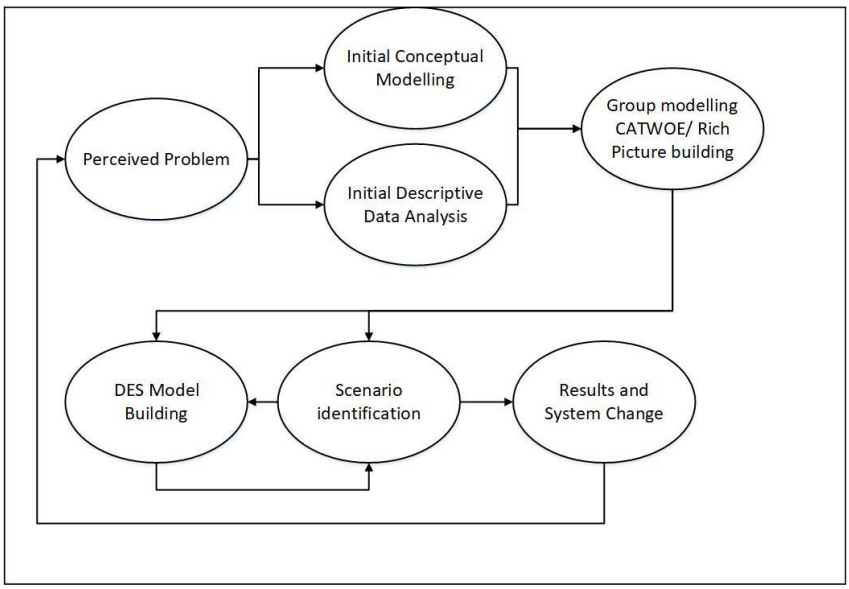


Figure 1 *Framework used to create decision support tool*

Stakeholders at both clinical and managerial levels participated in model development and scenario identification. Process maps were created of the workflow when scanning a suspected or confirmed COVID-19 case and those patients undergoing procedures potentially capable of generating aerosols. CT cases where the patient undergoes aerosol generated procedures (AGP) such as high flow oxygen therapy or open airway suction result in one hour of downtime before the CT room can be cleaned (HSE, 2020). Following augmentation of an earlier RP diagram, decision makers were prompted to consider feasible and desirable service improvements in the post COVID-19 context.

An existing model of the service created using the software AnyLogic (University Edition 8.4) included delays, queues, utilisation of resources, decisions, phone call interruptions, administrative activities and the patient journey. Observed triangular stochastic delays (minimum, maximum and mode) resulting from transportation and manual handling delays were incorporated into the model with different likelihoods for the inpatient and outpatient cohorts. Likewise, Poisson arrival schedules based on historical data were created for the inpatient and outpatient (including general practitioner (GP) patient) sources. A waiting list object was created to initiate the model with a queue of existing outpatients and GP patient requests which was based on the real waiting list and categorised by exam type. A third source was created with actual arrival times for one year of patient examinations. This was used to run the model in historical mode. In this mode historical arrival times and exam types were used to compare actual patient time in CT with model patient time in CT for the purpose of model validation.

Decisions regarding preparation, number and type of staff required, patient prioritisation and workflow were captured pre and post COVID-19. Schedules were designed for each scenario. When running stochastically, patients are assigned age, infectiousness, mobility and exam type attributes based on custom distributions for each which was based on an analysis of historical data (Conlon & Molloy, 2019). Figure 2 depicts a subsection of the DES model related to manual handling and infection control. In the figure, flexible staff are required to assist in manual handling of patients and time delays result when sourcing them. The patient is released from CT and the CT scanner is unavailable until the wrap up task which is cleaning is completed.

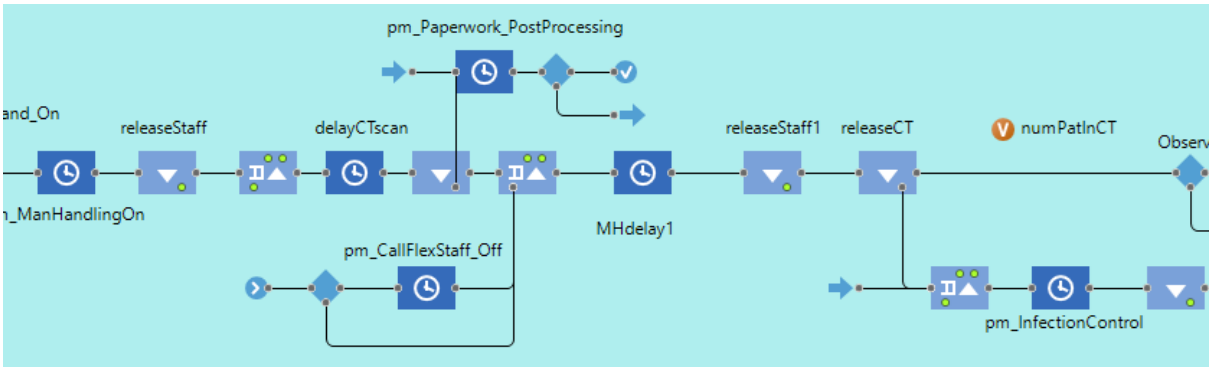


Figure 2 *Subsection of DES AnyLogic model*

The model included logic, animation and statistics/metrics pages. The statistics page provided a dashboard including scanner, radiographer and healthcare assistant utilization, consumption of staff time, and growth of the CT waiting list.

2.1 Repurposing the Model for COVID-19

The stochastic variations in time taken to scan suspected and confirmed, AGP and non AGP COVID-19 cases were observed. The CT workflow for these cases was documented. Cleaning staff and the CT clinical specialist were interviewed to verify the task times and workflow observations. A custom distribution table was used to allocate a COVID-19 infection status to Inpatients. Currently COVID-19 positive outpatients are not given appointments. The stochastic arrival of outpatient's was reduced to reflect the limited service. A database was created for cleaning times for cases where the patient was:

1. Not considered infectious and standard precautions taken,
2. Confirmed or suspected as having COVID-19 and where full PPE was required,

3. Confirmed or suspected as having COVID-19 and where full PPE required and AGPs were used.

Working closely with the clinical specialist and using the interview data and RP diagrams the salient factors for inclusion in the DES model were determined. Many tasks associated with CT service provision which were identified during interviews with staff were not included in the model. Such tasks include stock management, continuous professional development, logging of incident forms, training of staff and the vetting of outpatient CT requests. These were not included as to do so would have decreased the reusability and increased the model building complexity; this was not deemed practical and determined to be outside the scope of the model.

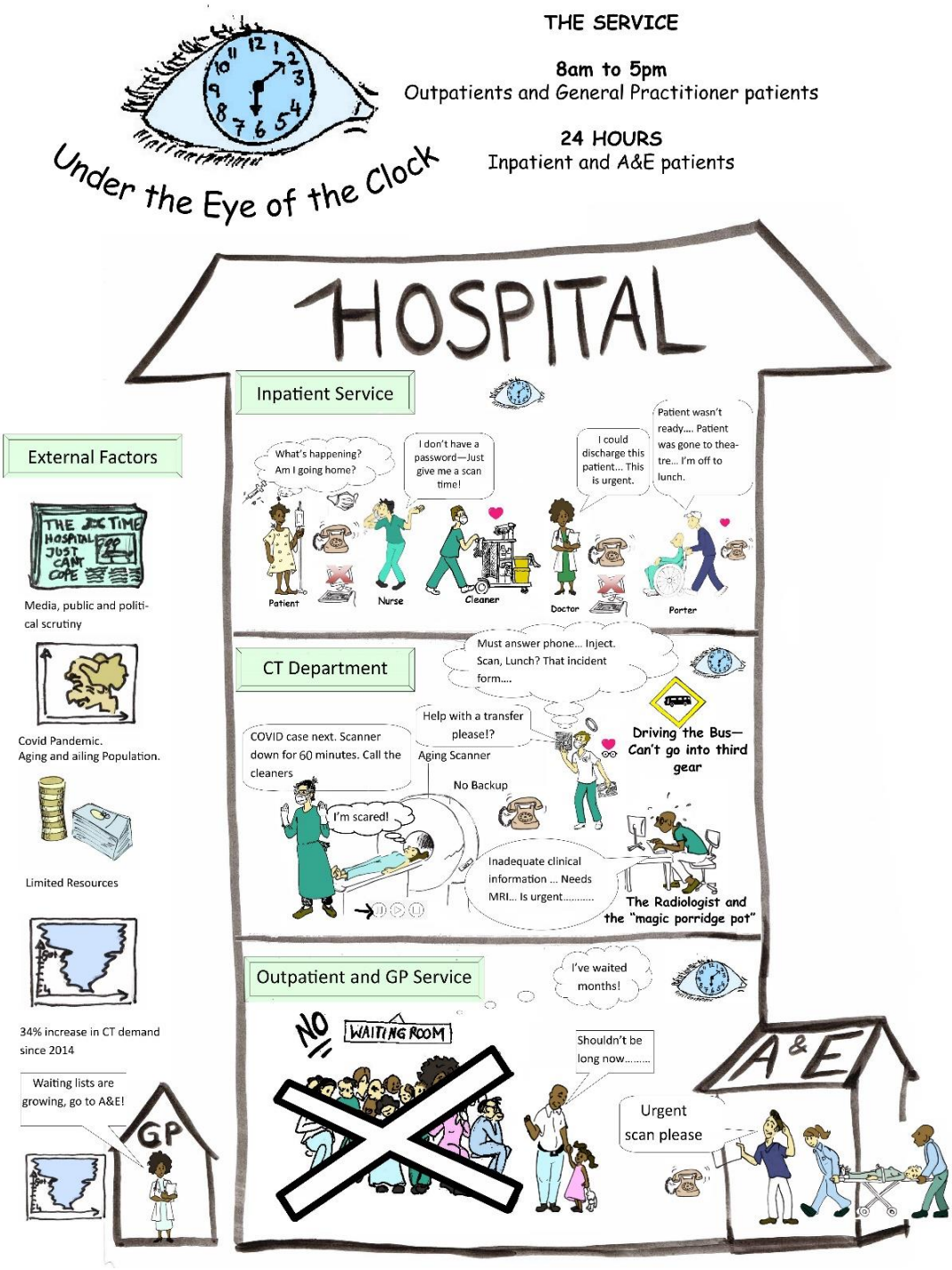


Figure 3 *Rich Picture of staff perceptions of service changes post COVID-19*

2.2 Model Validation

Verification and validation are executed throughout each phase of the simulation model development. Statistical validation compared model results with historical data using the patient time in radiology from arrival to scan end. This was compared for bias and standard deviation of errors between historical and model outputs. Validation also compared inpatient and outpatient arrival rates and the evolution of the CT waiting list with historical data. For each metric created, individual cases were examined, using logger data and the patient characteristics.

3 RESULTS

3.1 Explanation of the Rich Picture

In the RP, entitled "under the eye of the clock", a clear definition is made between the inpatient and outpatient sources by locating them on separate floors of the hospital (Figure 3).

Symbols are used to illustrate how the speed of work has been temporarily paused or slowed down. Waiting rooms are no longer crowded as GP and outpatient bookings are limited and patients phoned when the CT scanner is available. Staff shared how prior to COVID-19 everything was considered urgent, but since there is a "more clinical" prioritisation of work. Staff described how in the new COVID-19 environment staff have each other's backs and this is represented by the glasses. A heart indicates how relationships between radiology staff and cleaning staff have improved and staff are now on first name basis. A radiographer in full personal protective equipment (PPE) is seen waiting for the arrival of a COVID-19 patient. The department is closed to all other traffic during this waiting time and will not resume until the scan is complete and the patient has left the cleaned down department. The CT room remains vacant for one hour after a ventilated patient is scanned – again a clock features in the scene. The telephone features five times indicating the constant communications between ward staff, referrers, cleaner, porters and CT staff required to safely schedule and coordinate each CT scan.

3.2 Model changes post COVID-19

The workflow for Covid-19 AGP cases was mapped and is presented in Figure 4. The diagram includes the administrative preparatory work as well as the actual scanning of the patient. A questionnaire must be completed with ward staff to ensure that the patient is ready for their scan on arrival. Should the patient require observation additional staff must accompany the patient from the ward as staff wearing full PPE are not permitted in the CT console area. The questionnaire includes questions on their transportation method, details of intravenous (IV) access site, state of dress and the patient's resuscitation status in case of an event while in the department. Scanner and staff utilisation are decreased due to the sixty minutes time delay following AGP cases.

The following changes, identified as part of the RP diagramming exercise, were required to the existing DES model:

- Three additional phone calls pre-CT scan plus completion of COVID-19 questionnaire with staff nurse/referring doctor.
- Three additional phone calls post CT scan to alert ward of patient transit and to alert other departments that radiology is once more accessible for ultrasound and general x-ray patients.
- COVID-19 patients utilising aerosol generating procedures (AGP) such as ventilation, high flow oxygen or suction etc results in 1 hour downtime followed by 15 minutes cleaning and drying time.
- COVID-19 non AGP results in twenty minutes delay plus five mins cleaning and five minutes for drying (provided no delay sourcing cleaning staff).
- Change to OP schedule - OP bookings were decreased from ten examinations per day to four for a period of three months Match to May and again in January 2021.

The model parameters updated post Covid-19 are presented in Table 1.

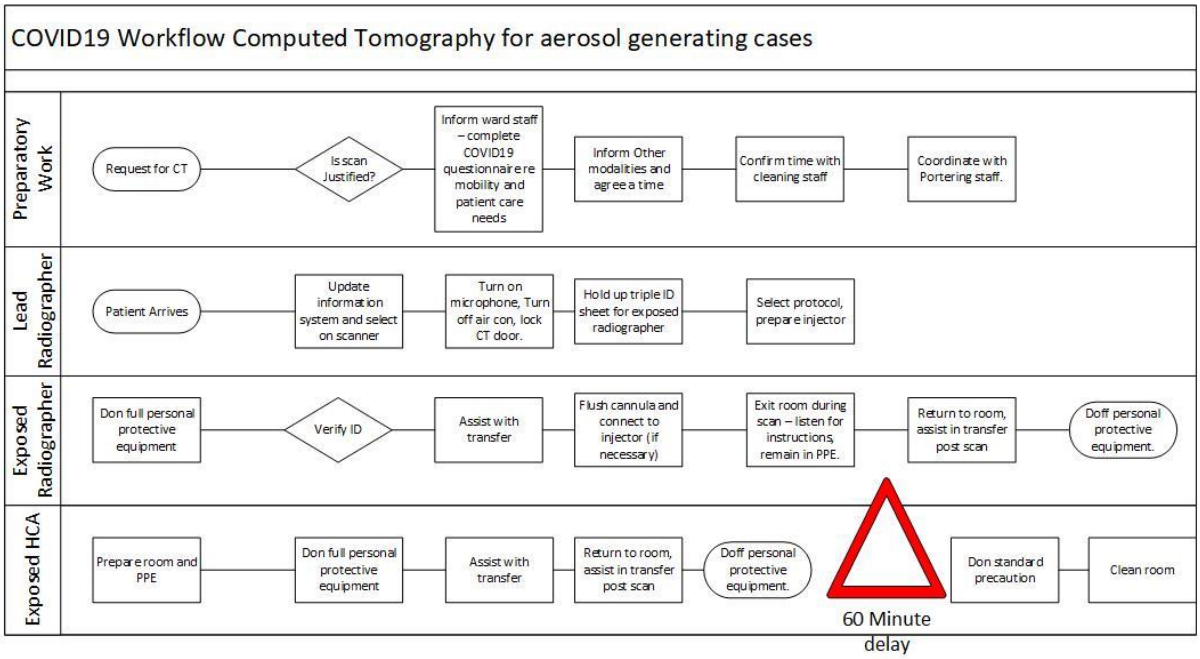


Figure 4 *Visio Diagram for CT workflow for COVID-19 positive case*

Table 1 *Tasks related to infection control*

Tasks for infectious patients	Staff required	Time taken (minutes)	Number of staff	Explanation of task
Standard precautions clean post scan	Assistant or Radiographer	7	1	Cleaning time (2 mins) plus required drying time (5 mins)
Confirmed or suspected COVID-19 non AGP clean post scan	Assistant or Radiographer cleans equipment, Hospital cleaners for floors and doors	10	1	Includes cleaning and drying time
Aerosol generating procedure Covid-19 clean post scan	Assistant or Radiographer cleans equipment, hospital cleaners touch points	10	1	NB Room downtime minutes required before clean = 60 mins
Don/doff standard precautions	Any staff member	1	1	Before and after each scan/activity
Don/doff full personal protective equipment	Any staff member	3	1	Before and after each scan/activity
Time required to call and wait for hospital cleaners	Cleaning staff	(0,15, 5)	1	triangular delay - minimum, maximum and mode time waiting for cleaning staff in minutes
Scan non contrast/oral	Radiographer	3	2	Previously 1 radiographer required
Scan involving IV contrast	Radiographer	7	2	

The scenarios identified in partnership with decision makers and the simulation results are presented in Figure 5.

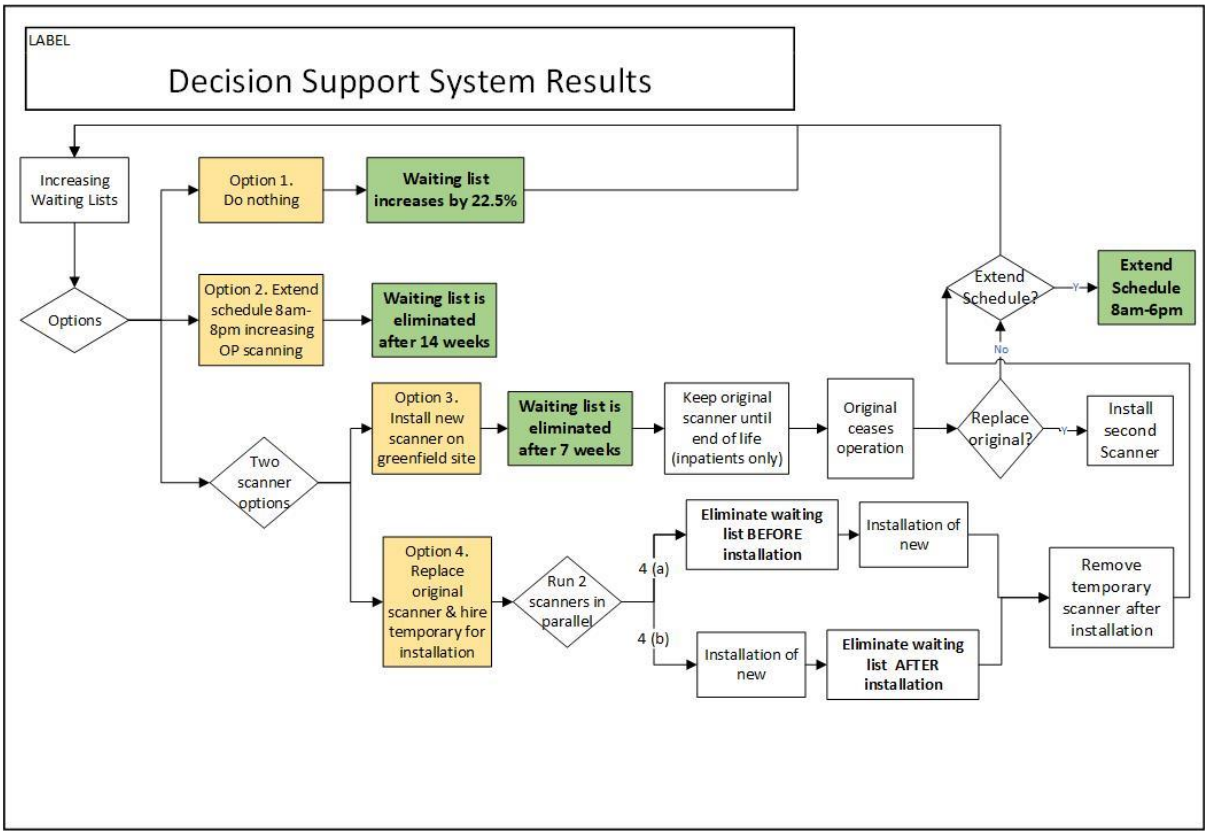
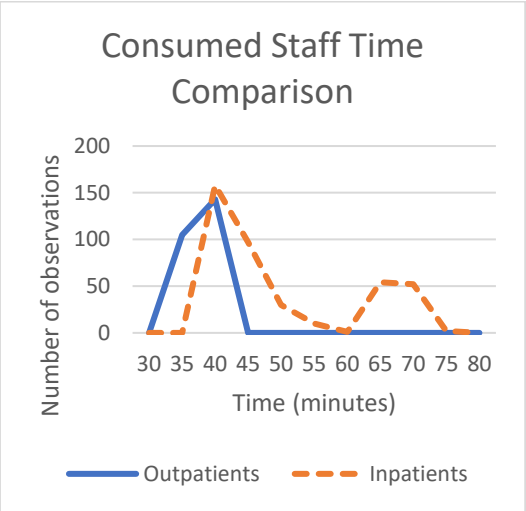


Figure 5 *Scenarios identified and simulation model results*

The pre COVID-19 consumed staff time (minutes) for exams requiring oral and IV contrast for 653 inpatient and outpatient exams were extracted from the model and are presented in Figure 6a. Post COVID-19, the average consumption of staff time for each examination category, for inpatients, was shown to vary as shown in Figure 6b.



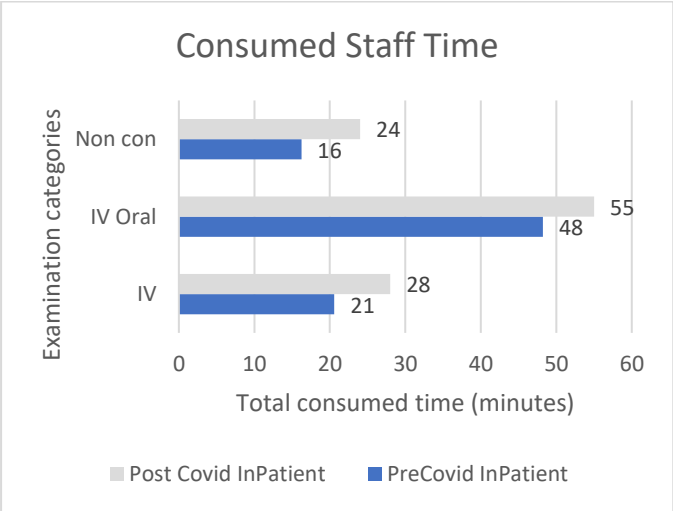


Figure 6a (left) *Variation in consumption of staff time by patient cohort pre COVID-19.* **Figure 6b (right)** *Effect of COVID-19 on consumption of staff time.*

3.3 DES model Validation data

The process was validated by comparing historical and model data for the patient time in system.

Table 2 *Validation Results*

Mean System Time and Errors for Different Exam Types(minutes)

Exam Categories	Mean Time in System - 95% CI	Mean Error - 95% CI
Overall	54.4±1.2	3.36±0.64
IV	33.2±1.6	-1.81±1.76
IVandOral	118.6±0.8	8.55±1.20
None	20.5±0.6	1.76±0.81
Oral	113.0±3.3	7.30±4.33
Procedure	48.7±2.7	2.81±3.01

4 DISCUSSION

The RP (Figure 3) captures the key features of the CT service, staff activities, the process, the environment, the delays, distractions, and some external factors contributing to workload and affecting service delivery. RP diagramming provided a strategic opportunity for meaningful decision-maker involvement allowing a tangible space to discuss and negotiate worthwhile, recommendations for change (Bell et al., 2019). While some argue that SSM is not a decision making tool and rather a post hoc measure to justify the status quo (Bergvall-Kareborn, 2002) it did, in this case study, arrive at a service improvement recommendation (Figure 5) that had not previously been considered. In scenario 3 a new CT scanner is installed on a greenfield site, with the original scanner running in parallel until it ceases operating. Following iterative consultations, it was decided to extend scenario 3 to consider the point in time when the original scanner would cease to operate. The decision, at this point, to replace the original or extend the schedule was included in the DST. The feasibility and desirability of a solution can influence whether it is implemented in practice and as has been reported in past literature the very process of designing and building the simulation model resulted in a greater understanding and appreciation of the behaviour of the service (Monks et al., 2014, 2016; Simon Dodds & Phillip Debenham, 2016).

Decision makers identified scenario 3 as the most feasible and desirable. Here a new outpatient-only scanner is installed with the original CT scanner allocated to the inpatient and emergency services. This scenario eliminates the waiting list seven weeks post installation and provides redundancy against service downtime while allowing separation of inpatient and outpatient cohorts. Segregation of infectious patient cohorts using designated scanners has been suggested as a measure to help hospitals deal with the COVID-19 pandemic (Huang et al., 2020; Sim et al., 2020). Scenario 3 allows the original scanner to continue to provide a service until it is no longer usable or cost effective to repair, thus maximising its potential. While no benefit was seen in decommissioning earlier than required, it was recognised that the hospital or OP service would experience disruption when the scanner finally ceased operations. The decision support tool includes this future scenario regarding replacement of the original scanner or extension of service hours.

In addition to specific recommendations, the study offered some general insights, in terms of consumed staff hours for inpatient and outpatient cohorts (Figure 6a). Inpatients have long been recognised as “schedule busters” and outpatients as schedule “buffers”, and recommendation made to separate these services (Boland, 2008; Murray et al., 2017; Reinus et al., 2000). The additional tasks associated with COVID-19 cases are provided in Figure 4 and Table 1 and were found to result in further work perturbations and increased consumption of staff time (Figure 6a and 6b). Industrial methods such as lean and six sigma advocate the reduction of process variation times and have applications in health (Womack et al., 2007; Young et al., 2004). This work recommends that inpatient and outpatient services be delivered and considered separately.

On a national level this works recommends the establishment of regional diagnostic hubs to provide a dedicated scheduled service for outpatient and GP diagnostic imaging. A scheduled service benefits

from increased efficiency, reduced variability, and a reduction in infection control related downtime. On the counter side vulnerable inpatients would benefit from not sharing waiting areas and corridors with outpatients and those who accompany them, improving their experience of privacy, dignity and avoidance of degradation (Murray et al., 2017).

When called upon in July 2020 the iterative nature of the framework (Figure 1) proved effective in adapting to the novel problem situation facing the radiology service delivery. This example of model reuse demonstrated how the initial modelling project provided an introduction to the capabilities of operations research and led to its becoming "baked in" as a decision support within the department (Ackoff, 2010). Benefits of the inclusion of decision makers in the research project included:

1. providing them with an opportunity to internalize research knowledge,
2. promotion of trust and consensus building and a more meaningful focus,
3. improvement of relationships,
4. higher likelihood of implementation (Harper & Pitt, 2004; Monks et al., 2016; Ross et al., 2003).

The radiology manager commented that while radiographers "naturally understood patient and process complexity, the model was useful when communicating with higher levels of decision makers".

4.1 Limitations and Future Work

Patient care remains largely unquantified and further research is recommended into the use of operational research methods for modelling patient care activities. It is recommended that future models allocate time for patient care or reference be made to it in model assumptions. The quality of reassurance, obtaining of informed consent, exam explanations and preparedness of the patient for diagnostic imaging can be eroded where additional workload is absorbed and time per exam reduced. As departments become busier we risk a "production-line" mentality that impedes compassionate care (Sinclair et al., 2016).

Following on from this work further application of the framework has been suggested for other imaging modalities in the department such as general x-ray and ultrasound.

5 CONCLUSION

The original objective of this work was to understand how the CT service has changed due to COVID-19 and to describe the repurposing of an existing DES model. Subjective and objective aspects of the impact of COVID-19 in terms of staff perception of their service, the workflow and the consumption of staff time have been quantified and graphically depicted. The creation of the RP provided an opportunity for staff to reflect on the current environment and service they provide and should be interesting to refer to in future times. COVID-19 has led to a decreased outpatient service and scanner availability, as well as increased workflow complexity and communications. The full extent of the impact of COVID-19 on the GP and outpatient populations is unknown. An approach using DES and SSM could prove successful to examine how best to manage future demand. This work recommends that inpatient and outpatient services be delivered separately and that, on a national level, regional diagnostic hubs be established to provide dedicated services for outpatient and GP diagnostic imaging.

ACKNOWLEDGMENTS

We would like to acknowledge the artistic contribution provided by Nora Zölzer-Bryce who produced the Rich Picture diagram. Nora is an artist and nurse with the Irish Wheelchair Association. Thanks, is also due to Caroline Hanrahan, Radiology Manager, Dr Parsons, Clinical Director and all the staff in Portiuncula University Hospital, clinical, clerical, and household for their participation.

REFERENCES

Ackoff, R. L. (2010). *Systems thinking for curious managers.* Triarchy Press.

Adam, E. J. (2006). Changes in the computed tomography patient population. *European Radiology, Supplement*, *16*(SUPPL. 4), 38–42. https://doi.org/10.1007/s10406-006-0186-1

Bell, S., Berg, T., & Morse, S. (2019). Towards an Understanding of Rich Picture Interpretation. *Systemic Practice and Action Research (2019)*, *32*, 601–614. https://doi.org/10.1007/s11213-018-9476-5

Bergvall-Kareborn, B. (2002). Enriching the model-building phase of soft systems methodology. (Research Paper). *Systems Research and Behavioral Science*, *19*(1), 27–49. https://go-gale-com.libgate.library.nuigalway.ie/ps/i.do?p=AONE&sw=w&issn=10927026&v=2.1&it=r&id=GALE%7CA83281263&sid=googleScholar&linkaccess=fulltext

Boland, G. W. L. (2008). Enhancing CT productivity: Strategies for increasing capacity. *American Journal of Roentgenology*, *191*(1), 3–10. https://doi.org/10.2214/AJR.07.3208

Booker, M. T., O'Connell, R. J., Desai, B., & Duddalwar, V. A. (2016). Quality Improvement with Discrete Event Simulation: A Primer for Radiologists. *Journal of the American College of Radiology*, *13*(4), 417–423. https://doi.org/10.1016/j.jacr.2015.11.028

Central Statistics Office. (2015). *Regional Population Projections 2016 - 2031 - CSO - Central Statistics Office*. http://www.cso.ie/en/releasesandpublications/er/rpp/regionalpopulationprojections2016-2031/

Conlon, M., & Molloy, O. (2019). Knowledge management in healthcare: Information requirements when creating a decision support tool in radiology. *IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *3*(Ic3k), 317–324. https://doi.org/10.5220/0008352703170324

Crowe, S., Brown, K., Tregay, J., Wray, J., Knowles, R., Ridout, D. A., Bull, C., & Utley, M. (2017). Combining qualitative and quantitative operational research methods to inform quality improvement in pathways that span multiple settings. *BMJ Quality & Safety*, *26*(8), 641–652. https://doi.org/10.1136/bmjqs-2016-005636

Currie, C. S., Fowler, J. W., Kotiadis, K., Monks, T., Stephan Onggo, B., Robertson, D. A., Tako, A. A., & Currie, S. (2020). *How simulation modelling can help reduce the impact of COVID-19*. https://doi.org/10.1080/17477778.2020.1751570

Gunal, M. M. (2012). A guide for building hospital simulation models. *Health Systems*, *1*(1), 17–25. https://doi.org/10.1057/hs.2012.8

Harper, P. R., & Pitt, M. a. (2004). *On the challenges of healthcare modelling and a proposed project life cycle for successful implementation*. *55*(6), 657–661. https://doi.org/10.1057/palgrave.jors.2601719

HSE. (2020). *Guidance on COVID-19* (Vol. 2).

Huang, Z., Zhao, S., Li, Z., Chen, W., Zhao, L., Deng, L., & Song, B. (2020). The Battle Against Coronavirus Disease 2019 (COVID-19): Emergency Management and Infection Control in a Radiology Department. *Journal of the American College of Radiology*, *17*(6), 710–716. https://doi.org/10.1016/j.jacr.2020.03.011

Monks, T., Robinson, S., & Kotiadis, K. (2014). *Learning from discrete-event simulation: Exploring the high involvement hypothesis*. https://doi.org/10.1016/j.ejor.2013.10.003

Monks, T., Robinson, S., & Kotiadis, K. (2016). Can involving clients in simulation studies help them solve their future problems? A transfer of learning experiment. *European Journal of Operational Research*, *249*(3), 919–930. https://doi.org/10.1016/j.ejor.2015.08.037

Mossa-Basha, M., Meltzer, C. C., Kim, D. C., Tuite, M. J., Kolli, K. P., & Tan, B. S. (2020). Radiology Department Preparedness for COVID-19: *Radiology* Scientific Expert Panel. *Radiology*, 200988. https://doi.org/10.1148/radiol.2020200988

Murray, T. E., Halligan, J. J., & Lee, M. J. (2017). Inefficiency, dignity and patient experience: Is it time for separate outpatient diagnostics? *British Journal of Radiology*, *90*(1080), 4–6. https://doi.org/10.1259/bjr.20170574

Oh, H.-C., Toh, H.-G., & Giap Cheong, E. S. (2011). Realization of Process Improvement at a Diagnostic Radiology Department with Aid of Simulation Modeling. *Journal For Healthcare Quality*, *33*(6), 40–47. https://doi.org/10.1111/j.1945-1474.2011.00133.x

Orsi, M. A., Oliva, A. G., & Cellina, M. (2020). Radiology Department Preparedness for COVID-19: Facing an Unexpected Outbreak of the Disease. *Radiology*, 201214.

https://doi.org/10.1148/radiol.2020201214

Rachuba, S., Knapp, K., Ashton, L., & Pitt, M. (2018). Streamlining pathways for minor injuries in emergency departments through radiographer-led discharge. *Operations Research for Health Care*, *19*, 44–56. https://doi.org/10.1016/j.orhc.2018.03.001

Reinus, W. R., Enyan, a, Flanagan, P., Pim, B., Sallee, D. S., & Segrist, J. (2000). A proposed scheduling model to improve use of computed tomography facilities. *Journal of Medical Systems*, *24*(2), 61–76. http://www.ncbi.nlm.nih.gov/pubmed/10895421

Ross, S., Lavis, J., Rodriguez, C., Woodside, J., & Denis, J. L. (2003). Partnership experiences: Involving decision-makers in the research process. *Journal of Health Services Research and Policy*, *8*(SUPPL. 2), 26–34. https://doi.org/10.1258/135581903322405144

Sim, W. Y., Chen, R. C., Aw, L. P., Abu Bakar, R., Tan, C. C., Heng, A. L., & Ooi, C. C. (2020). How to safely and sustainably reorganise a large general radiography service facing the COVID-19 pandemic. *Radiography*, https://doi.org/10.1016/j.radi.2020.05.001

Simon Dodds, A., & Phillip Debenham, S. (2016). Learning Anti-Chaos Design in Emergency Flow. *Journal of Improvement Science*, *32*, 1–14. http://www.journalofimprovementscience.net

Sinclair, S., Norris, J. M., McConnell, S. J., Chochinov, H. M., Hack, T. F., Hagen, N. A., McClement, S., & Bouchal, S. R. (2016). Compassion: a scoping review of the healthcare literature. *BMC Palliative Care*, *15*(1), 6. https://doi.org/10.1186/s12904-016-0080-0

Sterman, J. D. (2001). System Dynamics Modeling: Tools for Learning in a Complex World. *California Management Review*, *43*(4), 8–25. https://doi.org/10.1111/j.1526-4637.2011.01127.x

The Royal College of Radiologists. (2020a). Clinical Radiology UK Workforce Census 2019 Report. *Clinical Radiology*, *April*. http://www.rcr.ac.uk/publications.aspx?PageID=310&PublicationID=359

The Royal College of Radiologists. (2020b). Clinical Radiology UK Workforce Census 2019 Report. *Clinical Radiology*, *April*.

Van Lent, W. A. M., Deetman, J. W., Teertstra, H. J., Muller, S. H., Hans, E. W., & Van Harten, W. H. (2012). Reducing the throughput time of the diagnostic track involving CT scanning with computer simulation. *European Journal of Radiology*, *81*, 3131–3140. https://doi.org/10.1016/j.ejrad.2012.03.012

Womack, James, Jones, Daniel, Roos, D. (2007). *The machine that changed the world*. Simon & Schuster.

Young, T., Brailsford, S., Connell, C., Davies, R., Harper, P., & Klein, J. H. (2004). Using industrial processes to improve patient care. *BMJ (Clinical Research Ed.)*, *328*(7432), 162–164. https://doi.org/10.1136/bmj.328.7432.162

Zanardo, M., Martini, C., Monti, C. B., Cattaneo, F., Ciaralli, C., Cornacchione, P., & Durante, S. (2020). Management of patients with suspected or confirmed COVID-19, in the radiology department. *Radiography*. https://doi.org/10.1016/j.radi.2020.04.010

AUTHOR BIOGRAPHIES

MARY CONLON received a BSc (Hons) Radiography and Diagnostic Imaging from University of Wales, Bangor in 1998 and MSc Radiography from University College Dublin in 2008. She is currently a Senior radiographer in Ireland's Health Service Executive and PhD student in the National University of Ireland, Galway.

OWEN MOLLOY is lecturer and PhD supervisor at the National University of Ireland, Galway. He holds Doctoral (Industrial Engineering), Masters (Electronic Engineering) and Bachelors (Electronic Engineering) degrees from the National University of Ireland.

AVRIL HANLY received a BSc (Hons) Radiography and Diagnostic Imaging from University of Hertfordshire, and post graduate qualification in CT. She is the CT Clinical Specialist in Portiuncula University Hospital, Health Service Executive, with a special interest in continuous process improvement and applications of Lean in Healthcare.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.024

MODELLING PRE-SYMPTOMATIC INFECTIOUSNESS IN COVID-19

Dr. Russell Cheng

University of Southampton
Highfield, SO17 5BJ
United Kingdom
cheng@btinternet.com

Dr. Christopher Dye

University of Oxford
Mansfield Road, OX1 3SZ,
United Kingdom
Chrisdye56@gmail.com

Dr. John Dagpunar
University of Southampton
Highfield, SO17 1BJ
United Kingdom
jdagpunar@hotmail.com

Dr. Brian Williams
SACEMA
Stellenbosch University
South Africa
williamsbg@me.com

ABSTRACT

This paper considers SEPIR, an extension of the well-known SEIR continuous simulation compartment model. Both models can be fitted to real data as they include parameters that can be estimated from the data. SEPIR deploys an additional pre-symptomatic (also called asymptomatic) infectious stage not in SEIR but known to exist in COVID-19. This stage can also be fitted to data. We focus on how to fit SEPIR to a first wave. Both SEPIR and the existing SEIR model assume a homogeneous mixing population with parameters fixed. Moreover neither includes dynamically varying control strategies deployed against the virus. If either model is to represent more than just a single wave of the epidemic, then the parameters of the model would have to be time dependent. In view of this we also consider how reproduction numbers can be calculated to investigate the longer term overall result of an epidemic.

Keywords: Differential equation epidemic models, Parametric models, Effective Reproduction Number, Asymptomatic transmission

1 INTRODUCTION

A parametric SEIR model has been used by the authors in Dye et al. (2020) to compare the first wave of the COVID-19 epidemics in different European countries. In Dye et al. (2020) this model is fitted to data using the method of maximum likelihood estimation rather than perhaps the more widely-used Bayesian Markov-chain. The compartmental structure of the SEIR model is standard and does not include a specific compartment to represent the pre-symptomatic (also asymptomatic) infectiousness stage known to occur in those infected by COVID-19. We describe the SEIR and SEPIR models in Sections 2 and 3, focusing on the models themselves rather than on the effect of the epidemics on the affected countries. We discuss the fitting of these models to data, focusing on use of the maximum-likelihood method of estimation which produces (point) estimates of parameter values, as this gives an unequivocal specific model representation of the epidemic. In Section 3.2 we give a numerical example based on the first wave of the COVID-19 epidemic in Switzerland.

An important aspect of the basic maximum-likelihood method is that the parameters values are assumed to be unknown but fixed in value. Similarly, in the Bayesian case, the distributions of the parameters are not only unknown, but are assumed to be fixed. However different strategies varying over time have to be deployed in trying to contain a fast moving epidemic like that produced by COVID-19. This means that the model parameters do not remain constant but have to be time

dependent if the trajectory of the epidemic is to be correctly reproduced. Note that use of the models based on SEIR can be used in examining more than one wave, see for example Dagpunar (2020).

For reasons of space and simplicity we focus in this paper on 'first wave' behaviour of the SEPIR model, leaving for elsewhere discussion of situations where time-varying parameters might be used.

We do however discuss how progress of an epidemic is summarised by the *effective* reproduction number R_t, a dynamically varying version of R_0, the (basic) reproduction number. Theoretically, R_0 is unequivocally defined in terms of the idealized epidemic infecting a homogeneously mixed population with no control measures and assuming every member of the population is susceptible.

However when monitoring the progress of an epidemic R_t is more useful, and in lay terms, is generally referred to as the reproduction number. It can still be defined to be the expected number of persons infected by an infective individually, but is made time dependent because of changes in the management of the epidemic and in the susceptible proportion . We consider how to calculate R_t in Section 4 which also examines how R_0 itself can be calculated for the SEIR and SEPIR models as these can both be regarded as what are called SI^nRmodels, which include *n* multiple infectious stages, as described in Ma and Earn (2006), who discuss calculation of R_0 in these.

2 THE SEIR MODEL

The SEIR model has been described in the Supplementary Materials of Dye et al. (2020), but for ease of comparison with SEPIR model we give the description again here. The model is of a homogeneously mixed population with four compartments representing those who are susceptible, exposed, infectious and recovered (SEIR), as shown in Figure 1.

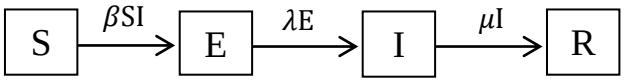


Figure 1. *The SEIR model. The compartments denote those in the population that are Susceptible, Exposed, Infected and Recovered.*

The variables *S*, *E*, *I* and *R* satisfy the ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{1}$$

$$\frac{dE(t)}{dt} = \beta S(t)I(t) - \lambda E(t) \tag{2}$$

$$\frac{dI(t)}{dt} = \lambda E(t) - \mu I(t) \tag{3}$$

$$\frac{dR(t)}{dt} = \mu I(t) \tag{4}$$

A convenient recent reference is Ma (2020) who uses a slightly different notation. Also, to highlight deaths due to the virus we divide those that recover well and those that die due to the virus. Thus the infectious individuals are divided into two compartments as illustrated Figure 2.

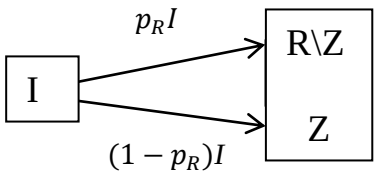


Figure 2. *Adjustment of the SEIR model where R is divided into two compartments, R \ Z, those that recover and Z, those that die; where p_R is the proportion that recover.*

More elaborate models can and have been developed. For example, see Dagpunar (2020) who extends R into additional compartments representing different outcomes of hospitalization

The SEIR model of Figures 1 and 2 are assumed to depend on certain parameters, initially assumed unknown. Fitting the model to data, is simply the process of estimating the parameters, either directly using data obtained from observing the epidemic, or from information obtained from other sources. Once the parameter values are estimated, the behaviour of the SEIR model is completely specified. The parameters are defined in Dye et al. (2020). To avoid repetition and confusion they are not discussed directly here as we shall be discussing the SEPIR model where a very similar set of parameters will be fully defined.

However we do point out here the time-delay parameter τ used to modify equation (4) to:

$$\frac{dR(t)}{dt} = \mu I(t-\tau). \tag{5}$$

We denote by $\boldsymbol{\theta} = (b1, b2, \dots, b_m)$, the vector of parameters, where m is the number of parameters. In Dye et al. (2020), $m = 9$. In the SEPIR model of Section 3, $m = 11$. With $\boldsymbol{\theta}$ given, the four differential equations (1), (2), (3) and (5) can be solved by numerical integration to give the trajectories

$$S(t,\boldsymbol{\theta}), E(t,\boldsymbol{\theta}), I(t,\boldsymbol{\theta}), R(t,\boldsymbol{\theta}), Z(t,\boldsymbol{\theta}) \quad \text{for } t = 1,2,\dots,M \tag{6}$$

where t is the day and M is the number of days of interest. We used the standard method of Maximum Likelihood (ML), as given for example in Cheng (2017), to estimate parameter values.

Here we outline the approach used to estimate the parameters from a sample of observed daily deaths. Let the sample of observed number of daily deaths be denoted by

$$\mathbf{Z} = \{z_t \;\; t = 1,2,\dots M\} \tag{7}$$

where z_t is the number of deaths on day t and M is the number of days observed. If the observations were made without error and the right parameter values are correct for $\boldsymbol{\theta}$, then the death trajectory $\{Z(t,\boldsymbol{\theta}) \;\; t = 1,2,\dots,M\}$ would match the observed deaths $\mathbf{Z}$ in (7) and the model would then be successful in explaining deaths.

To include statistical uncertainty in the model we assume instead

$$z_t = z(t,\boldsymbol{\theta}) + e(t) \quad t = 1,2,\dots,M \tag{8}$$

where $e(t)$ is random error. For simplicity the $e(t)$ are assumed to be normally and independently distributed (NID) with standard deviation σ, i.e.

$$e(t) \sim \text{NID}(0, \sigma^2), \text{ so that } \; z_t - z(t,\boldsymbol{\theta}) \sim \text{NID}(0, \sigma^2) \tag{9}$$

Note that σ is treated as a parameters so is included as a component of $\boldsymbol{\theta}$.

The logarithm of the distribution of the sample is then

$$\text{L}(\mathbf{Z}|\boldsymbol{\theta}) = -\,(M/2)\ln(2\pi) - M\ln\sigma - [1/(2\sigma^2)]\sum_{i=1}^{M}[z_t - z(t,\boldsymbol{\theta})]^2 \tag{10}$$

where $\mathbf{Z}$ is the random argument, and the parameters $\boldsymbol{\theta}$ are fixed. In ML estimation (MLE), this is turned on its head so that $\mathbf{Z}$ is simply the known sample of observations now regarded as fixed and we write L as $\text{L}(\mathbf{Z}|\boldsymbol{\theta}) = \text{L}(\boldsymbol{\theta}|\mathbf{Z})$) calling it the (log)likelihood to indicate that it is now treated as a function of $\boldsymbol{\theta}$. The ML estimator $\widehat{\boldsymbol{\theta}}$ is simply the value of $\boldsymbol{\theta}$ at which $\text{L}(\boldsymbol{\theta}|\,\mathbf{Z})$ is maximized. i.e.

$$\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\{(\text{L}(\boldsymbol{\theta}|\,\mathbf{Z})\}. \tag{11}$$

Nelder-Mead numerical search for the maximum was used. This goes through different $\boldsymbol{\theta}_i$ i=1, 2, 3,… comparing the different $\text{L}(\boldsymbol{\theta}_i, |\mathbf{Z})$ to find $\widehat{\boldsymbol{\theta}}$, the best $\boldsymbol{\theta}$.

To simplify the description of the estimation process, only fitting to deaths data, **Z** as in (7) has been described, but the method extends straightforwardly to include other data samples. For example

$$\mathbf{Y} = \{y_t \quad t = 1,2,\dots,M\} \tag{12}$$

where $\boldsymbol{y_t}$ is the number of prevalent active cases on day *t*. Fitting simultaneously to both **Y** and **Z** can be carried out by adding to the right-hand side of (10) a corresponding set of terms for **Y**

Each step of the Nelder-Mead optimization is summarized as follows. The trajectory of each of the variables *S*, *E*, *I*, *R*, as given in Equation (6) is calculated, for simplicity using Euler step-wise integration of the differential equations (1) – (4) , but with step-length 1/8th of day as a step length of 1 is quite inaccurate. These are essentially scale invariant so we can standardise the equations taking

$$(S + E + I + R) = 1.$$

This choice of 1 for the standardising constant makes all four variables *fractions*. Initial values are S(0, $\boldsymbol{\theta}$) = 1, and E(0, $\boldsymbol{\theta}$) a small quantity, these denoted by s_0 and e_0 respectively. However these are subsequently adjustable, so are treated as parameters. Also we set $I(0, \boldsymbol{\theta}) = R(0, \boldsymbol{\theta}) = 0$.

With the state variable trajectories obtained, the likelihood is then calculated. This requires values of all the parameters, in particular *N*, the population size, which is treated as a parameter. This is appears in calculating the differences $[(1 - p_r)NR(t, \boldsymbol{\theta}) - \bar{Z}(t)],\ t = 1, \dots, M$, used in the likelihood to compare the model deaths with the observed cumulative deaths $\bar{Z}$. Thus *N* is taken into account and can be changed in selecting the next $\boldsymbol{\theta}$.

3 THE SEPIR MODEL

3.1 Structure of the SEPIR Model

In the SEPIR model we introduce an extra compartment to the SEIR model in Fig. 1 changing it to Fig 3:

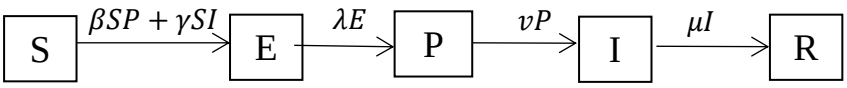


Figure 3. *The SEPIR model. The compartment P denotes those who are infectious but are pre-symptomatic whilst I denotes that are infectious and symptomatic*.

The original I compartment is now split into two with its first compartment, P, comprising those infectious who are pre, i.e. asymptomatic, and the second, I, comprising those infectious who display symptoms. The ordinary differential equations (1), (2) and (3) in the SEIR model are replaced by the differential equations (13), (14), (15) and (16), with (4) and (5) remaining unchanged. There are two terms in going from S to E, comprising: those infected by someone in P, with transmission rate β, and those infected by someone in I, with transmission rate γ. The reciprocal v^{-1} is the mean period someone spends in state (compartment) P whilst μ^{-1} is the mean period spent in I.

$$\frac{dS(t)}{dt} = -\beta S(t)P(t) - \gamma S(t)I(t) \tag{13}$$

$$\frac{dE(t)}{dt} = \beta S(t)I(t) + \gamma S(t)I(t) - \lambda E(t) \tag{14}$$

$$\frac{dP(t)}{dt} = \lambda E(t) - vP(t) \tag{15}$$

$$\frac{dI(t)}{dt} = vP(t) - \mu I(t) \tag{16}$$

We treat the quantities $\beta, \gamma, \lambda, v, \mu$ as parameters to be estimated. However we include six further parameters $t_0, e_0, s_0, \sigma, p_R$ and τ. These are all listed and defined in columns 1 and 2 of Table 1.

Some of the parameters can be given fixed predetermined values with the others fitted to data by Maximum Likelihood as described in the SEIR model in Section2.

3.2 Switzerland: A Numerical Example

Column 3 gives the parameter values when SEPIR was fitted estimating all 11 parameters by maximum likelihood using M = 109 days of data based on daily observations starting on 15 Feb 2020. Two series: Daily New Cases and Daily Deaths were used. The values of all the parameters are of interest. The parameter values for Switzerland are given in Table 1. We highlight two aspects.

Firstly, the SEPIR model, because the differential equations are scale invariant, can be standardised so that the population size is 1. However we allow the population size to be variable with the size estimated by allowing rescaling, when we maximize the likelihood. The estimated population size of 36,700 is remarkably small compared with the actual population size of 8.2 million. The main reason for the difference is that the model does not include a mechanism of epidemic control which we know took place in every country to prevent the spread of infection. The SEPIR model, which assumes a homogeneously mixed population can only allow for this by changing the population size. Moreover without examining regional records it may be that the outbreak in Switzerland was mainly confined to parts nearest Italy, the first European country to be badly affected by COVID.

Table 1. *Parameters of the SEPIR model with estimates for Switzerland*.

Symbol	Definition	Estimated value and 95% confidence interval
β	pre-symptomatic transmission rate	0.44 (0.435– 0.454)
γ	symptomatic transmission rate	0.137 (0.128– 0.147)
λ^{-1}	mean latent period in compartment E	1.19 (1.13– 1.26)
v^{-1}	mean pre-symptomatic period	4.40 (4.00– 4.62)
μ^{-1}	mean symptomatic period	14.0 (13.8– 14.3) days
t_0	number of days from start of epidemic before observations began	30 days (too small to measure)
e_0	initial proportion of individuals exposed	6.6 (5.6– 8.2) E-07
N	numerical size of exposed population	3.67 (3.60 – 3.74) E+04
σ	standard deviation of observational error	103 (82 – 106)
p_R	probability of an infective recovered well	0.943 (0.942 – 0.945)
τ	mean time between the end of infectiousness and recovering well or death	3.0 (2.8 – 4.0) days

Secondly we examine whether the SEPIR model gives any indication of the extent of the pre-symptomatic stage.

To do this, we first summarize what is already known about this stage by reporting the findings of He at al. (2020) who investigated the case histories of 77 infector-infectee pairs in each of which an infectious person, the infector, goes on to infect a susceptible person, the infectee.

Citing the mean incubation period as 5.2 days, He et al, (2020) estimate the serial interval to be 5.8 days. From this they infer that infectiousness starts 2.3 days after the onset of infection and peaks just 0.7 days before symptom onset, giving an estimated proportion of infections of 44% as occurring before the onset of infector symptoms. Infectiousness then declined within 7 days. Figure 4 is a schematic showing the infector-infectee relationship.

The estimate of He et al. (2020) that the proportion of individuals infected pre-symptomatically is 44% means, in our case, that the proportion $(\beta v^{-1})/(\beta v^{-1} + \gamma\mu^{-1})$ should therefore be this value at least approximately. From Table 1, the value is 50.4%. This is in accord with the higher pre-symptomatic infection proportions given by Tapiwa et al (2020): 48% in Singapore and 62% in Tianjin. The practical consequences of this finding is evident, with elaborate track and tracing required to identify pre-symptomatic infections.

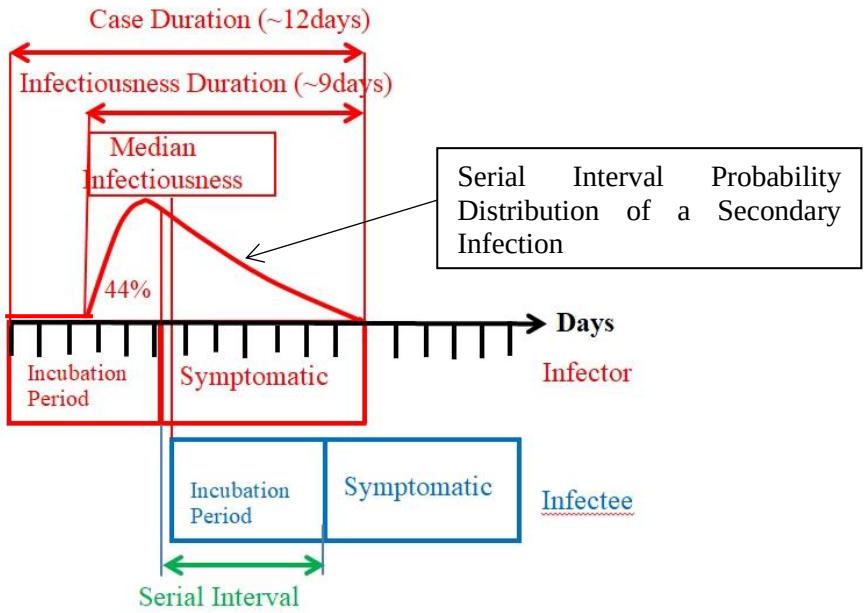


Figure 4: *Infector-Infectee Relationship as described by He et al (2020).*

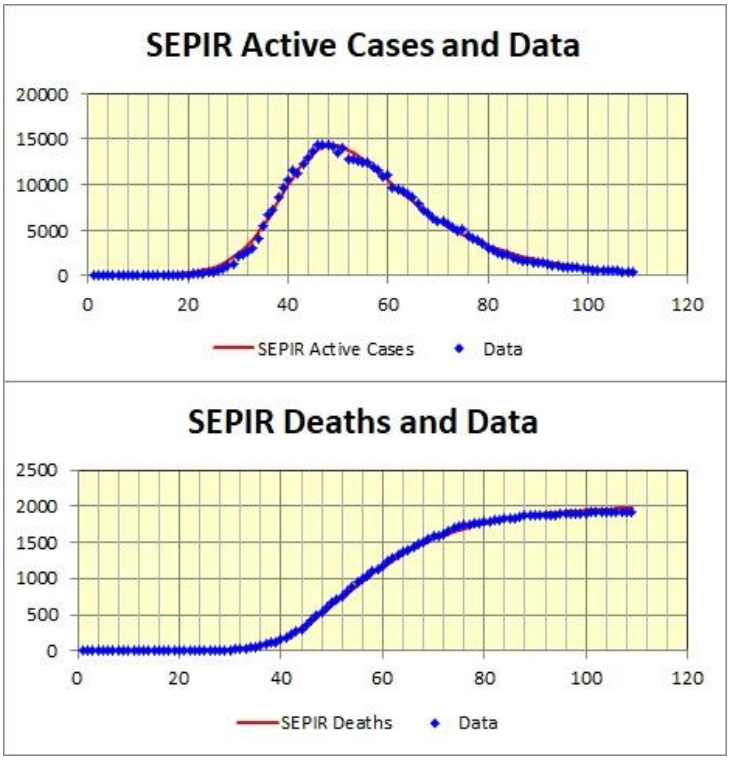


Figure 5: *SEPIR Active Cases and Cumulative Deaths fitted to Swiss Data. Horizontal axis is days with day #1 = 15th February 2020.*

The quality of the fit achieved by the SEPIR model is illustrated in Figure 5 where the Active Cases and Cumulative Deaths curves obtained by fitting the model to both data sets simultaneously are plotted with their corresponding data.

The SEPIR model was then repeatedly fitted to independent parametric bootstrap replications of the actual observed active cases and cumulative deaths data. As described in Subsection 4.1.3 of Cheng (2017), confidence intervals for the parameters can be obtained from the bootstrap parameter

estimates. For illustration, the resulting 95% confidence intervals for each of the fitted parameter estimates are reported in Column 3 of Table 1, using 500 bootstraps.

Charts of fitted SEPIR trajectories provide an easily understood way to display results. For example the fitted SEPIR cumulative deaths trajectory (red) is displayed in Figure 6 together with the observations (black). The method described in Section 4.3 of Cheng (2017) can be used to provide a confidence band for any model trajectory. For example we have a bootstrap cumulative deaths trajectory corresponding to each bootstrap sample. These are plotted (in grey) in Figure 6 giving a bundle of trajectories, with 95% confidence limits (green and blue). Only 250 bootstrap are depicted.

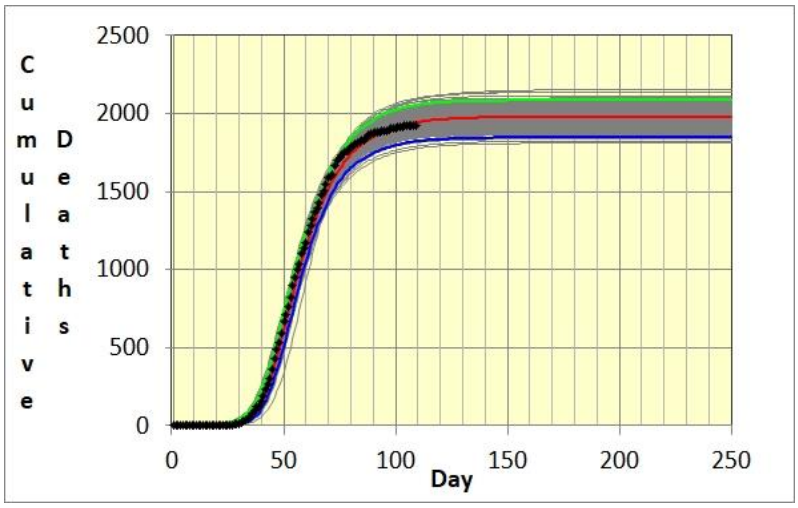


Figure 6: *SEPIR Fitted Cumulative deaths trajectory (red) for Swiss data obtained from 109 observations (black). Upper (green) and lower (blue) confidence limits.*

4 R_T THE EFFECTIVE REPRODUCTION NUMBER

The SEIR and SEPIR models are both SI^nR models with n multiple infectious stages as defined in Equations (6a)-6(d) in Ma and Earn (2006), whose Equation (7) gives formulas for the Reproduction Number, R_0, in the models. (The Reproduction Number R_0 is simply, but precisely, defined as the number of susceptible individuals that an infectious person will go onto infect when the epidemic first starts, assuming that the population is homogeneously mixed.) For the SEPIR model, using the parameters already appear in Equations (13)-(16), we have, in our Swiss example, that

$$R_0 = \beta/v + \gamma/\mu = 0.44 * 4.4 + 0.13 * 14 = 3.85. \tag{17}$$

which seems not implausible.

As mentioned in the Introduction, in practice R_t, the effective reproduction number, is more useful as, throughout the epidemic, it can be continually used to gauge how well control strategies are working. The theoretical basis underlying the calculation R_t is well described by Ma (2020). We have

$$R_t = \frac{c(t)}{\int_0^\infty c(t-u)w(u)du}, \tag{18}$$

where $c(t)$ is the incidence curve of new cases at time t and $w(u)$ is the *serial interval probability distribution* of a secondary infection; so that $w(u)du$ is the probability that an infectious individual (the infector)) infects someone else (the infectee) in the time period (u, $u+du$). This probability distribution of the time between the infections of an infector-infectee pair is depicted in Figure 4.

The denominator in Equation (18) measures how the new cases at time t arise from those infected prior to time t. The epidemic clearly is rising or falling depending on whether the numerator is larger or small than the denominator, with equilibrium when they are equal. Thus R_t has the critical

reproduction property of R_0 but moreover is dynamic, so that it can be used to gauge the progress of the epidemic as this develops.

It turns out that the formula (18) is quite robust so that the serial interval distribution does not have to be estimated all that accurately. In fact Germany, early during its first COVID-19 epidemic wave, used the simple denominator $c(t\text{-}4)$. Cori et al. (2013), though using a Bayesian approach, examined various empirically obtained serial interval distributions drawn from different epidemics. In Dye et al. (2020) the authors used a discretized and shifted gamma distribution $g(t),\ t = 1,2,\ldots,12$ to represent the serial interval distribution $w(u)$ that is shown as the red curve in Figure 4, calculating the denominator as

$$D = \sum_{u=1}^{12} c(t-u)g(u). \tag{19}$$

Figure 7 depicts R_t calculated using this formula for Switzerland when c(t) is a daily 7-day moving average of new cases.

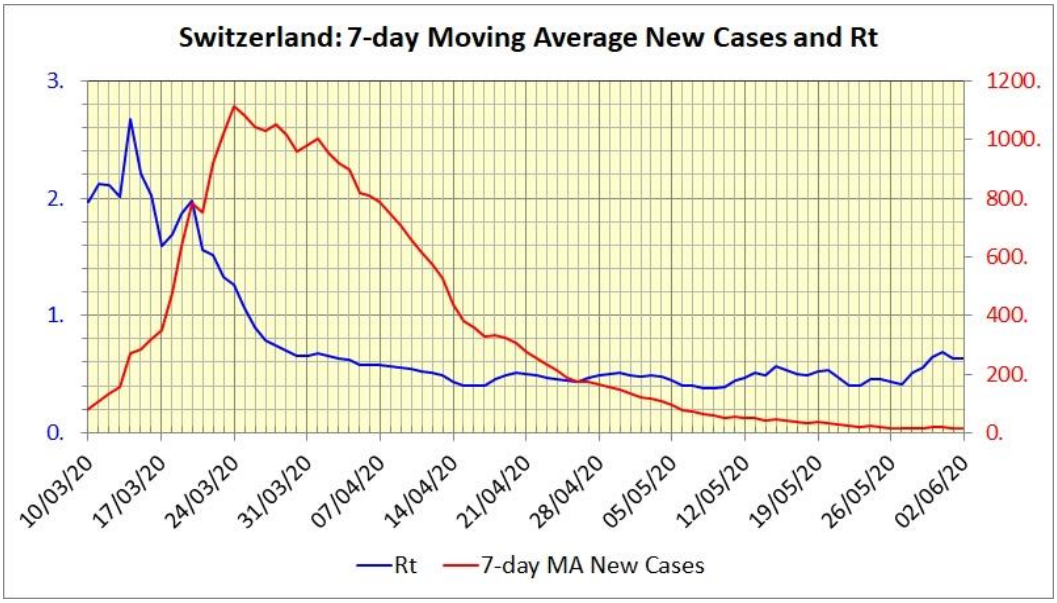


Figure 7: Chart of the effective R_t calculated using the formula in Equation (19) for Switzerland where $c(t)$ is the 7-day moving average of new cases.

An important point is that how the epidemic ultimately ends depends on R_0, not R_t. Until the arrival of vaccines all controls, lockdown, hand-washing, social distancing and so on, affect only β, moreover only $temporarily$. Thus R_t, which depends on this temporary β, varies as the controls vary, and so is a simple gauge of how current controls are doing. However once controls are removed, β returns to its original value as in R_0, so that the epidemic returns, causing another wave. As determined by Kermack and McKendrick (1927) for the SIR model, the ultimate end, when a given number of susceptibles have been infected, is determined by R_0; the remaining uninfected susceptibles, $S(\infty)$, satisfying the so-called final-size relation:

$$\ln\,[S(0)/S(\infty)] = R_0[1 - S(\infty)/N]. \tag{20}$$

This final-size relation also applies to the SEIR and SEPIR models as these are both SI^nR models with multiple infectious stages as defined in Equations (6a)-6(d) in Ma and Earn 2006, whose Equation (7) gives formulas for R_0 in the models.

Vaccination works differently by moving susceptibles directly to the R compartment so that the effective population shrinks. This is most clearly seen if immunization takes place before the epidemic so that a proportion p of the population is immunized (see Brauer et al. 2019). Then R_0 would shrink to $R_0(p) = \beta_B N(1-p)/\mu$. (Here β_B is the β in Brauer et al. 2019 which assumes mass

action incidence where an infectee makes $\beta_B N$ contacts sufficient to transmit infection in unit time.) $R_0(p)$ is less than 1 if

$$p > 1 - 1/R_0,$$

when there would be no epidemic. Thus not all of the population needs immunization to prevent the epidemic. For example if $R_0 = 3$ over two thirds of the population would need immunization. This clearly shows that the only long-lasting control is vaccination, At time of writing, the race is on in the UK, with covid-19 continuing to infect large numbers, but with vaccination rates being accelerated to counter this. Also the new variants suggest that a basic reproduction number of R_0 as high as 5 is not out of the question. That would mean that under no Non-Pharmaceutical Interventions, 80 % of the population would need vaccination before the growth rate became negative. With such a high R_0, almost the entire population would need vaccination for the disease to be eliminated, and even this assumes that new cases are not imported from other countries. It seems likely then that Covid-19 will be with us for a long time and effective test, trace, and isolation will be needed in the long term.

5 CLOSING REMARKS

In conclusion the SEPIR model is an extension of the well-known SEIR model. Both are particular cases of the more general SI^nR model with multiple infectious stages. For Covid-19, the SEPIR compartment, P, is used to represent those infected pre-symptomatically. Compared with the SEIR model this is an important improvement, as shown in our example. This latter is based on real data, and clearly shows the large part played by pre-symptomatic infection in the case of Covid-19. The practical implication is that Covid-19 control strategies need to recognise and deal with pre-symptomatic transmission to be truly successful. Our formulation of the SEPIR model includes adjustable parameters which can either be given or fitted to observed data. An important aspect of our parametrisation is that it allows estimation of an initial susceptible proportion that is less than unity rather than supposing that the susceptibles comprise the country's whole population, as is usually assumed. This relaxes the assumption of homogeneity of virus transmission throughout the whole population, as this assumption may not be reasonable, especially in the early stages of an epidemic where the number of individuals infected is initially small.

It should be noted that our transmission rates, though estimated, are supposed constant rather than time dependent. This latter would be needed to model changing management of the epidemic. This could explain why the estimates of some of the biological parameters are rather different from those observed in some other studies.

We end with two caveats.

Firstly we have not yet examined in detail the robustness of the maximum likelihood optimization used to fit the model. In our numerical example we chose the first wave of the epidemic in Switzerland because the data corresponded well to the characteristics of the SEPIR model. However even in this example alternative good fits can be achieved with combinations of parameter values different from those reported in Table 1. Thus, in practice, comparison with parameter estimates obtained in other ways should always be made where possible to assess when the estimates obtained can be relied on.

Secondly, the simplicity of models such as SEIR or SEPIR. means that the practical usefulness of using them on their own, in isolation, is limited. The models are idealizations of the way the epidemic behaves and of population behaviour. Thus control policies are not modelled nor their influence on population behaviour. Indeed lack of homogeneity of population behaviour is an important factor that has to be addressed in implementing control policies because these latter have to recognize the issues they give rise to, for the population as a whole to be willing to follow them. At this time of writing, resurgence of the virus has taken place and more virulent virus strains have appeared, but balanced by the availability of vaccines, This all requires a national control policy which is fair. On this basis, it is not unfair as at time of writing and as adopted by present national policy to go in "earlier and harder and stay longer than might be thought necessary" even in areas with low prevalence. The alternative of delaying, all too easily is likely to ultimately inflict on such areas similar damage to that currently experienced by high tier areas

Ideally a detailed model allowing for local differences is required, but seems unrealistic given the speed of changing events. . However, less complicated models like the present SEPIR model may be helpful in informing decision makers who may otherwise only have time to use just simple common sense in making their decisions.

REFERENCES

Brauer F, Castillo-Chavez C, Feng Z. (2019) Simple Compartmental Models for Disease Transmission. In: *Mathematical Models in Epidemiology*. *Texts in Applied Mathematics*, vol **69**. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-9828-9_2

Cheng R C H (2017). *Non-Standard Parametric Statistical Inference*. Oxford University Press, Oxford.

Cori A, Ferguson N M, Fraser C and Cauchernez S (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* **178(9)**: 1505-1512.

Dagpunar J S (2020). Sensitivity of UK Covid-19 deaths to the timing of suppression measures and their relaxation. *Infectious Disease Modelling* **5:** 525-535 https://doi.org/10.1016/j.idm.2020.07.002

Dye C, Cheng R C H, Dagpunar J and Williams B G (2020). The scale and dynamics of COVID-19 epidemics across Europe. *R. Soc. Open Sci.* . **7**: 201726 https://doi.org/10.1098/rsos.201726

He L, Lau E H Y, […], Leung G M (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**: 672-675.

Kermack W O and McKendrick A G (1927). A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. Ser. A* **115**: 700-721.

Ma J and Earn D J D (2006). Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology* **68**: 679-702.

Ma J (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling* **5**: 129-141.

Tapiwa G, Kremer C, Chen D, Torner A, Faes C, Wallinga J and Hens N (2020). Estimating the gerneration interval for coronavirus disease (COVID-19) based on symptom onset, March 2020. *Euro Surveill*. **25(17**): pii=2000257.

AUTHOR BIOGRAPHIES

RUSSELL CHENG retired from the University of Southampton in 2007 where he had been Head of the Operational Research Group, having held previous positions at Cardiff University and the University of Kent at Canterbury. https://www.southampton.ac.uk/maths/about/staff/rchc.page

CHRISTOPHER DYE FRS, FMedSci, has held positions at the London School of Hygiene and Tropical Medicine, the World Health Organization and Gresham College London. He is currently Professor of Epidemiology at Oxford University http://en.wikipedia.org/wiki/Christopher_Dye

JOHN DAGPUNAR retired from Edinburgh University in 2008. He is Visiting Research Fellow in Mathematical Sciences at the University of Southampton. His research interests are in simulation, financial mathematics, health studies, and reliability.
https://www.southampton.ac.uk/maths/about/staff/jd2y15.page

BRIAN WILLIAMS is Senior Research Fellow at the South African Centre for Epidemiological Modelling and Analysis (SACEMA) having held the position of Epidemiologist at the World Health Organisation from which he retired in 2008.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.025

SUPPORTING THE VENTILATOR CHALLENGE DURING THE COVID-19 PANDEMIC WITH DISCRETE EVENT SIMULATION MODELLING

Matthew Windeler
Michael Higgins
Giles Thomas

Ford Motor Company
Dunton Technical Centre, Essex, UK.
mhiggi37@ford.com
mwindele@ford.com
gthom242@ford.com

ABSTRACT

A Discrete Event Simulation (DES) model was developed by Powertrain Operations Manufacturing Engineering (PTME) in support of the Ventilator Challenge production line at a Ford Motor Company facility. The facility needed to be installed and commissioned both quickly and efficiently, with DES being one of the key tools utilised to achieve this. A flexible, dynamic input model with Excel-based data input and results dashboard enabled rapid changes to the model, enabling a clear understanding of the impact of evolving factors such as operator efficiency, labour requirements and part quality throughout the project's duration. By adopting different simulation concepts the model was applied across the entire product lifecycle, from conceptualisation to decommissioning. Through the development and implementation of this model, the characteristics of the production facility were better understood, allowing targeted engineering actions to improve productivity. This project also highlighted some key learnings for applying DES within rapidly changing manufacturing environments.

Keywords: Discrete Event Simulation, Manufacturing, Ventilator, COVID-19

1 INTRODUCTION

In March 2020, the MRC Centre for Global Infectious Disease Analysis at Imperial College London released a report describing how their epidemiological modelling predicted that unmitigated, the demand for critical care (ICU) hospital beds (with mechanical ventilation capability) in the UK due to COVID-19 would be in excess of 30 times the current NHS capacity (Ferguson et al., 2020). It was apparent that in addition to 'flattening the curve' it was equally necessary to 'raise the line', and with a global shortage of medical ventilators - and existing manufacturers' production capabilities insufficient to meet even a fraction of demand - the Government requested assistance from businesses to support in the production and supply of ventilators to the NHS (Department for Business, Energy & Industrial Strategy, 2020).

In response, VentilatorChallengeUK - a consortium of leading UK automotive, aerospace, and medical engineering companies - was formed, committed to delivering on formal orders from the Government for over 10,000 medical ventilators (Ventilator Challenge UK, media information notice: https://ventilatorchallengeuk.com, accessed February 2021). Ford Motor Company was part of this consortium, and many employees across Ford of Europe - including individuals and teams from Ford Powertrain Manufacturing Engineering (PTME) - were redeployed from key company projects to collaborate full-time on this initiative, throughout its duration.

While every other group responding to the Government's request attempted to develop entirely new prototype ventilators, VentilatorChallengeUK focussed instead on proven existing designs, and how production could be upscaled. Elements of the Penlon AV-S anaesthesia ventilator were redesigned to

allow more rapid build, and with a viable process established, multiple new facilities were set up to manufacture parts, and assemble the new ESO 2 'Emergency Ventilator' in larger numbers (Penlon, VentilatorChallengeUK and the ESO 2 Emergency Ventilator: https://www.penlon.com/Blog/May-2020/Penlon,-VentilatorChallengeUK-and-the-ESO-2-Emerge, accessed February 2021). Ford undertook the responsibility for the assembly and testing of these ESO 2 ventilator units (and front screens) and repurposed a facility at the Ford Dagenham plant in order to support this. As the consortium's key message stated, "every ventilator built has the potential to save a life", and it was therefore of paramount importance that the facility was installed quickly, and was capable of meeting the required capacity as soon as possible. Every available resource was put at the team's disposal in order to achieve that goal.

Proving out proposed manufacturing facilities using Discrete Event Simulation (DES) models has been a fundamental requirement for all Ford PTME projects over the last decade, and simulation modelling of such facilities has been an approach used within PTME as early as the 1970s (Ladbrook and Januszczak, 2001). By utilising the modelling skills developed for automotive manufacture and applying them to the production of ventilators, the intention was to establish the capability of the planned facility per the design, and continue to optimise it during the commissioning phase. By leveraging real data collected from the facility and feeding it back into the DES model, the production team could make informed decisions through an iterative, data-driven process.

The aim of this paper is to demonstrate how DES was applied at different phases of the project, working within a compressed time frame and with non-ideal systems, to support the effective delivery of the EOS 2 ventilator facility.

2 METHODOLOGY

The preferred process simulation software application of the PTME team is Lanner's Witness Horizon (Lanner, Witness Simulation Software: https://www.lanner.com/en-us/technology/witness-simulation-software.html, accessed February 2021) modelling studio, and it is within this technology that the team had the required expertise. As such, it was the platform used for all simulation modelling conducted during this project.

The ESO 2 ventilator facility could be considered as two distinct areas, 'Assembly', and 'Test', with these areas then split further into smaller zones as summarised in Figure 1. Each zone was characterised by a sequence of process steps assigned to manual work stations. Each unit, or part thereof, was required to pass through each step of the process. An initial model was constructed, based upon the part flow and process defined by the engineering team, with provision for removal, regression, repair and reinsertion of parts at various stages of the process, as required.

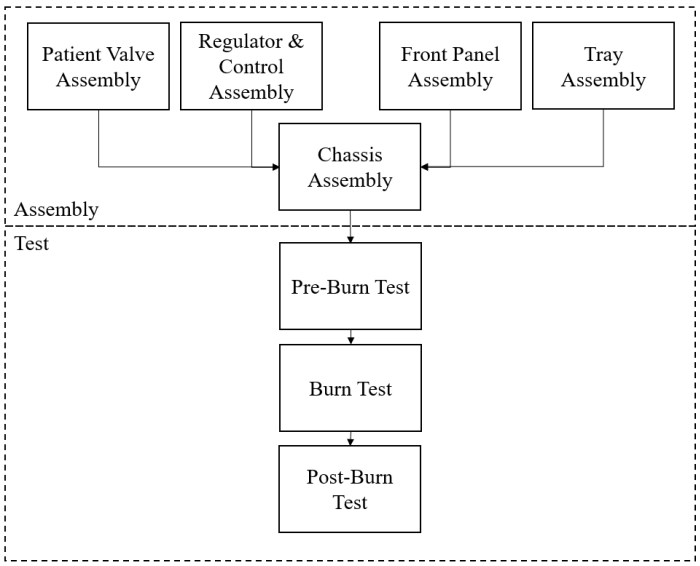
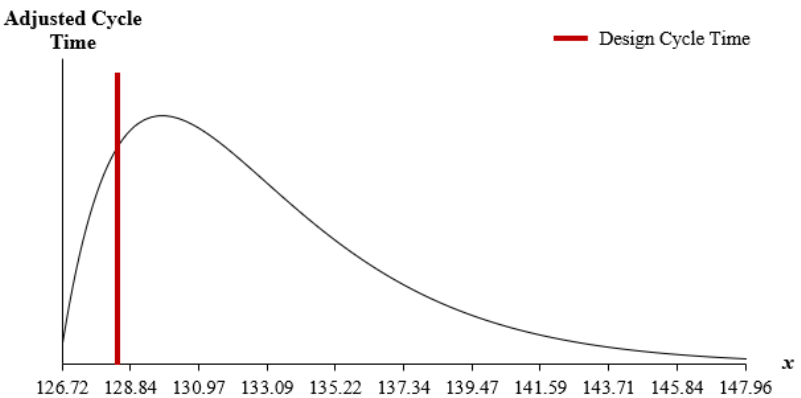


Figure 1. Simple overview of the main ESO 2 ventilator facility zones

3 IMPLEMENTATION

The first challenge faced by the simulation team was the lack of a stable system description. While the wider engineering team were working with the ventilator OEM (Penlon), understanding and refining the manufacturing process, the system design was in a constant state of flux. Recognising this rapid process evolution – additional stations, workload rebalancing, modified part flow, buffer relocation – would be an inherent feature of this project, the model was built to allow maximum flexibility and adaption of these changes as they were made. By creating a model with many variables and arrays that could be initialised at model start time (T=0) directly from an Excel workbook 'Front End', itself with links to numerous data sources, changes to the model could be made very quickly and easily, in addition to opening up further potential for model experimentation, data analysis, and optimisation.

As well as layout and flow process, there were other key input data to consider for the model too. Each station had a design cycle time (the time the station was engineered to run at). However, with the objective of the model being to better understand the inherent potential variability in the system, entering a single value for cycle time would not be adequate. For conventional models, Ford's established practice is to utilise cycle time distributions based upon recorded real-world data, either from monitoring systems on the line being modelled or a suitable surrogate operation (in the event of an entirely novel system) (Higgins, 2013). For this project, however, neither of these options were possible from the outset, so a suitable alternative was to be established. By introducing an Erlang probability density function with appropriate selection of k-value, a portion of each station cycle time was distributed relative to the design cycle time to introduce random variation about the mean, skewed toward likely overycycling - an Erlang distribution is commonly used to represent the time to complete a complex task (Robinson, 2007).This can be seen in Figure 2. The mean (CT), ratio (CTa/CTb), and k-value inputs for the Erlang were controllable by independent model variables, enabling sensitivity analysis to be conducted on the degree of overcycles observed. There were also stations that could generate rejected parts. These were initially based on estimates from the test specialist and were presented as a percentage likelihood of a part passing through a station without issue. By using dynamic model variables to control this too, sensitivity analysis could be performed on rejection rates throughout the model run.



$$Cycle\ time\ distribution = CTa + Erlang(x\ \times CTb, k)$$

Figure 2. Erlang distribution modification of design cycle time (k = 2)

As previously mentioned, the model flexibility was a key consideration. While initially the model was built on assumptions for cycle time and reject rate, there were concurrent workstreams assigned to implementing data collection methods within the physical facility. As production commenced and progressed, a dataset would begin to grow that could then be fed back into the model to hone model accuracy and increase confidence in reported results and recommendations.

Despite the high number of initial assumptions made and the continually evolving process during the initial model build, both the model approach and inputs were discussed and agreed with key

stakeholders as early as possible to ensure viable model progression and buy-in to the results. Whenever necessary, further detailed reviews were carried out when significant changes occurred to ensure the model remained an accurate-as-possible representation of the facility. The resulting simulation model was utilised throughout the lifecycle of the project to help characterise and optimise the line.

3.1 Planning Phase

In the planning phase, while the process was being developed and provisions were being made for repair areas and labour, the simulation was utilised to validate decisions and identify areas of the line that could be problematic, or might prove to be problematic should assumptions made during design exceed given limits.

Cyclic verification of the model was first undertaken to ensure the model had been built without error, and that results were accurately reflecting expectation from the input data. Once complete, this permitted cyclic experiments to be run (i.e. all constraint parameters turned off) with distributed cycle times considered. As a result the benchmark production rate could be quantified in terms of jobs per hour (JPH) for each area. At this stage it was already possible to identify where cycle time reductions would provide the biggest return – especially important when considering the limited engineering resource and time available. Figure 3 shows the JPH average and distribution for each zone within the Assembly area.

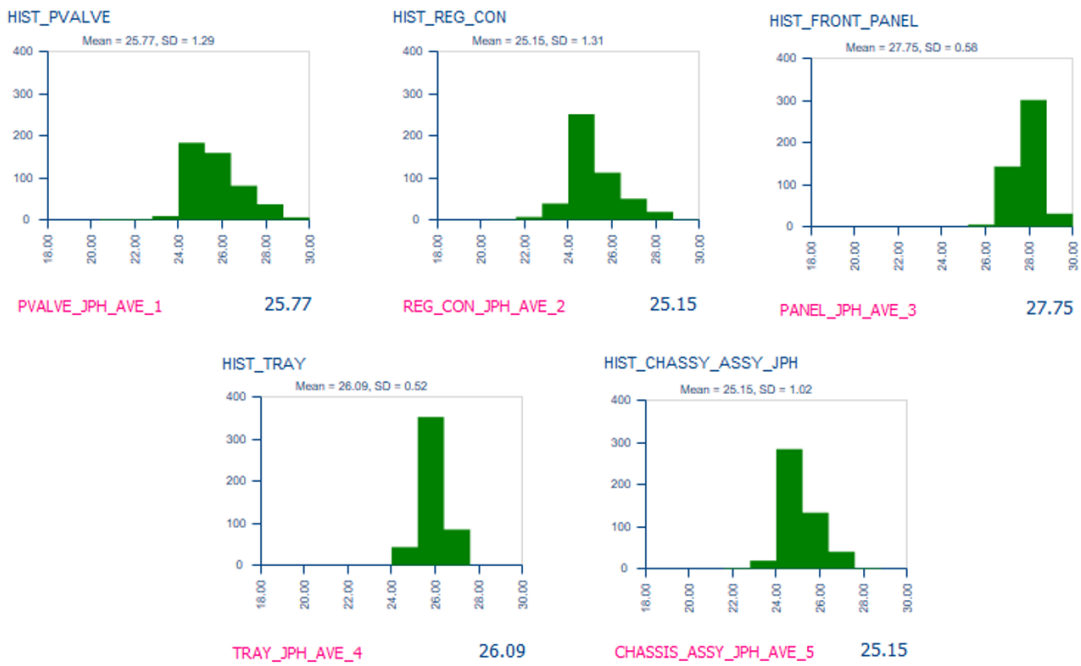


Figure 3. JPH distribution histograms for various areas in the model

Based on these results, information was passed back to the process team demonstrating to them that at a cyclic level the 'regulator and control' zone was constraining the assembly area, and that mitigation would require focus on reducing the cycle time of two stations. Further advice detailing prudent locations for buffers was also provided, in order to keep the constraint areas of the line producing parts.

With the introduction to the model of part rejection rate, direction on how to setup part regression based on initial quality assumptions could be provided. It quickly became clear that rejection rate was going to be a far more significant issue than cycle time, although the data available to support the model would prove to be difficult to secure. Sensitivity analysis on parts rejected per hour and scenario comparison for regression provided clarity on how best to setup the regression strategy. For example, it was shown that given an estimated 10% rejection rate from a station in the assembly area, the system would not be able to cope should the repair and re-test procedure take place in-station, whereas the introduction of a single repair bay (that included the capability to re-test before being returned to the

line) would enable the constraint to be overcome. Sensitivity analysis would show that this strategy could cope with up to reject rates as high as 15% before becoming the system constraint once more.

Similar analysis was undertaken for the test area of the line, while also considering how labour resources should be allocated to the repair areas to maximise efficiency. The compound effect of rejection rates at each station through the process – a total of 17 different work stations each with their own rejection rate across the three areas within test - allowed for prediction of parts per hour into the repair areas to be provided. Sensitivity analysis around the station cycle time, reject rates and repair times could simulate the expected utilisation of labour for repairs and how best to create a labour resource strategy. Initial analysis demonstrated that labour could be shared between repair bays without having a negative impact to overall throughput, so initial estimates could be revised, and labour allocated elsewhere.

3.2 Commissioning Phase

After utilising the model throughout planning for estimation of how the line could be expected to perform, and warning of potential areas to closely monitor, it was then used in a slightly different manner once the line had actually started producing ventilators. By shifting away from a non-terminating simulation, experiments could be run to demonstrate what the expected output of the line over a specific period of time would be, and this information was key for understanding requirements for incoming parts as well as when complete ventilators could be shipped to the next stage of the assembly process.

To model projected efficiency improvements over time, dynamic modifiers were applied to both the Erlang cycle time distribution and the station reject rate. By adopting the Monte Carlo concept of simulation (Law and Kelton, 2007) and running multiple replications utilising disparate pseudo-random number streams for sampling, the variation due to randomness in the model could be captured and considered accordingly. This is representative of what would be observed within the facility, as the operators become more familiar with tasks, they are able to complete them with less variability, leading to reduced cycle times and improved quality. The foundation for this analysis was based on the same assumptions made during the planning phase with respect to cycle time and reject rate, and the incremental efficiency improvements aligned to the standard Ford Powertrain processes. It also made it possible to activate additional work stations as they were commissioned and increased the capacity of the line, as well as modelling behaviour between different shifts and shift patterns.

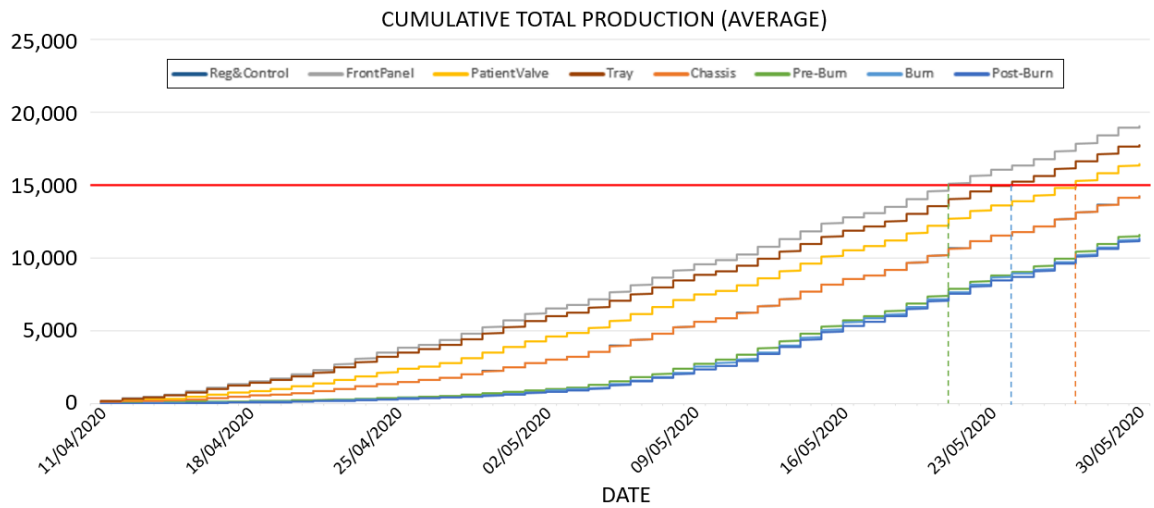


Figure 4. Predicted production over time from each area

An example of the model output can be observed in Figure 4. This chart shows expected output overtime for each of the areas across assembly and test aggregated across 6 replications, showing how production would increase over time and when forecasted totals could expect to be reached within each area. Utilising the simulation in this way was more powerful than trying to calculate throughput

mathematically as it enabled the variation in the system to be more clearly captured. By running scenarios for multiple replications, the number of parts produced could be predicted within confidence intervals so that plans could be made to cover different scenarios within the range (Figure 5).

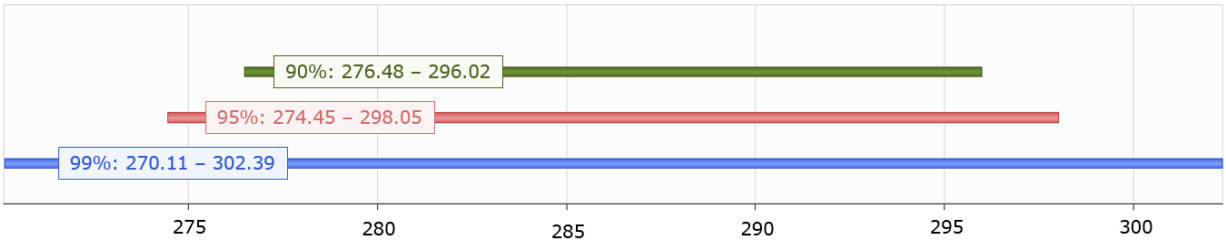


Figure 5. Estimated parts produced, with confidence intervals within a specified time period

3.3 Production Phase

As production began to ramp up and the output increased in line with predictions, it became clear that some early model assumptions had not been realised. Although cycle times had improved in line with expectations, external factors (beyond the control of Ford) had resulted in part quality issues in excess of expectation, and a new, more comprehensive regression and repair strategy was required. Again, DES was used to support this development.

Data collection with respect to ventilator repairs provided valuable insight into how units were managed whilst away from the mainline. As this pool of data grew, a fuller understanding of the offline regression process was developed. In conjunction with the test station operator repair guides, which detailed the process, duration, and outcome scenarios of tests conducted in-station, it became possible to develop probability routing distributions for every test station on the line.

When a unit entered a test station it was now possible to determine the probability of whether it would pass all tests, require re-testing, require regression or repair (and where that would be conducted), what the next step in its journey through the line would be (pass to the next station, return to a previous operation, redirect to another area, or take away for offline regression), as well as the time it would take to undertake these steps.

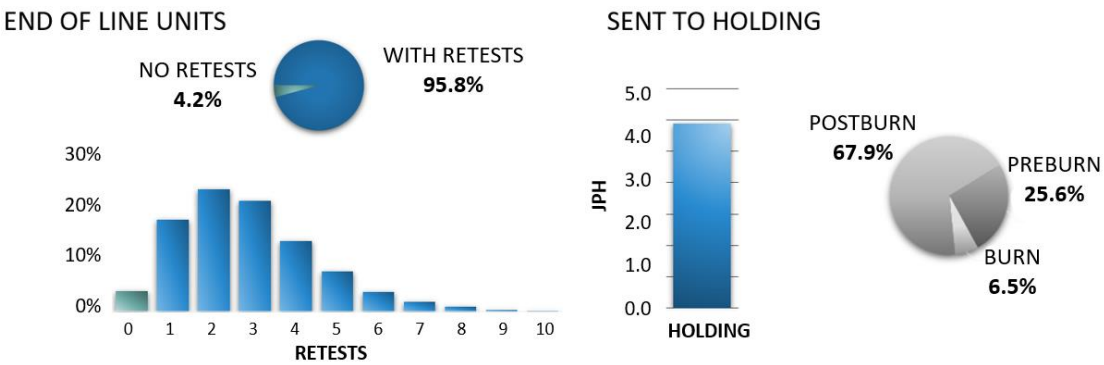


Figure 6. Example screenshot from simulation results dashboard detailing test/repair data

By enhancing the model with these routing probability distributions (which, in line with the continued ethos of rapid response could be modified as soon as new data was available), the impact to production of different strategies could be predicted with confidence. Various solution scenarios put forward by stakeholders across the project could be modelled to see which would result in the biggest improvements to end of line throughput, and the effects viewed in real time during a simulation run from the results dashboard of the Excel 'Front End' (see Figure 6).

3.4 Decommissioning Phase

As the end of the project neared, focus shifted away from production optimisation and onto how the facility could be decommissioned. With the simulation model able to derive dock-to-dock times (also known as total cycle times, the cumulative time taken to complete a process, including wait time and inventory time) at each stage, it could be determined when units/parts had to commence specific processes in order to ensure they were completed in time.

Figure 7 shows a feature of the model whereby a decommission date and date could be entered, and the corresponding date and time that the last parts ought to be loaded to the line in that area, in order to complete production before decommission. This was useful information for the production team as it could support scheduling of parts, allocation of labour and detail exactly when different areas of the line could be switched off.

	DATE	TIME
DECOMMISSION:	Thu 2 July	09:00

AREA	LAST UNIT IN	LAST UNIT OUT
PATIENT VALVE	Wed 1 Jul, 17:01	Wed 1 Jul, 17:11
REGULATOR CONTROLLER	Wed 1 Jul, 11:18	Wed 1 Jul, 13:42
FRONT PANEL	Wed 1 Jul, 12:38	Wed 1 Jul, 13:10
TRAY	Wed 1 Jul, 12:45	Wed 1 Jul, 13:07
CHASSIS	Wed 1 Jul, 15:11	Wed 1 Jul, 17:16
PRE-BURN	Wed 1 Jul, 17:17	Wed 1 Jul, 19:20
BURN	Wed 1 Jul, 19:20	Thu 2 Jul, 05:23
POST-BURN	Thu 2 Jul, 05:23	Thu 2 Jul, 07:46

Figure 7. Example screenshot from simulation results dashboard detailing decommissioning dates/times

4 CONCLUSION

Throughout all stages of the project, from conceptualisation to decommissioning, the Ford EOS 2 ventilator facility was supported by Discrete Event Simulation. The developed DES models were able to accurately represent the interference and combination of events that could occur; especially in comparison to static mathematical models. Having a simulation model available through all phases of the project enabled a comprehensive understanding of both current status and future performance, and allowed productivity to be maximised through data-driven decision making.

Notwithstanding, the project did highlight areas where improvements could be made. Due to the nature of the project, the constant rapid evolution of the facility often outpaced the ability to capture and model these changes – despite the inherent flexibility designed into the model to address this need. With non-stop production, and wide changes being implemented on decisions taken by those ‘on the ground’ on an hour-to-hour basis, at times it was simply not possible to effect the level of responsiveness required from the simulation.

A key aspect of this was data capture. The speed at which the facility was installed did not provision for an adequate data infrastructure to be installed. At the outset, there was no data available meaning substantive assumptions had to be made, and despite the engineering expertise of those making the assumptions, these were not always correct. As data later became available once production increased, it was not always reliable nor accessible in a suitable timeframe. A key takeaway from this project has been the criticality of viable data, available in a timely fashion, in order to support a simulation model’s development.

Those issues aside, the project was still considered a success for the PTME Simulation team. It had been demonstrated that a valid DES model could be swiftly built and used to support a low volume, ‘pop-up’ manufacturing facility. Going forward it gives confidence that for similar projects the team

has the experience and insight to specify the requirements of such a facility at a conceptual stage, and maintain a strategic position to enable the execution of data-driven engineering decisions.

REFERENCES

Department for Business, Energy & Industrial Strategy. (2020). *Call for businesses to help make NHS ventilators*. https://www.gov.uk/government/news/production-and-supply-of-ventilators-and-ventilator-components, accessed February 2021.

Ferguson NM, et al. (2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. Imperial College COVID-19 Response Team. https://doi.org/10.25561/77482

Higgins, M (2013). *Fitness of simulation: advancing the application of discrete event simulation within the automotive industry*. MPhil Thesis, Cardiff University. http://orca.cf.ac.uk/61450/

Ladbrook, J. and A. Januszczak (2001). *Fords Power Train Operations – Changing the Simulation Environment*. The Proceedings of the 2001 Winter Simulation Conference Washington, pp 863-869

Law, A. M. and Kelton, W. (2007) *Simulation Modelling and Analysis*. Boston: McGraw-Hill

Robinson, S. (2007). *Simulation the Practice of Model Development and Use*. 2007, UK: John Wiley and Sons.

AUTHOR BIOGRAPHIES

MATTHEW WINDELER graduated with a Master of Science in Materials Science and Engineering in 2014. He has worked with Ford Motor Company since 2014, initially gaining experience within Powertrain Manufacturing Engineering (PTME) as a tooling engineer within Process Engineering, before taking on his current role as Senior Simulation Engineer in 2018. He is the lead engineer responsible for delivery of simulation models to support manufacturing process feasibility and commissioning, as well as engaging with academic partners to further develop application of the Simulation toolsets within Ford. His email address is <mwindele@ford.com>.

MICHAEL HIGGINS graduated with a Bachelor of Science in Mathematics and Physics in 2011 and in 2014 gained an MPhil(Eng) through his research into the application of Discrete Event Simulation within the automotive industry. He has worked with Ford Motor Company since 2012 where his current role is within Powertrain Manufacturing Engineering (PTME) as Simulation and Process Optimisation Leader. He has gained a holistic view of powertrain manufacturing and has utilised simulation as a tool across several functions within PTME including plant and production engineering roles. He is currently researching and developing methods to increase usability, and widen the application of, the Ford Simulation Toolset. His email address is <mhiggi37@ford.com>

GILES THOMAS B.Eng(Hons) is a Manufacturing Engineer and freelance simulation specialist, with over 20 years' experience in DES within the automotive and aerospace industries. Now working with Ford Motor Company, he is responsible for development of their FASTER rapid model building application, as well as leading various simulation projects across the UK and Europe, and delivering training in the use of discrete event simulation tools and techniques. He is currently studying for an MSc in Computer Science with Data Analytics. His email address is <gthom242@ford.com>

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.026

SIMULATING THE SPREAD OF COVID-19: A CASE STUDY ON WUHAN

Md Zamanur Rahman
w*1757464@my.westminster.ac.uk*

Christopher Hitchin
w1706795@my.westminster.ac.uk

Kathryn Liddle
w1712301@my.westminster.ac.uk

Sifi Bahuleyan
w1746454@my.westminster.ac.uk

Olawale Betiku
w1623075@my.westminster.ac.uk

Thierry Chaussalet
chausst@westminster.ac.uk

School of Computer Science and Engineering
University of Westminster
115 New Cavendish St, London, W1W 6UW

ABSTRACT

We have a long history of pandemics. But the severity and fatality of COVID-19 concern the entire world and attracted the R&D, media, and billions of people worldwide. Under this purview, the study aims to simulate the peculiar spread of the COVID-19 virus, which allows us to see the impact on Wuhan's deaths and recoveries. The study considered the basic 'SIR' model using five scenarios: with 'all measures, social distancing, isolation, isolation and social distancing, and no measures' situations to understand the pattern of the spread of COVID-19. It found a significant impact on these different interventions. The death rate is seen as the highest with no support scenario and has reached the lowest with the isolation and social distancing scenario. Moreover, it also showed that any action taken significantly affects the number of people infected, the number of recoveries, and the number of deaths.

Keywords: Simulation modelling, SIR, COVID-19, Social Distancing, Isolation

1 BACKGROUND

On 31 December 2019, a cluster of pneumonia cases of unknown aetiology was reported in Wuhan, Hubei Province, China. On 9 January 2020, the China CDC (Centre for Disease Control and Prevention) reported a novel coronavirus as the causative agent of this outbreak, coronavirus disease 2019 (COVID-19), and 44 cases were reported to WHO by the national authority of China (ECDC, 2020). Later, it was spread to different territories of Wuhan, including Hubei as well as Thailand, Korea, and Japan, as reported by the WHO in its first situation report. After that, WHO updated Clinical Management Guidance for COVID-19, released their preparedness and responses when spread globally in their 52nd situation report. WHO declared a global health emergency on March 12 (WHO, 2020). It has now spread to 187 countries and over 3.84 million people with a fatality volume of over 0.269 million (Johns Hopkins CSSE, 2020). The severity and fatality of concern expressed by the R & D of the whole world have attracted a great deal of media attention, with billions of people going into lockdown across the globe (Buchholz, 2020)

2 LITERATURE REVIEW

We have a long history of pandemics, however we did not know when the pandemic would occur or how severe it would be. It could cause an outbreak resulting in millions of fatalities. Under the circumstances of the absence of reliable Covid-19 pandemic exposure systems, computer models have become important information tools for all concerns. They can help provide global insight into the outbreak's behaviour and spread of infectious diseases in a given population, with varied geographic and demographic features. There are different studies on epidemiological modelling approaches. Here is a summary of the literature review to find the proper approaches and parameters.

The study of mathematical models for the spread of infectious diseases is an important issue in epidemiology. Given the real world, a theoretical model of the spread of infectious diseases is proposed (Zhong et al, 2009; Lynch, 2020). The classic SIR model (Susceptible-Infectious-Recovered) is based on ordinary differential equations developed by Kermack and McKendrick in 1927. It was successful in predicting the behaviour of some epidemics (Hethcote, 2000). Another study (Daley, 2008) found that the spread of infectious diseases crucially depends on the pattern of contacts between individuals, whereas Stehlé et al (2011) found some limitations regarding person-to-person contacts, which is also found in Chen et al (2020). Another study (Vynnycky and White, 2010) found some critical features like basic and net reproduction numbers and the herd immunity threshold of infections, and the reasons for epidemics. The study by Cecconi and Barazzetti (2020) found that work, social relations, and leisure have an impact on the spread. They also found social distancing to slow down the rapid spread of COVID -19 in Italy.

Mac Hyman of Tulane University categorically mentioned in his Mathematical Modelling of COVID-19 the relation and importance of the number of tests with the spread of this pandemic (Mac Hyman, 2020).

A System Dynamics (SD) approach can help us understand the rapid spread of an infectious disease such as COVID-19 and generate scenarios to test the effect of different control measures (Bordehore et al., 2020). However, within a given population, diseases can spread at different rates over time due to the natural random nature of contact between individuals in the population. But here, SD has the limitations of no variation in output with a fixed rate of contact (Forrester, 1961). Subsequently, the simulations are repeated with different input parameters by applying a Monte Carlo simulation with no variation (Stan, 1987).

On the other hand, Khalil et al (2012) found that the variables used for using Agent Based Models (ABM) are social agent attributes, distribution of population, and patterns of agent interactions whereas Hack (2019) found that human mobility is a key element in studying the large-scale spatial transmission of infectious diseases and improving epidemic control. In the case of super-spreading MERS-CoV, simulations of the epidemic show proportionality to the super-spreading effect (Hossain et al, 2017). The rapid growth in computer power has enabled ABM to consist of autonomous "agents" that interact with each other and have varying characteristics (Lynch, 2020). There are situations for which ABM can offer distinct advantages to conventional simulation approaches (Macal and North, 2006). It is demonstrated that the dynamic spatial interactions within the population lead to high numbers of exposed individuals (Perez and Dragicevic, 2009) whereas Chen et al (2020) mentioned that simulation can be used to predict the spread of the disease.

3 PROJECT GOAL AND SYSTEM DESCRIPTION

The primary purpose of this project is to simulate the strange territory spread of the COVID-19 virus, deaths, and recoveries. This will allow us to see the potential impact of Wuhan and understand how control measures affect the virus' spread.

3.1 Flow Diagram

The classic SIR model (Hethcote, 2000) assumes that individuals transfer between categories with a certain probability where β is controlling how much the disease can be transmitted through exposure, determined by the chance of contact and probability of disease transmission γ = how much the disease can be recovered in a specific period. Our proposed model is an extension of the classic SIR model Our population is divided into six categories: Susceptible, Infected, Isolated, Not Isolated, Recovered, and Death (Figure 1) where each was changing over time with a given probability to make the model more realistic.

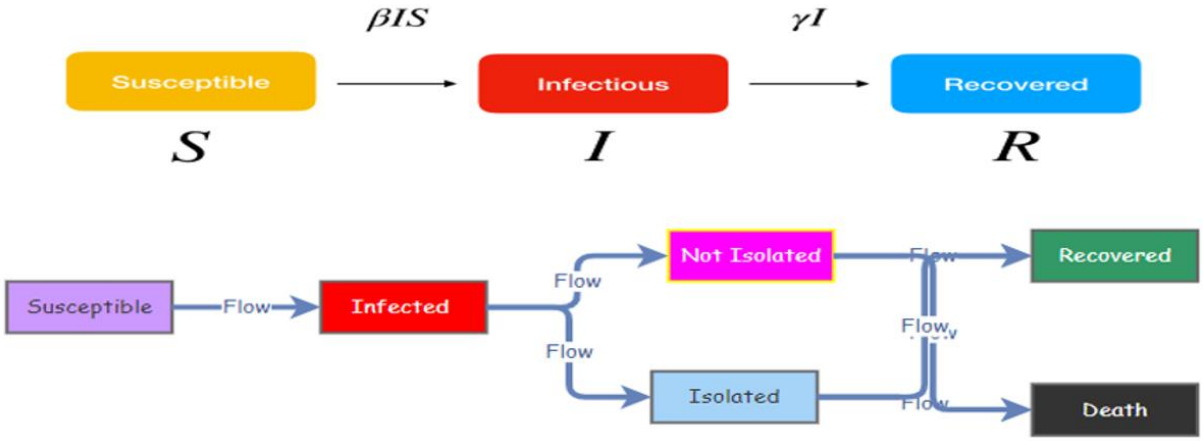


Figure 1 *Classic SIR and Proposed SINIRD Model Flow Diagram*

3.2 Assumptions

- The population is closed in that no additions are made to the susceptible population as births and immigration are ignored. The only way an individual leaves the susceptible group is by becoming infected.
- Once someone enters the recovery population, immunity is assumed so they cannot re-enter the susceptible population.

3.3 Key Variables

Since the model is based on the SIR model, the parameters are also based on three phases. The parameters are self-explained in the diagram. In the first phase, the parameter is with the parameter susceptible, which is the assumed population of Hubei. In the next phase, we included the infected stages are split down into three stages where a healthy person is infected, without and with symptoms for keeping them in Isolation. Recovery and Death are the 3rd phases for both isolated and not isolated states. There are 17 variables among which is Healthy, Infected, Isolated, Not Isolated, Recovered and Dead are the 6 states of the population: Social Distancing, Transmission rate, Vaccine, Total Ill, % Isolation, Incubation, Recovery Rate, Illness Duration (for recovery), Illness Duration for (for dead) Hospital Beds, Death Rate are the 11 key agents. There is a specific formula behind the variables. Despite being suggested by a literature review (Mac Hyman, 2020) that the number of the test has a strong relationship with the spread of COVID-19 but due to lack of data, these parameters are not considered due to some data constraints.

3.4 Various Types of Simulation in Epidemiological Modelling

In a search to understand the behaviour of infectious diseases spread model and predict the pattern of diseases through a population, several attempts were initiated.

The earliest accounts were carried out in 1927 by Kermack and McKendrick (Hethcote, 2000). Following the SIR model, other physicians tried SEIR (Susceptible–Exposed–Infectious–Recovered) and ISEIR (Immunized–Susceptible–Exposed–Infectious–Recovered) models (Hethcote, 2000). However, mathematical models did not consider factors such as variable population structure and dynamics of daily individual interactions, which drove more realistic modelling results (Bonabeau, 2002). To overcome these limitations, ABM come up with extra leverage tracking the effect of social interactions on individual entities.

Several studies found the advantages of using ABM, which consists of a population of agents, an environment, and a set of rules managing agents' behaviour (Perez and Dragicevic, 2009). Each agent has two components: a state and a step function. The agent state describes every agent's attribute values in the current state. The step function creates a new state (usually stochastically) representing the agent attributes at the next step. The great benefit of agent-based models is that these models allow epidemiological researchers to do a preliminary "what-if" analysis to assess systems' behaviour under

various conditions and evaluate. This is an alternative control strategy to adopt to fight epidemics (Perez and Dragicevic, 2009).

ABM also helps answer the issue of validation (Oberkampf and Trucano, 2002). Unlike System Dynamics (SD), which uses a top-down approach to model the system in ABM simulations, the system is "brought about" by carrying out lower level interactions between the agents. For this reason, ABM is beginning to be used in a range of fields, including biological simulations and social sciences, representing people as interacting agents in environments. ABM simulations can produce different output results for each run based on knowledge of the local interactions of the underlying agents and without making any changes to the input parameters. A study by Ahmed et al (2013) shows the influence and effect of variation within these two distinct simulation paradigms and shows that the ABM simulation of the epidemiological SIR model is more effective at capturing the natural variation within SIR compared to an equivalent model using SD with Monte-Carlo simulation.

4 KEY DATA

The 2019 Coronavirus (COVID-19) has turned into a global pandemic with unprecedented challenges for the worldwide community. Understanding the state of the disease and planning for future trajectories relies heavily on data on spread and mortality. But the unfortunate thing is the official data coming from various countries are highly unreliable (Stevens, 2020; Ghaffarzadegan and Rahmandad, 2020).

4.1 Data Source

The primary data source used was the number of cases, recoveries, and deaths published daily by the World Health Organisation and Johns Hopkins University and broken down by country and state (CSSEGIS and Data, 2020). Our study focused on Wuhan, where the outbreak started. As of 30 March, 10.4% of cases were in China, and 83.2% of these were in Hubei. Other information, such as the incubation period, hospital beds, and when lockdown began, were also used to inform parameter values (ECDC, 2020; CSSEGIS and Data, 2020; Wu Pei Lin and Lin, 2020, Reuters, 2020). We model the outbreak in Wuhan with individual reaction and governmental action (holiday extension, city lockdown, hospitalisation, and quarantine) based on some parameters of the 1918 influenza pandemic in London, United Kingdom (Lin et al., 2020).

4.2 List of major assumptions

- All humans are susceptible to 7.8 billion (as of February 2020) (Chamie, 2020).
- No one is immune to the disease as it is a zoonotic virus (it originates from another, yet unknown, animal).
- Those who recover are immune to the disease (at least if there is an outbreak). Seasonal human coronavirus produces immunity to these viruses which last longer than that of seasonal influenza but assume it is not permanent (Bai, 2020)

4.3 Fitted Distribution and p-value

As we know, an ABM is a computer programme that implements a Complex Adaptive System (CAS) by simulating its behaviour. The CAS describes the probability distribution of outcomes for each vector of inputs x and equation parameters p, and the ABM simulates the probability distribution (Blume, 2015). In our model, % of case change has exponential Distribution, whereas the death rate has got the beta distribution.

5 PROPOSED MODEL

The proposed ABM model involves 6 population states and 11 agents' rules which govern the behaviour of the agents (see Figure 2). Agents represent the human population, in which each agent is involved in a sequence of daily basis activities according to the agent's social environment.

5.1 Correct usages of model

ABM is widely used in population health, especially with modelling the spread of infectious disease. A bottom-up approach helps us in providing global insight into the spread of contagious disease outbreaks' behaviour in a given population and transmission dynamics irrespective of geographic, demographic features, and social structure (Perez and Dragicevic, 2009).

5.2 Limitations

- There is a chance of fake news and unreliable data we use – much of this has been circulated on social media (Purohit, 2020; Ghaffarzadegan and Rahmandad, 2020).
- There is also a proportion of unreported cases as people may not feel the worst symptoms (Neergaard, 2020).
- Unable to validate or compare the results with ABM using insight maker.
- Frequent changes in data are unstable due to the condition of the agent's behaviour and time series
- The data is inaccurate; people might not declare their symptoms to avoid being isolated, quarantined, or hospitalised.
- ABMs can be more challenging to analyse, understand, and communicate than traditional analytical/ mathematical models (Wainwright and Mulligan, 2013).
- Unable to consider the social activity level, daily movement, spatial location, infection time, social type, and agent social networks (Khalil et al, 2012).

5.3 Model Layout

As mentioned earlier in the flow diagram, our model contains six states an agent can be: susceptible, infected, not isolated, isolated, dead, and recovered. Specific parameters affect the probabilities of agents moving between states. The parameters were changed to investigate how they impacted the spread of COVID-19, deaths, and recoveries. Figure 2 shows a conceptual ABM model diagram of the interactions of the different variables and parameters considered.

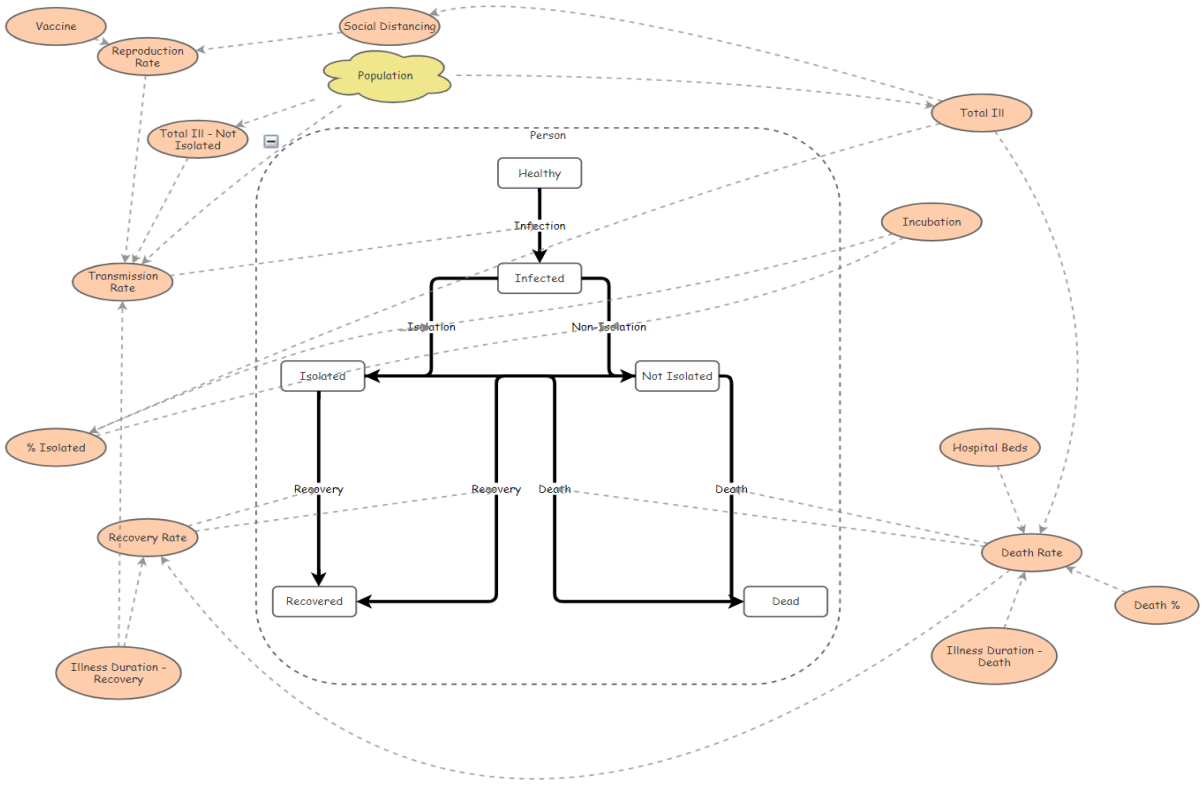


Figure 2 *Conceptual Model Diagram*
(Available at https://insightmaker.com/insight/211859/DZR-Model)

5.4 Scenarios considered

In the study, five scenarios were considered where the control measures were social distancing and isolation and the medical resources were the number of beds to understand the pattern of the spread of COVID-19 as mentioned in Table 1.

Table 1 *Intervention Scenarios*

Scenarios	Description
A	All measures
B	Social distancing
C	Case isolation
D	Case isolation and social distancing
E	No measures

We ran the model with a close population of 500 agents for 740 days and initially 1 infected agent. The behaviour over time of each variable is shown in Figure 3. Of particular interest is the decrease in the healthy population, the increase in the number of recovered and dead, while these variables eventually stabilise after about 14 months (420 days).

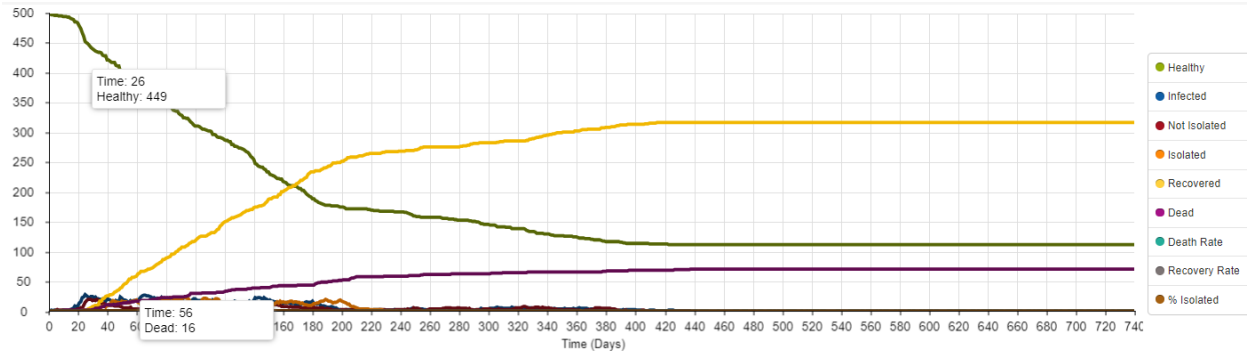


Figure 3 *Illustration of the Demo Model*

5.5 Results

This was done under five scenarios of pandemic COVID-19 where we considered different populations and different lengths of the pandemic with various interventions: no support, all, social distancing only, Isolation only, social distancing, and isolation only. The number of deaths (556) is the highest with no support scenario, lowers with isolation and social distancing scenarios, and is the lowest with all measures (243). Results are shown in Table 2.

Table 2 *Results for all scenarios*

Scenario A (with all measures)	Scenario B (social distancing)	Scenario C (case isolation)	Scenario D (case isolation & Soc Dist)	Scenario E (no measure)
Highest number infected: 26 Highest number isolated: 19 Recovered: 839 Death: 243 Length of the pandemic: 752 days	Highest number infected: 24 Recovered: 1061 Death: 317 Length of the pandemic: 998 days	Highest number infected: 56 Highest number isolated: 56 Recovered: 791 Death: 245 Length of the pandemic: 537 days	Highest number infected: 28 Highest number isolated: 18 Recovered: 977 Death: 257 Length of the pandemic: 1024 days	Highest number infected: 150 Recovered: 1344 Death: 556 Length of the pandemic: 280 days

5.6 Validation

In our project, given that historical data is available for the epidemic spread in the Hubei region, we used data validity to validate our model.

Model validation was carried out in R: we compared the predicted number of deaths and recoveries (scenario A) in the Hubei region with the actual case data available. We found a strong correlation between the model and the Hubei data for both the number of deaths (Corr = 0.959) and the recoveries (Corr = 0.958). We further tested the model by performing a t-test. For both recovered individuals and deaths, the p-value was <0.00001, indicating a highly significant result. These results overall confirm that our model is a valid representation of the pandemic spread in the Hubei region.

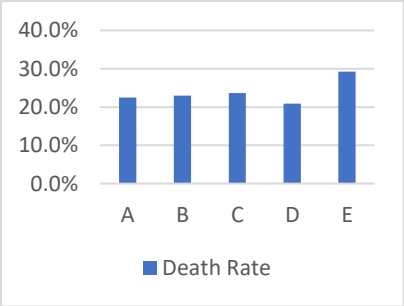
6 OUTPUT ANALYSIS BY COMPARING SCENARIOS AND HYPOTHESIS TESTING

It may be mentioned that different interventions have a significant impact on the infected, recovered, isolated, and not isolated and death scenarios which is the beauty of our model. We see that the death rate is the highest as expected with no support scenario and the lowest with the isolation and social distancing scenario (Table 3 and Figure 4).

Table 3 *Output Analysis with different scenarios*

SN	Intervention Support Name	Mean Infected	Mean Isolated	Var. Infected	Recovd.	Dead	Total Cases	Death Rate
A	Social Distancing + Case Isolation + Extra Hospital Beds	8	5	40.213	839	243	1082	22.5%
B	Social Distancing	8	0	39.164	1061	317	1378	23.0%
C	Case Isolation	8	7	183.19	791	245	1036	23.6%
D	Case Isolation + Social Distancing	10	6	38.733	977	257	1234	20.9%
E	None	11	0	754.47	1344	556	1900	29.3%

Figure 4 *Scenarios Death Rate*



Hypothesis testing showed that any action taken has a significant effect on the number of people infected, the number of recoveries, and the number of deaths. This was because testing all scenarios against the base scenario (no effects) rejected the null hypothesis that these were equal to the volumes for scenario E. However, it should be noted that there is some volatility in the results. Some scenarios that we would expect to give relatively better performance (A vs. D) did not provide these results.

A potential next step would be to run more simulations so that there would be less volatility. However, it does appear that the acts of case isolation and social distancing together do affect, suggesting that government policies as they are should work (Strochlic and Champine, 2020).

6.1 Sensitivity Analysis of Death and Infection Rate

Further simulations were run based on Scenario A to examine the sensitivity of the death rate (Table 4). This was done by reducing the death rate in this scenario by 10%. Instead of changing the death rate, the infection rate was reduced by 10% (Table 5). This implies that reducing the infection rate is a key measure to avoid the spread of the virus, with measures such as social distancing. This gives a sensitivity of 0.91 for the death rate, showing that parameters that change the death rate do come through in the actuals.

Table 4 *Sensitivity Analysis of the Death Rate*

Scenario	Death Rate
All measures	22.5%
All measures, death rate reduced by 10%	20.4%

Table 5 *Sensitivity Analysis of the Infection Rate*

New infection rate	0.374
Previous infection rate	0.4328
Sensitivity	0.86414

7 CONCLUSION

The field of computational epidemiology has arisen as a new branch of epidemiology to understand epidemic transmission patterns and to help in planning precautionary measures. The proposed model

simulates the effect of pandemic COVID-19 outbreaks in Hubei, China. The model can be easily customised to study the pandemic spread of any other infectious disease by merely adjusting the model parameters. Deployment of a proper combination of control strategies can limit pandemic chaos and reduce fatalities and substantial economic damage. Further work on the proposed model includes other parameters like the number of tests per population, impact with open-air treatment (Vitamin D impact), the optimum number of medical staff and resources, and age-gender-medical history to decode the pandemic outbreak waves.

REFERENCES

Ahmed, A., Greensmith, J., and Aickelin, U. (2013). Variance in System Dynamics and Agent-Based Modelling Using the SIR Model of Infectious Disease. *ResearchGate*. Available from https://www.researchgate.net/publication/247153200_Variance_in_System_Dynamics_and_Agent_Based_Modelling_Using_the_SIR_Model_of_Infectious_Disease [Accessed 4 May 2020].

Bai, N. (2020). How the New Coronavirus Spreads and Progresses – And Why One Test May Not Be Enough. *How the New Coronavirus Spreads and Progresses – And Why One Test May Not Be Enough | UC San Francisco*. Available from https://www.ucsf.edu/news/2020/02/416671/how-new-coronavirus-spreads-and-progresses-and-why-one-test-may-not-be-enough [Accessed 22 April 2020].

Blume, L. (2015). *Agent-Based Models for Policy Analysis*. National Academies Press (US). Available from https://www.ncbi.nlm.nih.gov/books/NBK305903/ [Accessed 4 May 2020].

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99 (suppl 3), 7280–7287. Available from https://doi.org/10.1073/pnas.082080899 [Accessed 5 May 2020].

Bordehore et al. (2020). Understanding COVID-19 spreading through simulation modelling and scenario comparison: preliminary results. | medRxiv. Available from https://www.medrxiv.org/content/10.1101/2020.03.30.20047043v1.full.pdf+html [Accessed 21 April 2020].

Buchholz, K. (2020). Infographic: What Share of the World Population Is Already on COVID-19 Lockdown? *Statista Infographics*. Available from https://www.statista.com/chart/21240/enforced-covid-19-lockdowns-by-people-affected-per-country/ [Accessed 22 April 2020].

Cecconi, F. and Barazzetti, A. (2020). Agent-based simulation models are applied to social behaviours determining the dynamics of pandemics. *UCM*. Available from https://ucm.edu.mt/en/agent-based-simulation-model-applied-to-social-behaviors-determining-the-dynamics-of-pandemics/ [Accessed 22 April 2020].

Chamie, J. (2020). World Populations: 2020 Overview | YaleGlobal Online. Available from https://yaleglobal.yale.edu/content/world-population-2020-overview [Accessed 22 April 2020].

Chen, B. et al. (2020). Visual Data Analysis and Simulation Prediction for COVID-19. 19.

CSSEGIS and Data. (2020). *CSSEGIS and Data/COVID-19*. Available from https://github.com/CSSEGISandData/COVID-19 [Accessed 22 April 2020].

Daley, D.J. (2008). *Epidemic Modelling: An Introduction*, Revised ed. edition. 1st paperback edition, reprinted 2005: Cambridge University Press.

ECDC. (2020). Disease background of COVID-19. *European Centre for Disease Prevention and Control*. Available from https://www.ecdc.europa.eu/en/2019-ncov-background-disease [Accessed 22 April 2020].

Forrester, J.W. (1961). *Industrial dynamics.* Cambridge, Mass.: M.I.T. Press.

Ghaffarzadegan, N. and Rahmandad, H. (2020). Simulation-based Estimation of the Spread of COVID-19 in Iran. *medRxiv*, 2020.03.22.20040956. Available from https://doi.org/10.1101/2020.03.22.20040956 [Accessed 21 April 2020].

Hack, J. et al. (2019). Epidemic Spreading in Urban Areas Using Agent-Based Transportation Models. *ResearchGate*. Available from https://www.researchgate.net/publication/332284749_Epidemic_Spreading_in_Urban_Areas_Using_Agent-Based_Transportation_Models [Accessed 7 March 2020].

Hethcote, H.W. (2000). Mathematics of Infectious Diseases. *SIAM Review*, 42 (4), 599–653. Available from https://doi.org/10.1137/S0036144500371907 [Accessed 9 March 2020].

Hossain et al. (2017). Modelling of a Super-Spreading Event of the Mers-Corona Virus during the Hajj Season using Simulation of the Existing Data. *ResearchGate*. Available from https://www.researchgate.net/publication/320736530_Modeling_of_a_SuperSpreading_Event_of_the_MersCorona_Virus_during_the_Hajj_Season_using_Simulation_of_the_Existing_Data [Accessed 21 April 2020].

Johns Hopkins CSSE. (2020). Coronavirus COVID-19 (2019-nCoV). *https://systems.jhu.edu/*. Available from https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [Accessed 22 April 2020].

Khalil M. et al., K. (2012). Agent-based modelling for pandemic influenza in Egypt. *ResearchGate*. Available from https://www.researchgate.net/publication/224136840_An_agent-based_modeling_for_pandemic_influenza_in_Egypt [Accessed 29 February 2020].

Lin, Q. et al. (2020). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases*, 93, 211–216. Available from https://doi.org/10.1016/j.ijid.2020.02.058 [Accessed 3 April 2020].

Lynch, P. (2020). Coronavirus: Modelling the evolution of a viral outbreak. *The Irish Times*. Available from https://www.irishtimes.com/news/science/coronavirus-modelling-the-evolution-of-a-viral-outbreak-1.4197391 [Accessed 21 April 2020].

Macal, C. and North, M. (2006). Tutorial on Agent-Based Modelling and Simulation PART 2: How to Model with Agents. *Proceedings of the 2006 Winter Simulation Conference*. December 2006. Monterey, CA, USA: IEEE, 73–83. Available from https://doi.org/10.1109/WSC.2006.323040 [Accessed 29 February 2020].

Mathematical Modeling of COVID-19: 4/17/20 Update with Mac Hyman. (2020). Available from https://www.youtube.com/watch?v=hlzXMFyuhwU [Accessed 3 May 2020].

Neergaard, L. (2020). Coronavirus Death Rate Still Uncertain as Mild Cases Go Unreported. *HuffPost*. Available from https://www.huffpost.com/entry/coronavirus-death-rate_n_5e5ff494c5b6f4b39568be17 [Accessed 22 April 2020].

Oberkampf, W.L., and Trucano, T.G. (2002). Verification and validation of computational fluid dynamics. *Progress in Aerospace Sciences*, 38 (3), 209–272. Available from https://doi.org/10.1016/S0376-0421(02)00005-2 [Accessed 4 May 2020].

Perez, L., and Dragicevic, S. (2009). An agent-based approach for modelling dynamics of contagious disease spread. *International Journal of Health Geographics*, 8 (1), 50. Available from https://doi.org/10.1186/1476-072X-8-50 [Accessed 6 March 2020].

Purohit, K. (2020). Misinformation and fake news spark India's coronavirus fears. Available from https://www.aljazeera.com/news/2020/03/misinformation-fake-news-spark-india-coronavirus-fears-200309051731540.html [Accessed 22 April 2020].

Reuters. (2020). Wuhan closes makeshift hospital as new coronavirus cases in China drop sharply. *Reuters*, 2 March. Available from https://uk.reuters.com/article/uk-health-coronavirus-china-toll-id UKKBN20P023 [Accessed 22 April 2020].

Stan, U. (1987). "John Von Neumann and the Monte Carlo method." *Los Alamos Science, 15*, (special issue), 131-136. Available from https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-88-9068 [Accessed 3 May 2020].

Stehlé, J. et al. (2011). Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9 (1), 87. Available from https://doi.org/10.1186/1741-7015-9-87 [Accessed 3 May 2020].

Stevens, H. (2020). Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve" - Washington Post. Available from https://www.washingtonpost.com/graphics/2020/world/corona-simulator/ [Accessed 21 April 2020].

Strochlic, N. and Champine, R.D. (2020). How some cities 'flattened the curve' during the 1918 flu pandemic. *History*. Available from https://www.nationalgeographic.com/history/2020/03/how-cities-flattened-curve-1918-SPANISH-flu-pandemic-coronavirus/ [Accessed 22 April 2020].

Vynnycky, E. and White, R.G. (2010). An introduction to infectious disease modelling. Available from http://anintroductiontoinfectiousdiseasemodelling.com/chapters/chapter-1/ [Accessed 22 April 2020].

Wainwright, J. and Mulligan, M. (2013). Environmental Modelling: Finding Simplicity in Complexity, 2nd Edition | Wiley. *Wiley.com*. Available from https://www.wiley.com/en-gb/Environmental+Modelling%3A+Finding+Simplicity+in+Complexity%2C+2nd+Edition-p-9780470749111 [Accessed 22 April 2020].

WHO. (2020). sitrep-ncov. Available from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf [Accessed 22 April 2020].

Wu Pei Lin and Lin, as told to W.P. (2020). 'The city is suffocating': a diary of the Wuhan coronavirus lockdown. *The Guardian*, 31 January. Available from https://www.theguardian.com/world/2020/jan/31/the-city-is-suffocating-diary-of-the-wuhan-coronavirus-lockdown [Accessed 22 April 2020].

Zhong, S., Huang, Q. and Song, D. (2009). Simulation of the spread of infectious diseases in a geographical environment. *Science in China Series D: Earth Sciences*, 52 (4), 550–561. Available from https://doi.org/10.1007/s11430-009-0044-9 [Accessed 3 May 2020].

AUTHOR BIOGRAPHIES

MD ZAMANUR RAHMAN recently received an MSc in Business Intelligence and Analytics (BI&A with distinction from the University of Westminster, UK. He also holds an MBA, MCom, MPhil, and PhD (IBA, DU). Before that, he worked in different private and public universities in Bangladesh including IBA, the University of Dhaka as a professor in the area of Accounting and Research as well as in Civil Service of the Government of Bangladesh (GoB) and different foreign funded (CEC, UNDP, DFID) projects.

CHRISTOPHER HITCHIN received a BSc (Hons) in Mathematics from the University of Warwick in 2014. He is currently in the process of completing an MSc in BI&A at the University of Westminster. He is currently a Pricing & Underwriting Analyst at Direct Line Group, one of the UK's leading general insurance companies, with previous experience as a Systems Analyst for an IT platforms provider specialising in the life insurance industry.

KATHRYN LIDDLE received a BSc in Ecology and Wildlife Conservation from Bournemouth University in 2013 and an MSc in BI&A (Distinction) from the University of Westminster in 2020. She has worked as an analyst in various organisations including Tesco, Skanska, and Biztory.

SIFI BAHULEYAN received a BSc in Psychology from Brunel University in 2013 and is currently in the process of completing an MSc in BI&A. She has been working for several years at NHS organisations, currenty as quality improvement coach at Camden and Islington NHS Foundation Trust.

OLAWALE BETIKU received a BSc in business studies with first-class honours at the university of West London, 2016. He is in the process of completing an MSc in BI&A at the University of Westminster and is currently a successful entrepreneur, forex trading, stocks, and cryptocurrency.

THIERRY CHAUSSALET received a PhD in Probability and Stochastic Processes from North Carolina State University (USA) in 1993. He founded the Health and Social Care Modelling Group at the University of Westminster in 1998, became Professor in 2007, and has been course leader of the MSc BI&A since 2010. His research interests are around intelligent data-driven methods to model management and decision making processes. He serves on the Editorial Board of various healthcare modelling and informatics journals. He has been a member of the EPSRC Peer Review College, an expert evaluator for the EU FP7 ICT programme, and is a member of the NIHR Peer Review panel.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.027

USING TRANSIENT SIMULATIONS TO IMPROVE FIELD SERVICE SYSTEMS FOR SEMICONDUCTOR MANUFACTURING

Dr. John W. Fowler

Arizona State University
P.O. Box 874706
Tempe, Arizona 85287 USA

Dr. Esma S. Gel

Arizona State University
P.O. Box 875906
Tempe, Arizona 85287 USA

Dr. Cem Vardar

Arizona State University
P.O. Box 875906
cem.vardar@asu.edu

ABSTRACT

Simulation models of complex real world systems require long computation times for producing reliable estimates of the system performance measures of interest. However, it has been observed in the simulation-based optimization literature that the order and ranking relationship between solutions often become significant very early in the simulation runs, even during the transient state. In this paper we propose a method for using very short transient simulation runs to compare different alternatives which can be quite useful in metaheuristics such as genetic algorithms. We present initial experimentation results for a simple problem system (M/M/1) and a field service system design problem from semiconductor industry.

Keywords: Transient Simulation, Simulation-based Optimization, Heuristic

1 INTRODUCTION

Analysis of large real world systems is a difficult task due to the complexities that arise when there is randomness inherent in the system. However, some degree of randomness is a common and unavoidable characteristic among almost all real-world systems. Simulation modelling has been widely used as a descriptive analysis tool to obtain performance measure estimates of such systems under given system configurations.

Although computer simulation modelling is a highly effective tool for descriptive modelling, analysts often need to search for model decision parameters that either maximize or minimize one or more performance measures of the system. This type of a problem is referred to as a simulation optimization problem. Fu (2001) formally defines a simulation-based optimization problem as finding a configuration or a design that minimizes the objective function:

$$\min_{\theta \in \Theta} J(\theta) = E[L(\theta, \omega)]$$

where θ represents the (vector of) decision variables, J(θ) is the objective function, ω represents a sample path (simulation replication), and L(θ,ω) is the sample performance measure estimate.

The difficulty of solving this problem is that J(θ) is an implicit function of the decision variables (i.e., θ) and an observation of J(θ) can only be obtained through an execution of the simulation model. Often, this requires long computation times, particularly when initial transient bias is significant. Particularly in the semiconductor industry, simulation models used for various purposes, such as

solving supply chain and wafer fab operation problems, are generally quite large and require long computation times to obtain reliable estimates of the performance measures of interest.

Since simulation modelling is traditionally a descriptive analysis tool, in practice, simulation execution parameters such as the level of detail in the model, run length, number of replications, and warm up period are usually set to achieve a sufficiently accurate and precise evaluation of the given system configuration. Most simulation optimization techniques proposed in the literature accept these parameter settings as given and build upon this assumption. That is, they assume that an execution of a simulation model $L(\theta,\omega)$ provides an unbiased estimate for $J(\theta)$. Observe that this assumption is inherent in the formal definition of the simulation optimization problem. However, getting an unbiased estimate of $J(\theta)$ typically requires long computation times especially when the variability inherent in the system is significant, since the accuracy of the confidence interval around $J(\theta)$ cannot improve faster than $1/\sqrt{N}$. In addition the existence of a transient bias for most real world system adds significant computation burden for estimating the steady state performance.

In this paper, we investigate the use of short transient simulation runs to compare solution alternatives in the presence of design-dependent bias and estimation noise. We propose a method that improves the efficiency of using short transient simulations to estimate the ranking of the solution alternatives. Our method uses data collected during the simulation run at different time intervals to predict design-dependent estimation error and ranking of solutions. We demonstrate our method using experimental results on an M/M/1 queueing system and a problem from semiconductor manufacturing that deals with designing a maintenance service system for wafer fabrication facilities.

2 RELEVANT LITERATURE

Our approach is motivated by the seminal paper, Ho et al. (1992), which proposes a concept called 'ordinal optimization'. Ordinal optimization finds a good, better or best solution, instead of trying to accurately estimate the performance of the systems. Ordinal optimization can be complementary to the current simulation optimization techniques like the one proposed in Boesel et al. (2003) by reducing the massive search space into a manageable size before applying sophisticated performance evaluation techniques.

Ho et al. (1992) demonstrate that the order relation between systems often becomes significant very early in the simulation runs and without making many replications. In their experiments, they show that promising solutions could be differentiated from inferior solutions with very little simulation computing time.

The research that followed Ho et al. (1992) can be generally classified in two threads: applications of ordinal optimization ideas and theory of ordinal optimization ideas. Ganz and Wang (1994), Ho and Larson (1995), Wieseltier et al. (1995), and Yang et al. (1997) apply ordinal optimization ideas to various real world problems. On the theory side, Dai (1996) and Xie (1997) prove that the convergence rate of order of two systems can be exponential as the simulation effort expanded increases. Lau and Ho (1997) formalize the idea of ordinal optimization with the definition of alignment probability, which represents the probability of having a specified number of "good enough" solutions in a selected subset of the design space. The authors define the concept and tabulate alignment probabilities for the horse race selection rule for different forms of order performance curves. The basic assumption is that the noise in the observations does not depend on the alternative.

Yang et al. (1997) investigates various options when this important assumption is violated. They propose a method for estimating the design-dependent noise by a linear regression approach. They use this method in two real world examples and demonstrate improvements over traditional ordinal optimization techniques. Yang and Lee (2002) relaxes the assumption of design-dependent noise and propose methods of selection when the noise is design-dependent but known.

In the ordinal optimization literature, short simulations are one of the methods used for selecting a set of good enough alternatives using a fixed simulation run length which has to be specified beforehand. To best of our knowledge there has been no study about how to choose and/or adapt the simulation run length for making the comparisons. In this paper we propose a heuristic which uses variable run lengths to compare the alternatives based on the information that can be obtained from the initial observations in the simulation. In addition we consider comparison of only two alternatives with respect to each other rather than a selection of a subset of good enough alternatives from a large

set. This case is particularly important for metaheuristics used in commercial simulation optimization packages where comparison of two solutions has to be made repetitively.

In the traditional simulation literature there is a vast amount of research papers which mainly study how to deal with the initial transient bias for estimating steady state mean. A comprehensive review of the methods for dealing with transient bias can be found in Law (2014) and progress is still being made. The main focus of most of these studies is how to remove and/or minimize the effects of transient stages of a simulation run where the main objective is to estimate steady state performance rather than utilizing the information generated during the transient stages of the simulation. However, in some recent studies the idea of using transient simulation data to draw inferences about steady state behavior is being explored. For example, Voss et al. (2005) utilizes maximum likelihood estimators for the mean of an autoregressive process to construct confidence intervals for steady state mean from transient simulation data. Although the method proposed is asymptotically valid for linear autoregressive series, they do not work well for models where the initial transient bias is significant such as an M/M/1 queueing system loaded over 50% utilization. For realistic systems the order of autoregression has to be also estimated by using the highly variable transient data which results in poor coverage for the confidence intervals constructed. Our paper differs from this study by focusing mainly on the ranking of two different alternatives using transient data rather than estimating their respective steady state performance.

3 PROPOSED METHOD

The idea of using short transient simulation runs to compare two alternatives is based on the observation that ranking relationship between systems usually becomes significant much before the systems reach their respective steady states in a simulation run (Ho et al. 1992). That is, if we only need the ranking between two different alternatives and are not interested in estimating the steady state performance of the systems we can use much shorter simulation runs for making the correct comparison. Estimating the ranking with a short computation time rather than spending the computation budget on estimating the steady state performance could particularly be useful in simulation-based optimization techniques that utilize metaheuristics such as genetic algorithms and simulated annealing.

Based on this observation, a method that could be applied is estimating the ranking between different alternatives based on their observed transient performance. The observed transient ranking of two alternative systems from a single simulation run might not reflect the true steady state ranking due to two reasons. The first one is the inherent simulation variability which might be rather high in short transient simulations due to correlated observations and small sample size. We have some degree of control on this type of error through the length and number of replications we use for comparing the system alternatives. The second reason is true average transient ranking of two systems might not reflect the true steady state ranking relationship. That is, when comparing two alternative systems, the alternative with the worse steady state performance might have a better expected transient performance. The usefulness and efficiency of this method depends on the transient behavior ranking relationship between the two systems to be compared. Below, we define three groups of systems for categorizing this relationship.

1) Dominating ranking relationship: The average transient ranking between two systems reflects the true steady state ranking of the systems at any simulation length. Comparison of two systems with a dominating relationship can be made at any simulation length without making a systematic error. That is at each simulation length the true average transient ranking reflects the true steady state ranking. When comparing these systems with this type of ranking relationship based on a single replication we can make an incorrect selection only due to simulation variability.

For example, consider following two M/M/1 system alternatives

Alternative 1: $\rho = 0.75$ with $\lambda = 0.5$ and $\mu = 0.667$

Alternative 2: $\rho = 0.80$ with $\lambda = 0.6$ and $\mu = 0.75$

Figure 1 shows the average time in system for both alternatives based on the number of customers simulated after starting in an empty and idle state. Observe that, if we want to select the system with lower time in system by simulating two systems starting in an empty and idle state, we can use the

transient observations of time in system with any simulation length because alternative 1 has a lower expected time in system than alternative 2 even in the transient stages starting with in an empty and idle state.

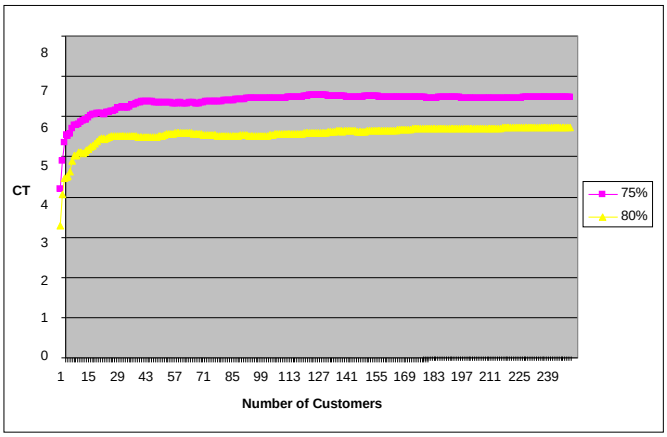


Figure 1 Average time in system for two alternatives with dominating ranking relationship

Problems with alternatives that have a dominating transient ranking relationship are the best candidates for using short transient simulations for estimating steady state ranking.

2) Single crossover ranking relationship: There is a cutoff simulation length point after which the observed ranking between two systems reflects the steady state ranking of the systems and before which the observed ranking is the reverse of the steady state ranking.

For example consider following M/M/1 systems:

Alternative 1: $\rho = 0.65$ with $\lambda = 0.250$ and $\mu = 0.384$

Alternative 2: $\rho = 0.85$ with $\lambda = 0.497$ and $\mu = 0.584$

Figure 2 shows the average time in system for both alternatives based on number of customers simulated after starting in an empty and idle state. If we want to select the system with lower time in system by simulating two systems starting in an empty and idle state and make the decision based a simulation of the first 10 customers starting in an empty and idle state we would be making a systematic error. Although alternative 1 has a larger average time in system in initial phases of the transient state, it has a higher average time in system than alternative 2 in steady state. For these types of systems, only the observations after the crossover point should be used for estimating the steady state ranking.

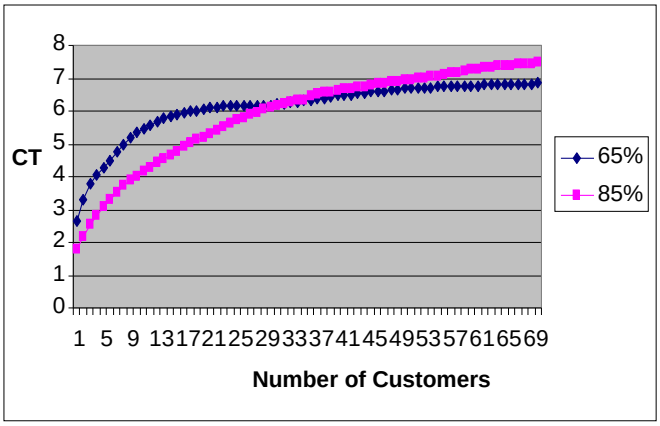


Figure 2 Two alternatives with single crossover ranking relationship

Problems where different solution alternatives that have a single crossover relationship can still be good candidates for applying this method if the crossover point is very early in the transient stage or can be accurately predicted. This type of relationship is often observed when comparing two alternatives for real world problems.

3) Multiple crossover ranking relationship: The ranking relationship in the transient state between alternatives changes multiple times until the systems reach at steady state. These types of systems are not good candidates for using transient simulations. Queueing systems with non-trivial batching policies or systems that include periodic queueing elements potentially belong to this group.

For alternatives that have a dominating ranking relationship any ranking observation made during the transient period can be expected to give a correct inference about the steady state ranking relationship of the alternatives. An incorrect ranking can be only observed due to inherent simulation randomness. In a similar fashion, for alternatives that have a single crossover relationship only the ranking observations made after the crossover point can be expected to give correct inference. However, in metaheuristics such as genetic algorithms, incorrect decision in the exploration (early) phase are not overly harmful.

A difficulty that needs to be considered for using transient simulations for estimating the steady state ranking of two systems is the high variability in the transient solutions due to a low number of observations. Even for alternatives that have a dominating or single crossover ranking relationship the comparison decisions that are made based on initial observations may not be reliable due to the fact that comparisons are made based on a single replication. A method that uses transient simulation runs for comparing alternatives should include a mechanism for controlling high variability in the transient observations. In steady state simulations with independent replications (or independent batch means) confidence intervals are used to control the level of error in the inferences drawn from simulation results. Since observations in short transient simulations are highly correlated, straightforward confidence interval methods cannot be applied exactly. However, in the following sections of this paper we demonstrate that confidence intervals that assume independence of observations could be useful for approximately controlling the error. We demonstrate in Vardar (2006) that using common random numbers in different alternatives further improve the control of error.

3.1 Variable length transient ranking heuristic

In this section, we define the proposed heuristic. The basic idea of our heuristic relies on making short simulations of two systems to be compared, and comparing the performance difference between the two systems at different run lengths. At each comparison point we calculate a statistic using the observations made until that point. Based on this statistic, we either decide to take additional observations from both systems or stop and make a decision on the ranking of the alternatives based on the observations obtained so far. This type of dynamic data collection from simulation is common in the ranking and selection literature but it has mainly been used for steady state simulations (Picthitlamken et al., 2006).

We take observations in batches of k from both systems that will be compared. k could be in terms of numbers of arrivals to the system or in terms of simulated time. $\bar{x}_i^n$ is the average of all the observations taken from system *i* in the *nth* batch. $\bar{X}_i^n$ is the cumulative average of alternative *i* after the *nth* batch. We denote the difference between the systems in batch *n* with $\bar{y}^n$ ($\bar{y}^n = \bar{x}_i^n - \bar{x}_j^n$) and the cumulative difference between the systems after batch *n* with $\bar{Y}^n$ ($\bar{Y}^n = \bar{X}_i^n - \bar{X}_j^n$). We start with taking 3 batches of observations from each alternative and construct a pseudo-confidence interval with half length (*HL*) width using the following formula.

$$HL = t_{1-\alpha/2,n-1}\left(\frac{s_n}{\sqrt{n}}\right)$$

Here $t_{1-\alpha/2,n-1}$ is the upper critical value for *1-α/2* critical value for the *t* distribution with *n-1* degrees of freedom. s_n is the sample standard deviation of the batch average differences ($\bar{y}^1, \bar{y}^2, ..., \bar{y}^n$) calculated using the formula.

$$s_n^2 = \frac{\sum_{t=1}^{n} (\bar{y}^n - \bar{Y}^n)^2}{n-1}$$

If $|\bar{Y}^n| > \frac{HL}{r}$ then we stop and make a decision on the ranking of two alternatives based on their current observed performance. In this formula *r* is a scaling factor to control the average run length used for making the comparisons. If this inequality is not satisfied we take one additional batch from each alternative and calculate the updated *HL* and compare it with the updated cumulative difference. We continue to iterate until a ranking decision is made.

The pseudo code for our heuristic can be seen below.

```
Select k, α, r
Take 2 initial batches from both systems (2*k observations)
n=2
madeDecision = False
While Not madeDecision
    take 1 additional batch from both systems (k additional observations)
    n=n+1
    calculate ȳ^n , Ȳ^n , and s_n^2
    calculate HL = t_{1-α/2,n-1}(s_n/√n)
    if |Ȳ^n| > HL/r then madeDecision = true
end While
Select the best system based on the sign of Ȳ_ij^n
```

Statistic *HL* is calculated using the formula for the half length of a confidence interval with α level of confidence with *n* independent observations. Since we are using very small batch sizes and collecting data during the transient period observations are correlated. This often results in the underestimation of s_{n+} and invalid confidence interval half lengths. However, this is not a major problem since our objective is to compare performance of two alternatives rather than estimating a precise confidence interval of the performance difference.

We also use a scaling factor, *r*, to control the average run length used for making the comparisons. When we pick a small *r* more evidence is required for making a comparison which results in longer runs on the average. Whereas when *r* is big comparison decision can be made with shorter runs. As expected longer runs result in higher probability of correct selection. Thus, *r* should be chosen based on the correct selection requirements of the application and the available computational budget.

4 COMPARISON OF VARIABLE AND FIXED RUN LENGTH FOR THE M/M/1 SYSTEM

We use the M/M/1 queueing problem used in Yang and Lee (2002) as one of our test problems. In this problem, we would like to find the M/M/1 system that gives the minimum average cycle time from a set of 1,000 different alternatives. The alternatives are determined by varying the arrival rate λ from 0.01 to 1.01 in steps of 0.001 and the service rate μ changes accordingly to keep the utilization of the system constant at ρ. Note that, in this setting we know the optimal alternative: λ=1.01 is the best possible configuration, since it minimizes cycle time. (Recall that cycle time is equal to $1/(\mu - \lambda)$ in an M/M/1 system.) We consider ρ =0.65 and 0.95 as two different versions of this problem.

Our objective in the experimentation is to estimate how well our heuristic performs when comparing two solutions picked randomly from the different alternatives in the solution space. Metaheuristics such as genetic algorithms and simulated annealing are widely used in commercial simulation optimization packages and require the implementation of repetitive pairwise comparisons between solutions during the search for an optimal solution.

Although this is a very simple system, transient bias particularly for higher ρ values is hard to eliminate. For example, for the test problem with ρ=0.8 and λ =0.5, we might need to run the simulation until 10,000 customers are served before we start to get a reliable estimate of the cycle

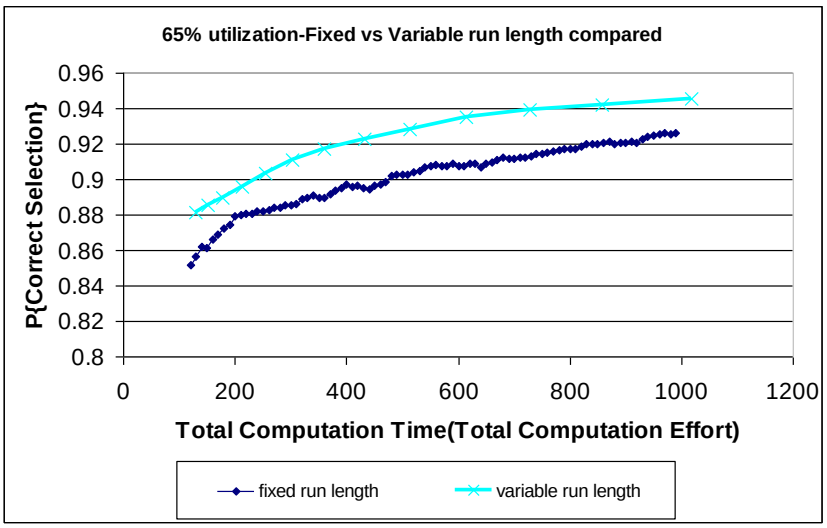


Figure 3 Fixed run length vs. variable run length simulation at 65% utilization of M/M/1 system

time. Whereas, when comparing two solutions, we can achieve probability of correct selection values as high as 80-90% with just 100 customers (i.e., 1/100th of the time required to get a reliable estimate of cycle time). Metaheuristics used in commercial simulation optimization software are fairly robust to this level of noise (Boesel *et al*., 2003).

We now present results from experiments that we have conducted using two different methods that use transient simulation data to compare two alternatives. The first method is the fixed run-length method in which we take a prespecified number of observations from the system and make the decision based on the observed performance of the two systems that are compared. This corresponds to the traditional ordinal optimization approach where the simulation run length specified beforehand. The second method is the variable run length heuristic described in section 3. The variable run length heuristic compares the statistic calculated based on the observations obtained so far with the cumulative difference between the two alternatives. If the absolute value of the statistic is greater than the absolute value of the difference between two alternatives, the comparison is made based on the observed performance at that point. Otherwise, more observations are taken from the simulation.

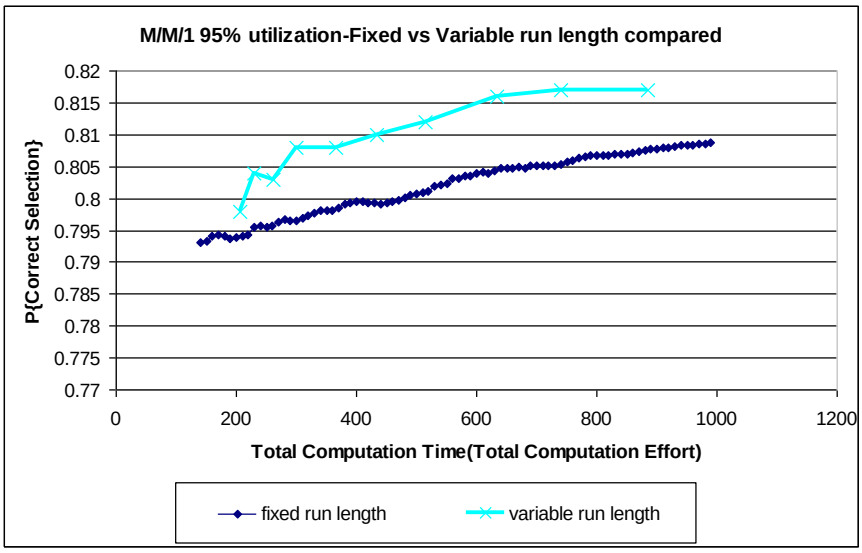


Figure 4 Fixed run length vs. variable run length simulation at 95% utilization of M/M/1 system

In the experimentation, we pick two solutions at random from the solution space (1,001 solutions). For the fixed run length method, we make a new comparison each time 10 additional customers are served and record the success rate of correctly identifying the better solutions. This way we cover the complete range of possible run lengths that can be used with the fixed run length

method. For the variable run length method, we use a batch size of 10 (k=10) and a 95% confidence level for the pseudo confidence interval. We change the run time coefficient parameter, r, between 0.1 and 1.4 to generate similar total computational run time to the fixed run length method. We pick 1,000 solution pairs and perform 30 replications for each pair. The following graphs (Figures 3 and 4) show the estimated probability of correct selection for two different utilization cases ρ=.65 and ρ=.95. The computation effort on the x-axis is number of customers.

For the same computation effort variable run length method has a higher probability of correct selection for all utilization and all computation times. The improvement of the variable run length heuristic could be seen better if average run length required for the same probability of correct selection is considered. For example to achieve a 90% correct selection for ρ=.65 with the fixed run method requires around 400 customers on the average whereas for the variable run length heuristic, the same probability of correct selection can be achieved with ~200 customers – around 50% reduction in the total computational time. Furthermore, the run time improvements are much higher as the utilization increases.

5 FIELD SERVICE SYSTEM LOCATION AND CAPACITY PROBLEM

Our second problem deals with strategic field service planning in the semiconductor manufacturing industry. In this problem, a field service provider has to decide the location and type of the regional service centers to open and the number of service engineers (of different types) to minimize the total fixed regional service center opening, personnel, travel, technology can contractual penalty costs. The problem is further complicated with capability of using a technology called remote diagnostics, which enables the service provider to respond to some portion of service requests remotely. We investigate an instance of this problem with seven customers to serve, four possible regional service center locations, two different service engineer types and two different types of regional service centers (as shown in Figure 5). The nodes in the figure represent the queueing network of service engineers handling service requests from fabs. Each solution represents a possible alternative for the system with respect to which service centers are opened, how many service engineers are employed at each location at each level, and the assignments of fabs to service centers. The objective is to determine the solution that minimizes the expected total cost of the field service provider. Due to second order congestion effects, the expected cost of different solution alternatives can only be estimated using a simulation model. This system is a very complicated system with high variability. To reach steady state and get a reliable estimate for the total expected cost, 10,000 to 20,000 hours of operation has to be simulated. A more detailed description of the problem can be found in Vardar *et al.* (2007).

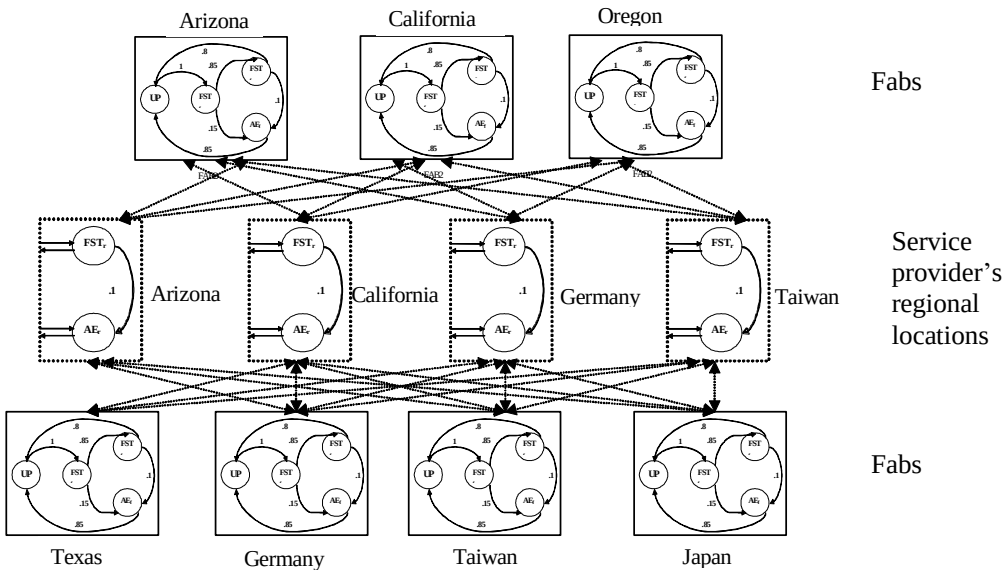


Figure 5 Field service system design problem

We now present results from initial experiments that we have conducted with two methods that use transient simulation data to compare two solution alternatives for the field service system design problem. We, again, compare the fixed run length method and variable length heuristic described above. For the variable run length method, we use a batch size of 96 hours (k = 96hours) and a 95% confidence level for the pseudo confidence interval. We change the run time coefficient parameter, r, between 0.1 and 1.4 to generate similar total computational run time to the fixed run length method. In the experimentation, we pick two solutions at random from the solution space of 1,000 random solutions and make 30 replications for each pairwise comparison. We note that the results are similar to the results in the M/M/1 experiments.

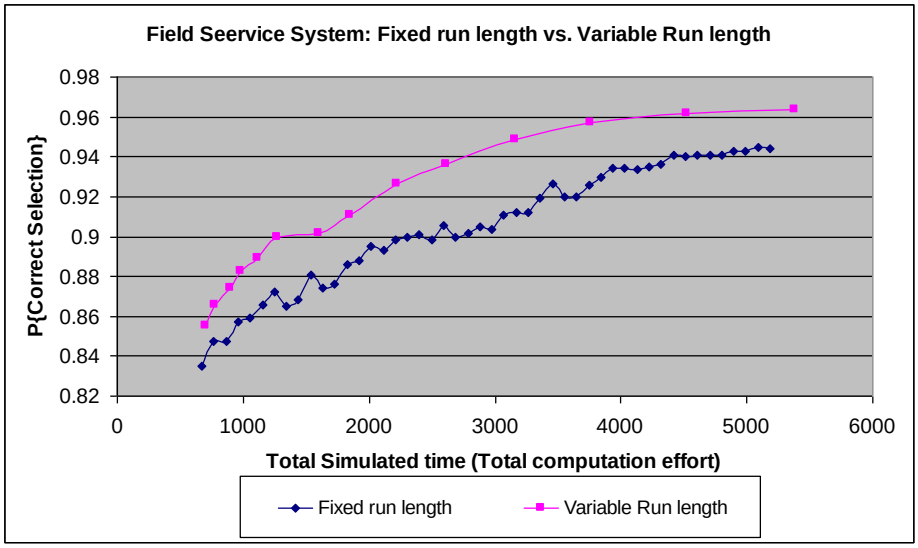


Figure 6 Fixed and variable run length heuristic compared for the field service design problem

6 CONCLUSIONS AND DISCUSSION

In this paper, we have presented a new method for using transient simulation runs to compare different solution alternatives for simulation-based optimization. Although transient simulation runs do not provide good estimates for the performance of alternatives, they can provide fairly consistent and useful inferences about how an alternative performs compared to another alternative. This type of inference is particularly useful in metaheuristics commonly used in commercial simulation-based optimization packages where this type of comparison between alternatives has to be made repetitively. In some cases, adequate correct selection rate can be obtained from transient simulation using 1/100th of run time needed for reaching steady state.

We presented initial experimentation results for two different problems: an M/M/1 system and a field service system design problem from the semiconductor industry. Our dynamic method provides the same level of correct selection with significantly less simulation run time in all cases presented. The threshold and step size parameters used in our method have an effect on the performance of the algorithm. Vardar (2006) explores ways to find effective values of the parameters in our method, independent of the problem type. It also modifies the variable run length heuristic to consider whether or not the sample paths of the transient solutions of the two solutions are trend converging or diverging which is helpful, particularly in the case of the single crossover ranking relationship.

Until recently, simulation-based optimization was not a practical tool to be used in problems that are faced by the semiconductor industry, since these problems often required the use of complex simulation models and long run times. However, advanced metaheuristics and the use of ordinal optimization ideas can open the way for successful use of simulation-based optimization in semiconductor industry for operational and strategic problems.

REFERENCES

Boesel J, Nelson B L, and Kim S H (2003). Using ranking and selection to 'clean-up' after simulation optimization. *Operations Research* **51**: 814-825.

Dai L Y, (1996). Convergence properties of ordinal comparison in the simulation of discrete event dynamic systems. *Journal of Optimization Theory and Applications* **91(2)**: 363–388.

Fu M C (2001). Optimization for simulation: theory and practice. *INFORMS Journal on Computing* **14 (3)**: 192-215

Ganz A and Wang X (1994). Efficient algorithm for virtual topology design in multihop lightwave networks. *IEEE/ACM Transactions on Networking* **2(3)**: 217-225.

Ho Y C, Sreenivas R S, and Vakili P (1992). Ordinal optimization of DEDS. *Discrete Event Systems: Theory and Applications* **2**: 61-88.

Ho Y C and Larson M E (1995). Ordinal optimization approach to rare event probability problems. *Journal of Discrete Event Dynamic Systems* **5(2–3)**: 281–301.

Lau T W E. and Ho Y C (1997). Universal alignment probability and subset selection for ordinal optimization. *Journal of Optimization Theory and Applications* **93(3)**: 455–489.

Law A (2014). *Simulation modeling and analysis* 5th edition. McGraw-Hill, New York.

Pichitlamken J, Nelson B L, and Hong L J (2005). A sequential procedure for neighbourhood selection-of-the-best in the optimization via simulation. *European Journal of Operational Research* **173**: 283-298.

Vardar, C. (2006). A novel simulation optimization approach for a remote diagnostics capable field service system design problem. *Arizona State University Dissertation*.

Vardar C, Gel E S, and Fowler J W (2007). A framework for evaluating remote diagnostics investment decisions for semiconductor equipment suppliers", *European Journal or Operational Research* **180**: 1411-1426.

Voss P A, Willemain T R, Haddock J, (2005). Estimating the steady-state mean from short transient simulation. *European Journal of Operational Research* **162**, 403-417.

Wieseltheir J E, Barnhart C M, and Ephremides A (1995). Ordinal optimization of admission control in wireless multihop voice_data network via standard clock simulation. *Journal of Discrete Event Dynamic Systems*, **5**: 243–280.

Xie X L (1997). Dynamics and convergence rate of ordinal comparison of stochastic discrete event systems. *IEEE Transactions on Automatic Control*, **42(4)**: 586-590.

Yang M S, Lee L H and Ho Y C (1997). On stochastic optimization and its applications to manufacturing. *Lectures in Applied Mathematics*, AMS-SIAM, **33**: 317–331.

Yang M S, and Lee L H (2002). Ordinal optimization with subset selection rule. *Journal of Optimization Theory and Applications* **113(3)**: 597–620.

AUTHOR BIOGRAPHIES

JOHN FOWLER is the Motorola Professor of Supply Chain Management in the W.P. Carey School of Business at Arizona State University. His research interests include discrete event simulation, deterministic scheduling, and multi-criteria decision making. He has published more than 125 journal articles and more than 100 conference papers. https://isearch.asu.edu/profile/36406

ESMA GEL is an Associate Professor in the School of Computing, Informatics and Decision Systems Engineering in the Ira A. Fulton Schools of Engineering at Arizona State University. Her research focuses on the use of applied probability techniques for modelling, design and control of production systems and supply chains, with particular emphasis on workforce engineering. Her work has been published in leading journals and funded by the National Science Foundation, as well as industrial partners such as Intel, IBM and Infineon. Gel is a member of INFORMS, IIE and ASEE. https://isearch.asu.edu/profile/257696

CEM VARDAR is a Principal Optimization Engineer at Carvana. He designs, develops, tests and maintains decision support systems based on mathematical models that run on massive data sets. He approaches decision support system design from an engineering perspective and not merely as scientific problems. https://www.linkedin.com/in/cem-vardar-70119514/

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.028

SIMULATION BASED DECISION SUPPORT SYSTEM TO DETERMINE PRODUCTION QUANTITY FOR A LOW SHELF LIFE PRODUCT

Mr. Amit Kumar

ITC Infotech India Ltd.
Cox Town, Bangalore 560005
amit.m@itcinfotech.com

Mr. Krishna Chaitanya Pappu

ITC Infotech India Ltd.
Cox Town, Bangalore 560005
krishnachaitanya.pappu@itcinfotech.com

Mr. Sumit Kumar

ITC Infotech India Ltd.
Cox Town, Bangalore
kumar.sumit@itcinfotech.com

ABSTRACT

In this paper we show how a Decision Support System (DSS) using simulation of a low shelf life pharmaceutical product's supply chain was created. The main technique that determines the production quantity is based on a simulation-based optimization approach. The system allows the planners to quickly see the impact of various strategies and changes in policies and take the best decision.

The medicine is for a critical disease and hence the service level must be maintained upwards of 99.9%. To maintain this high service level, especially when the shelf life of the medicine is three weeks is a challenge for any supply chain. The tradeoff between service level and inventory scrapping cost (due to medicine expiry) is optimized using a simulation-based optimization approach.

Keywords: Discrete event simulation modelling, Anylogic, Inventory optimization, Low shelf life

1 INTRODUCTION

The traditional optimization methods such as mixed integer programming, dynamic programming are well suited for strategic and tactical level decision making. Whereas for many operational and tactical level decision making in a complex real-world scenario, optimization methods are sometimes not practical. This is mainly because of the limitations of deterministic optimization methods of not being able to consider variability or uncertainty inherent in the system. Another challenge is that many of the real problems are combinatorial optimization problems which are NP-Hard. In such a situation simulation becomes a viable tool for problem solving. Hybrid models that combine simulation and optimization are also becoming more popular and demonstrating their efficacy (Muhammed Ordu et.al., 2020)

The client is a world leading pharmaceutical company with headquarters in Europe. For a medicine which treats a terminal illness, the shelf life of the medicine is as low as 21 days. Since this is a critical medicine delivered directly to the hospitals the service level needs to be maintained at upwards of 99.9%. The obvious risk is of write-offs. The client was facing the challenge of high scrapping due to medicine expiry.

We delivered a solution based on the Anylogic simulation software (Borshchev A and Grigoryev I, 2020) that would take multiple uncertainties into account and simulate the supply chain with these uncertainties while optimizing the inventory to achieve the desired service level and minimizing the write-offs cost.

The organization of this paper is as follows. In section 2 we give a literature review for similar problems. In section 3 we give a detailed description of the problem. In section 4 we explain our solution methodology and in section 5 we share the results of our work.

2 LITERATURE REVIEW

The management of the perishable products supply chain has gained traction in the past decade (Amorim et al., 2011). There has been research to develop sustainable solutions to handle product perishability, in the areas of manufacturing, storage, packaging, and transport operations (Accorsi et al., 2017). Most of the studies have focused on a common objective of improving the overall supply chain performance.

Looking into the area of manufacturing of perishable products, most of the research was focused on addressing the issues like supply fluctuations, improving quality, improving productivity and production technologies with a little attention to wastage at production (Yared Lemma et al., 2014)

When it comes to inventory decisions, the models related to inventory, which decays in terms of its utility over time are analysed. Zhaotong Lian and Liming Liu (2001) proposed a heuristic approach for continuous review of perishable inventory systems. QinglinDuanT and WarrenLiao (2013) proposed a simulation optimization methodology for inventory management of perishable products by considering order-up-to policy approach for handling highly perishable products. Xiaojun Wang and Dong Li (2012) proposed a dynamic pricing model to reduce food spoilage waste and maximise profit through a pricing approach based on dynamically identified shelf life. While the research on inventory models of deteriorating items has increased greatly over the last years, there is still some limited attention to inventory loss reduction decisions of perishable products in uncertain environments

This paper focuses on loss reduction in perishable product supply chains by determining the quantity of production to minimize the inventory scrapping cost. We have used simulation-based optimization approach to create a DSS.

3 PROBLEM DESCRIPTION

The supply chain for the medicine looks as shown in Figure 1. There are two manufacturing facilities, two distribution centers and numerous hospitals where the medicine should reach. Most of the demand in Europe, African and part of the Asian regions is served from Europe, with one of the manufacturing facilities and distribution centers located in Europe, while the rest of demand from US, Canada and rest of the Asian regions are served from manufacturing facility and distribution centers located in US.

Figure 1 *Supply chain overview*

Since the shelf life is only three weeks, the production is done based on forecast and then the product is pushed to the distribution centers. The flow from distribution center to hospital happens on actual order. The limited shelf life of the medicine is due to natural decay of the medicine. As an example, a 10-day old medicine can be used for a patient with weight around 100 kg, whereas a 15-day old medicine can be used for a patient with weight less than 70 kg. After 21 days the medicine can only be used for patients below 45kg, and hence we consider them expired (the numbers mentioned are only for illustration).

The manufacturing facility produces various other products also in the same production line on a production campaign basis. The production campaign details are known in advance. We know the days in a week (typically one or two days) when the concerned medicine is produced. The decision support system needs to output what quantity should be produced on those days. The production capacity is enough and hence no production capacity constraint is necessary.

The medicine needs to be shipped to the hospital. For the purpose of lead time calculation for planning each country is divided into one or more delivery zones depending on country size. The lead

time information is maintained and used at the zone level. Each hospital is assigned to a zone. Forecast is done at a week and zone level.

There are various sources of variability in this supply chain. The production is done based on forecast which comes with a forecast error. The weight of patient and the weekday of the treatment cannot be determined in advance and hence have uncertainty. The production can have batch failures. The transportation lead times have variability.

4 SOLUTION METHODOLOGY

The Enterprise Resource Planning (ERP) system provides master data such as data of the Stock Keeping Units (SKU), mapping of which SKU can be supplied in which country, historic lead times from manufacturing facility to Distribution Centre (DC) and DC to delivery zones. Transactional data such as proposed lead times for both legs of distribution based on weekday of shipment, production campaigns, current inventory snapshot, forecasts are also read from ERP database. In Figure 2 we provide an overview of the technical architecture of our solution.

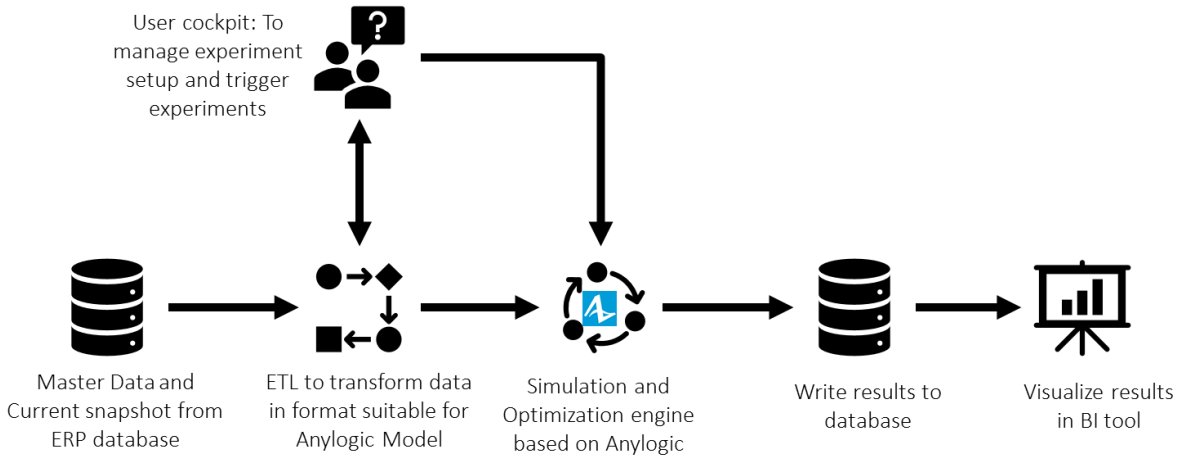


Figure 2 *Technical architecture of solution*

User Cockpit (excel based UI) facilitates users with options to choose the horizon of simulation, whether to run in pure simulation mode or digital twin mode. Digital twin mode will initialize experiment with current snapshot information whereas pure simulation model will use a warmup period. The length of the warmup period is also controlled from the user cockpit. Other options include choosing simulation or simulation + optimization, consider or adjust forecast biases and uncertainties, choose among production and distribution strategies etc. Once user is satisfied with the settings of the experiment, he can trigger the experiment from user cockpit itself. This will first launch the Extract Transform and Load (ETL) phase which will pull data from ERP and transform it and write data into Anylogic database in a form that the model will use. Once the ETL phase is complete, the ETL will communicate this back to cockpit logic and now cockpit will trigger the simulation experiment. Additionally, ETL will also perform all the checks and validations on the data. If any errors or warnings are found it will write the messages to the log file.

In Simulation phase, the simulation model runs the scenarios chosen by the user and on completion pushes the output results to database table.

The database loads the data from tables to views required for visualisations in a Business Intelligence (BI) Tool. The dashboard provides Key Performance Indicators (KPIs) visualisation to analyse the demands, production plans, write offs, compare statistics across the different scenarios and identify the parameter values that provides optimal write-offs and service level.

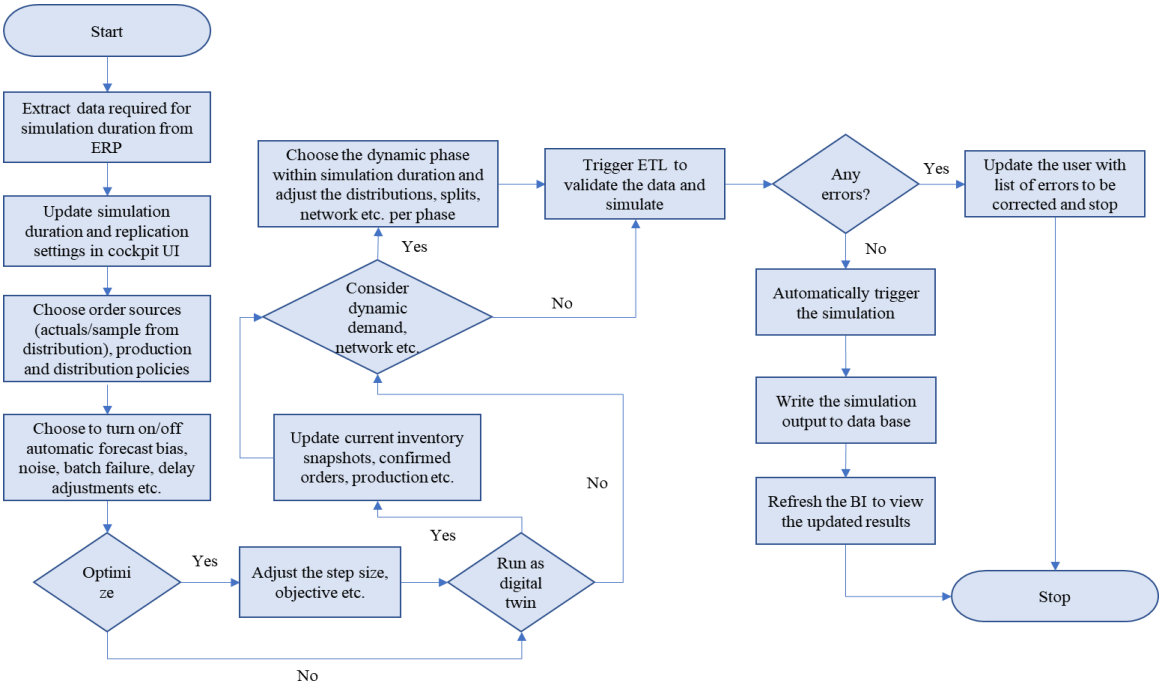


Figure 3 *Solution flow chart*

4.1 Overview of production policy

The production policy determines the quantity to be produced on production days by considering the inputs like production campaigns, confirmed orders, starting inventory, demand forecasts, zone splits(historical split of a country's demand at delivery zone level), weekday splits (historical split of demand at weekday level at hospitals), picking priorities etc. The planning is done on a weekly basis to determine the quantity to be produced for each SKU in the corresponding weeks. Different policies or production logic were tried and with each policy certain parameters were chosen to fine tune the policy using simulation-based optimization approach.

While the confirmed orders have all the required details like patient weights, days of treatment, zones of treatment etc. the forecasts are converted into daily orders by assigning these weights, treatment days, zones etc. through probabilistic sampling.

On the supply side, the variables that have impact on the service level have been determined by running the simulation over multiple replications by fixing the other attributes. The variables that had major impact on the service level have been included for supply side planning

The supply-side uncertainties considered are like batch failures, shipment delays, shipment damages etc. and demand side uncertainties like demand surges, high patient weights etc. are considered while planning for the safety quantity in addition to the cycle quantity.

The orders generated from forecasts are netted using available inventory first, followed by allocating them to the freshest production campaign to maintain highest serviceability and taking care of some demand side uncertainties like patient weights and treatment days. The production quantity per campaign is finalised by aggregating all the orders planned against each campaign and the quantities aggregated against the campaigns from immediate next week are frozen and sent for production. Hence, on the Monday of Week0 (current week), we plan and freeze the production in Week1 (next week).

The safety quantity represents the extra days of coverage that will be kept. This is determined at a country level and these are the parameters that will be optimized to fine tune the production policy selected.

4.2 Stock optimization and scenario planning

By changing the settings in user cockpit, we can perform optimizations as well as simulate and evaluate multiple scenarios. Some of these scenarios are explained below:

Stock optimization: Users can select among few different production policies and for each policy determine safety quantity at a country level that provides minimum write-offs while maintaining the highest service level.

Validate production plans: Users can validate the production plans by generating the demand sampled from a distribution and simulating it over multiple replications and evaluating the output. The demand will have the details of delivery zone, patient weight and treatment date.

Switch on/off uncertainties: Users can turn on/off multiple uncertainties while planning for production to obtain a robust output. Ex: consider/adjust lead time distributions, consider/adjust production batch failures, consider/adjust shipment delays, consider/adjust forecast bias, consider/adjust forecast noise etc.

Future projections: Users can simulate and project the sales into long term/short term future by using dynamic demand distributions for different periods in future catering to seasonality/trends in demand.

Evaluate risks by dynamic network changes: Users can evaluate dynamic network changes by switching the production and distribution sources and destination zones during special weeks/manufacturing facility closures etc. for specific time periods.

4.3 Simulation based optimization approach

The general framework of a simulation-based optimization approach is provided by C. Almeder and M. Preusser (2007). We use a similar approach. The Anylogic simulation software also provides an optimization engine based on OptQuest. We use the OptQuest engine as a black box for optimization. Doing multiple simulation runs and analysing the output of these runs we can see the impact of variation. In each simulation run different numbers will be sampled wherever we are sampling from a probability distribution such as in lead time, patient weight, forecast quantity etc. These runs are called replications. Over several such runs we can analyse any metric that we are interested in, such as the service level. We can observe the mean value and the spread of this metric. In order to make the confidence interval smaller we need to run more replications.

We already discussed that the parameters in this optimization are the safety quantity that need to be produced at a country level. This safety quantity is in terms of extra days of cover. Since the OptQuest optimizer is a meta-heuristic based optimization engine we also need to define the range in which the optimizer will search for the optimal value and the step size. For example, for a particular country, we want the optimizer to search the days of cover from 1 day to 5 days with a step size of 1 (or a non integer value for partial days of cover). This means that for this country the optimizer has choice of 1,2,3,4 or 5. Each time the optimizer fixes the values of these parameters, we call it an iteration. The number of iterations and the number of replications to be run during optimization are the hyper parameters that can be configured from the user cockpit. We found in our case 100 replications and 500 iterations lead to a good solution. This number was arrived at using a trial and error approach. It is important to understand that the optimizer chooses a different set of parameters for each iteration and for each iteration (with the same value of the parameters) multiple replications are run. The overall approach is summarized in Figure 3.

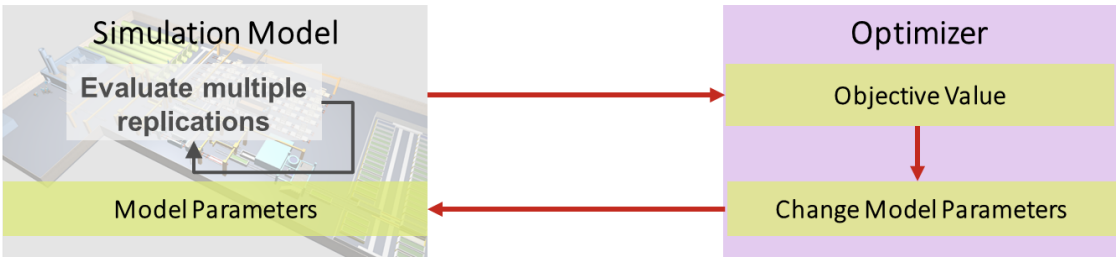


Figure 4 *Simulation based optimization approach*

Since the total number of countries is large, we do not explicitly optimize for all the countries. Using the pareto principle (also known as the 80/20 rule) we arrived at eight countries that cover 80% of the demand. We optimize the days of cover for these eight countries and we use a ninth parameter which represents the days of cover in all other countries. In total we had nine parameters to be optimized. This way the problem was made tractable in an acceptable amount of time.

5 RESULTS AND FUTURE SCOPE

The most important result from our decision support system is shown in Figure 4. We show here the output of the simulation-based optimization experiment with 100 iterations, where each iteration has 100 replications. Each dot represents a single iteration with the values of the parameters selected by the optimizer. The y-axis shows the service level and the x-axis shows the number of medicines scrapped for that iteration. While the optimizer does output the best iteration among all the iterations, we found it was better to select the best iteration visually by looking the plot in Figure 5. Each iteration will have 100 values for service level and corresponding scrap count as we run 100 replications. Hence each iteration will not be a single point but will have a spread in service level as well as in scrap count. We have shown the average values across the replications without the spread in Figure 4 and with spread of service level in Figure 5. There will also be a spread in the write-offs which is not shown here. This is one major reason that we select the best iteration visually. We can discard any iterations that are not robust (have a high spread of either service level or scrap count or both).

Through our improved production policy we could show a reduction in scrapped medicine by 28% as compared to the current performance without impacting the service leve1 as shown in Figure 4.

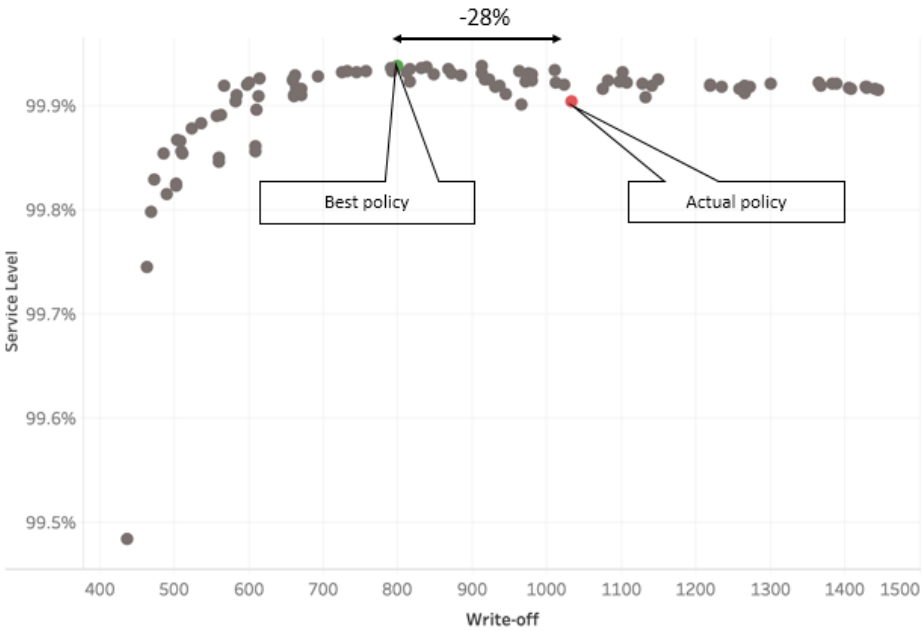


Figure 5 *Trade-off between service level and write-offs*

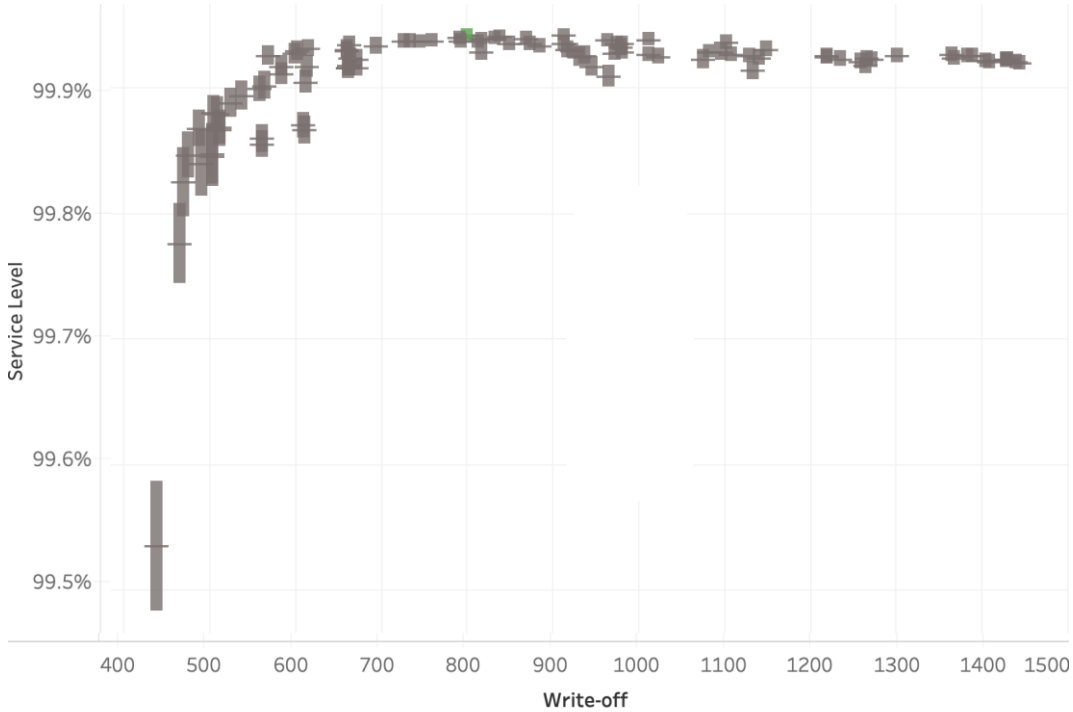


Figure 6 *Trade-off between service level and write-offs with variation in service level*

Once we have a simulation model of the supply chain created, we can try multiple ideas or scenarios to find an option that reduces cost or adds value in any other way. One of the scenarios we tried was to see the impact of lead time reduction. Another scenario tried was to delay the manufacturing a day. The current operations have a gap of 1 or 2 days between the date of manufactuing and date of shipment to cater to any delays/batch failures. A scenario of delaying the manufacturing by a day would increase the freshness of the batch on the day of treament there by enabling the ability to serve heavier patients with lesser number of medicines than earlier. One more scenario of potential saving by forecast bias reduction was tried. There are many countries which overforecast the demand just to avoid any stock outs.

For future work or further improvement in results we believe that in addition to using an optimization based approach it is possible to train a reinforcement learning based AI to make the decision of how much to produce. There has always been an interest in combining simulation with Artificial Intelligence(AI) techniques (Robert M et.al., 1987 and S Robinson et.al., 2005). With the latest infrastructure provided by companies such as Microsoft Bonsai as well as PathMind we feel that this approach should also be explored.

REFERENCES

Accorsi, R., Gallo, A., & Manzini, R. (2017). A climate driven decision-support model for the distribution of perishable products. *Journal of Cleaner Production*, 165, 917–929.

Borshchev A and Grigoryev I (2020). *The Big Book of Simulation Modeling: Multimethod Modeling with Anylogic 8*. Lightning Source Inc.

C. Almeder and M. Preusser (2007). A toolbox for simulation-based optimization of supply chains, *Proceedings of the 2007 Winter Simulation Conference,* Washington, DC, 2007, pp. 1932-1939, doi: 10.1109/WSC.2007.4419821

C. Kouki, E. Sahin, Z. Jemai, Y & Dallery (2013). Assessing the impact of perishability and the use of time temperature technologies on inventory management, *International Journal of Production Economics*, Volume 143, Issue 1, Pages 72-85

Muhammed Ordu, Eren Demir, Chris Tofallis & Murat M. Gunal (2020). A novel healthcare resource allocation decision support tool: A forecasting-simulation-optimization approach, *Journal of the Operational Research Society*, DOI: 10.1080/01605682.2019.1700186

OptQuest Optimization of Complex Systems, https://www.opttek.com/sites/default/files/pdfs/OptQuest-Optimization%20of%20Complex%20Systems.pdf accessed on 03/09/2020

Pedro Amorim, Carlos H. Antunes & Bernardo Almada-Lobo (2011). Multi-Objective Lot-Sizing and Scheduling Dealing with Perishability Issues, *Industrial & Engineering Chemistry Research 50(6)*, DOI: 10.1021/ie101645h

Robert M. O'Keefe & John W. Roach (1987). Artificial Intelligence Approaches to Simulation, *Journal of the Operational Research Society*, 38:8, 713-722, DOI: 10.1057/jors.1987.120

S Robinson, T Alifantis, J S Edwards, J Ladbrook & A Waller (2005). Knowledge-based improvement: simulation and artificial intelligence for identifying and improving human decision-making in an operations system, *Journal of the Operational Research Society*, 56:8, 912-921, DOI: 10.1057/palgrave.jors.2601915

S. Nahmias (1982). Perishable inventory theory: A review, *Operations Research*, vol. 30, no. 4, pp. 680-707, 1982

QinglinDuanT. WarrenLiao (2013). A new age-based replenishment policy for supply chain inventory optimization of highly perishable products, *International Journal of Production Economics*, Volume 145, Issue 2, Pages 658-671

Xiaojun Wang, Dong Li (2012), A dynamic product quality evaluation based pricing model for perishable food supply chains, *Omega*, Volume 40, Issue 6, Pages 906-917

Yared Lemma, Daniel Kitaw & Gulelat Gatew (2014). Loss in Perishable Food Supply Chain: An Optimization Approach Literature Review, *International Journal of Scientific & Engineering Research*, Volume 5, Issue 5

Zhaotong Lian & Liming Liu (2001). Continuous Review Perishable Inventory Systems: Models and Heuristics, *IIE Transactions* volume 33, pages 809–822

AUTHOR BIOGRAPHIES

AMIT KUMAR received a BTech in Electronics and Instrumentation Engineering from National Institute of Technology, Rourkela, India in 2008. He completed his Masters in Industrial Engineering from National Institute of Industrial Engineering (NITIE), Mumbai in 2012. He currently works as a principal consultant at ITC Infotech in Bangalore, India. https://www.linkedin.com/in/amit-kumar-71b8906/

KRISHNA CHAITANYA holds a Bachelor's in mechanical engineering from Osmania University, Hyderabad, India and Masters in Industrial Engineering from National Institute of Industrial Engineering, Mumbai India. He is currently working as principal consultant in ITC Infotech, Bangalore, India.

SUMIT KUMAR leads the Digital Solutions for manufacturing and supply chain at ITC Infotech, Bangalore, India. He has a BTech in Civil engineering from IIT Kharagpur and MBA in Operations Management from S.P. Jain Mumbai, India.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.029

ANALYSIS OF COMPLEXITY AND SIMULATION USAGE IN MANUFACTURING SMES

Dr. James Byrne

DCU Business School,
Dublin City University,
Glasnevin, Dublin 9, Ireland
james.byrne@dcu.ie

Dr. Paul Liston

Dept. of Accounting & Business Computing
Faculty of Business and Hospitality
Athlone Institute of Technology, Ireland
pliston@ait.ie

Prof. PJ Byrne

DCU Business School,
Dublin City University,
Glasnevin, Dublin 9, Ireland
pj.byrne@dcu.ie

ABSTRACT

Discrete event simulation (DES) is a powerful decision support methodology particularly suited to complex systems where the relationship between individual components is difficult to comprehend. This study aims to investigate DES usage and its relevance in manufacturing SMEs using a survey-based calculated complexity value for each organisation. A study was conducted to assess Irish manufacturing SMEs' perception of their organisation's complexity versus a survey-based calculated complexity value. Complexity levels were then used to determine the potential for DES usage. Respondents were also asked to respond on DES awareness and usage. This study suggests that manufacturing SMEs seem to underestimate the complexity of their systems and consequently the usefulness of DES. The poor uptake of DES among manufacturing SMEs can be partly attributed to a lack of awareness. This paper presents a call to action by the DES modelling community including practitioners, academics and software vendors in support of manufacturing SMEs.

Keywords: Discrete Event Simulation, DES, Complexity, Small and Medium Enterprises, SME

1 INTRODUCTION

Manufacturing small and medium-sized enterprises (SMEs) face a number of operational challenges such as scheduling, flexibility and lead time, customization, product returns, maintenance operations and the supply chain to note just a few. What is notable is that these challenges are not specific to companies of SME status but are also regularly encountered by larger organizations. However, the unique characteristics of manufacturing SMEs affect the conditions required to work with manufacturing strategies (Löfving et al., 2014). In addition, manufacturing SMEs are consistently challenged to improve efficiencies as they compete in a global market against large well-resourced multinational companies. In efforts to address such operational (and also strategic) challenges, larger organizations quite often evaluate the manufacturing technology landscape and the potential solution supports this makes available to them. Discrete Event Simulation (DES) is one such tool which can be applied in such situations. While some larger companies have gained from this methodology, smaller firms seem less likely to take advantage of its capabilities (Ingemansson et al., 2002). This low uptake by SMEs has occurred despite DES being available for a considerable period of time and its benefits being well publicized. Given the independent and peer-reviewed validation of the advantages of the use DES including dedicated academic journals, conferences and workshops on the subject there are a

number of open and as of yet unanswered questions in this regard: why is DES modelling not ubiquitous across the entire manufacturing sector? Do companies think that DES is not suited to their specific problem types? Is there an inherent aversion to this form of analysis? How many of those at the "coalface" are actually aware of DES? It is generally accepted that DES, although beneficial, can be costly and time consuming to undertake (Fowler and Rose, 2004; Johansson et al., 2008). It is proposed that the context in which DES offers greatest return is one where the system under investigation is sufficiently complex to justify analysis beyond standard spreadsheet-based calculations. Indeed, Robinson (2005) notes that "simulations are normally developed because a system is too complex to be represented in any other way". The challenge with this perspective is that 'complexity' is, in and of itself, difficult to define. There has been extensive research on the subject of complexity with many publications offering insight from different perspectives including that of manufacturing (e.g. Garzon et al. (2012); Lang et al. (2014)), and the topics of simulation and complexity have been considered together e.g. Zhang (2011); Aelker et al. (2013).

In this work, the authors aim to examine the level of complexity across a sample of companies in the Irish manufacturing sector and indicate how this can affect their potential use of DES. To this end, a brief overview of the literature on SME technology strategy and the broader topic of complexity is presented, followed by details of the data collection and analysis methodologies, the consequent findings and finally the conclusions drawn from the research.

2 LITERATURE REVIEW

The concept of complexity has been studied in great detail and the literature is comprehensive and varied in relation to the approaches and perspectives taken. Studies of complexity have been conducted across a large array of disciplines and areas in the manufacturing domain such as understanding the architecture of complexity (Simon, 1962); investigating the complexity of scheduling in SMEs (Garzon et al., 2012); determining the impact of manufacturing complexity on performance (Pradhan and Damodaran, 2009) and managing complexity (Götzfried, 2013). The meaning of the word complexity is considered to be vague and ambiguous (ElMaraghy *et al.* 2012) with others affirming that there is no real universally accepted definition or consistent interpretation of complexity (Aelker et al., 2013). In their attempts to understand complexity many have sought to determine how it can be quantified or measured (e.g. Modrak and Marton 2012) while others have focused on how it can be modelled and simulated (e.g. Cicirelli et al., 2011; Zhang, 2011). Some systems can be so complex as to be almost impossible to fully understand. Indeed Beer (1959) argued that in the world there exists a class of "exceedingly complex systems", which are in principle unknowable, we can never know them completely and they can always surprise us, including the brain, firm and economy (Pickering, 2004). Much work has been done in the area of complex adaptive systems (CAS) theory (e.g. see Anderson (1999), Lewin (1999), Choi et al. (2001)) in which a perfect understanding of the individual parts of a system does not automatically convey a perfect understanding of the whole system's behavior (Miller and Page, 2017). In all, a significant volume of complexity-related literature has been published. In determining the nature of complexity studies four categories have been identified in relation to the different types of complexity that can exist. The predominant categories of complexity that have emerged are static and dynamic complexity (Gabriel, 2008; Serdarasan, 2013) and *internal and external complexity* (Gabriel, 2008; Götzfried, 2013).

- *Static complexity* can be referred to as structural complexity (ElMaraghy et al., 2012), which is time-independent and intrinsic in the product and systems structure (Gabriel, 2008).
- *Dynamic complexity* relates to the unpredictability in the behaviour of a system over time and is thus time dependent and related to the operational behaviour of the company (ElMaraghy et al., 2012). Gabriel (2008) notes that the events in manufacturing systems that lead to dynamic complexity include machine breakdowns and quality failures.
- *Internal complexity* is related to the complexities experienced within the company and is considered endogenous (Götzfried, 2013). The facets which are deemed to be in control of the managers, such as product offerings, types and amount of equipment, degree of vertical integration, systems design and maintenance, reliability, quantity and timing of materials and tools all contribute to internal complexity (Gabriel, 2008).

- *External complexity* relates to those factors which are outside the control of the company and its managers, and include factors such as: customer expectations and demands; regulations and market and environmental changes; intensifying competition; external stakeholders such as customers, suppliers, distributors and regulatory bodies (Blome et al., 2014; Efthymiou et al., 2012; Gabriel, 2008; Götzfried, 2013).

Many features of manufacturing environments contribute to complexity and several frameworks for outlining the drivers of complexity have been devised (Aelker et al., 2013). Following a literature analysis this paper proposes six general complexity categorizations (determinants), see Table 1.

Table 1 *Determinants of Complexity*

Product/Parts	High product mix; Multiple part types made in the same facility/line; Multiple levels of subassemblies; Product features and capabilities; Electrical and mechanical components, software, and human-interfaces;
Process	Number of process technologies; Number of manufacturing steps; Batch processing; Automation: highly integrated with all levels of the enterprise; Layout of components in a system and connectivity between them; Number of routes through the factory; Information flow: internal, external and intra-plant; Planning and scheduling functions: strategies, number, content, timing & priority of documents for planning and scheduling, decision making; Precedence constraints for set of operations for producing a part type; Relationships between parts/components; Many elements with numerous but simple interrelations;
Equipment	Complex equipment that leads to high levels of preventive maintenance and downtime; Multiple products made on different machines; Control systems and software; Machine capabilities; Unpredictability over time;
People	Operator absenteeism; Human cognitive ergonomics;
Supply Chain	Delays and faults in raw materials; Resource unavailability;
Market	Complexity of market forces: global competition, turbulence, variety, short delivery, zero defects; Customer change requests; Shorter product lifecycles; More intense competition; Rising customer expectations; Social and environmental pressures; Government legislation and standards;

In terms of the measurement of complexity, Frizelle and Woodcock (1995) developed a mathematical model utilizing entropy as a measure of the complexity in a system. Their proposed method considers the elements of a system, the extent to which they interact and the degree to which each operational source contributes to the overall complexity of the firm. Validation of the method was presented by way of three case studies. Further validation of the entropy method was provided by Frizelle and Suhov (2008) where an additional three cases were tested. However, no evidence was found of widespread usage beyond these validation cases. Around the same time the Meyer and Foley Curley (MFC) approach was developed (Meyer and Curley, 1995). The MFC approach introduces the concepts of knowledge and technology complexity, uses a scoring methodology for defined system variables and was found to give good insight at a relatively low cost. In a study of the two approaches Calinescu et al. (1998) found that although the entropic method gave greater insight, it had a significantly higher input cost. Examples of researchers trying and failing to successfully measure complexity are also available. Gabriel (2008) proposed and tested a quantitative measure for the manufacturing complexity that results from system design, referred to as the internal static manufacturing complexity (ISMC) measurement. The ISMC was focused on product line complexity, product structure and process complexity components and consisted of eight measurable factors.

In studies associated with measuring complexity in manufacturing environments, there is often a written or unwritten assertion that complexity is a negative attribute and that the proposed measure of complexity can either be used to identify the less complex option when making decisions or as a yardstick for process improvement initiatives. Deshmukh et al. (1998) however argue that an increase in static complexity can in fact result in an improvement in system performance if the system is operated optimally. They note, that for a given part mix, static complexity can be increased by adding machines to the shop floor, changing the capability of some machines so that they can handle more

operations, and by changing the process plans for parts such that the operations can be processed in any sequence. The foregoing descriptions demonstrate only a small sample of the efforts to observe, assess and develop measures for gauging how the various components of a business environment can lead to complexity. However, despite the breadth of literature, little is published in relation to complexity in manufacturing SMEs and more specifically the role that DES for such SMEs may hold.

3 METHODOLOGY

This study aims to better understand why DES uptake is low, particularly amongst SMEs, and determine if there is justification for attempting to improve this metric. To gain the perspective of this business sector, the opinions and details of a representative sample of manufacturing SMEs based in Ireland were collected and studied. With the research objectives in mind, a survey was selected as the most appropriate data collection mechanism and a set of questions was compiled based on past DES study experience of the research team and the relevant literature. With no readily available dataset of potential respondents, a number of steps were undertaken prior to the survey being conducted. The first step involved defining the participant category of interest (i.e. companies that meet the SME criteria and operate in the manufacturing sector). Using these criteria, a dataset of 500 Irish manufacturing SMEs was created through contacting local and national enterprise boards and government bodies and reviewing media reports. Information was collated on company name, type, turnover (where possible) and contact details.

DES requires a certain amount of complexity in the system of interest in order for companies to realize the maximum value from the model. Without this level of complexity other tools such as spreadsheet modelling and queuing methods may provide a better solution for SMEs striving to obtain an understanding of, and evaluate proposed changes in, their processes (Chance et al., 1996; Fowler and Rose, 2004). Consequently, in addition to gaining details of the companies and their attitude toward DES, one objective of the survey was to establish a sense of the complexity within respondents' systems. As noted from the literature, defining, understanding and measuring complexity is not an exact science. Nonetheless, previous studies have built up a knowledgeable base from which to understand the concept of complexity. Bar Garzon et al. (2012) no other literature was found which focuses on the topic of SME complexity with most studies dealing with large organizations. Garzon et al. (2012) focused on scheduling complexity and does not offer a broader framework or set of measures to apply to manufacturing SME complexity. With no clearly defined set of criteria with which to measure SME complexity, reference was made in this study to the literature on the determinants of complexity (see Table 1) and the criteria used in past DES studies.

Given that companies were being studied remotely via survey, the time and effort required to engage in the in-depth complexity measurement processes (e.g. the entropy rate or MFC methods) was deemed prohibitive. Instead, a simpler two-pronged (perception versus calculation) approach was designed and implemented. The first element of this approach was a direct question asking respondents for their own perception of their own systems' complexity. With the exception of presenting system parameters that may contribute to complexity, no definition of complexity was offered which allowed respondents to determine their own perception of their own systems' complexity based on their own assumptions of the complexity they experience in their manufacturing system. Specifically, the question posed was "How complex do you think your production system is? (For example, in relation to the number of products, the range of processes, product crossovers, variability, etc.)" with the possibility of answering "very", "somewhat", or "not very complex". The second element consisted of a set of questions pertaining to the complexity determinants. The information sought (Table 2) reflected the fact that a company's complexity can be affected by both internal (e.g. batch sizes) and external factors (e.g. customer demand) and covers the main determinant categories of complexity in manufacturing as presented in Table 1. In the questions for batch size, demand and production variability, the respondents could choose from a three point scale and each point was mapped directly back to a complexity level of "very", "somewhat", or "not very complex". The two questions related to SKUs were open-ended to allow respondents to describe their own particular case and therefore categorization was required prior to mapping to a complexity level. Once the response for each question was mapped to a complexity level, it was possible to compare the average calculated complexity level with the perceived complexity level as SME self-assessment.

Respondents were also asked questions under three other broad categories, company overview, IT system usage and DES awareness.

Table 2 *Criteria requested to gauge manufacturing SME complexity*

Criteria	Information sought from survey respondents	Determinant
Batch size	Do batch sizes always remain the same, change somewhat or change constantly?	Products
Total SKUs	How many different SKUs exist in the system?	Products
Active SKUs	How many active SKUs are live at any one time?	Products
Customer demand	Does customer demand always remain the same, change somewhat or change constantly?	Market/Supply Chain
Production	Is production variability considered to be low, medium or high?	Process/Equipment/ People

The survey was distributed to all 500 Irish manufacturing SMEs on the developed distribution list and held open for 12 weeks. Companies were mostly contacted via email or, where no direct email was available, contact was made through the company contact form on their website. A follow up strategy was set in place whereby each company received a maximum of two further emails after the initial contact was made. Perhaps due to bad destination email boxes twelve invites did not successfully reach the companies. Out of the 488 emails that were successfully delivered, responses were received from 183 of the SMEs (37.5%). However, after initial review 111 fully completed surveys were found to be usable for the purpose of assessing complexity (22.5%). Out of the usable surveys, it is noted that over 85% of survey respondents occupied a "Director" level role, with a further 5% operating at "Senior Management" level, with the remaining 9.5% of respondents operating at a level below "Director" or "Senior Management" (with described roles such as: accounts personnel, administrators, secretaries, laboratory analysts).

4 RESULTS

An overview of the companies surveyed is provided in Table 3, where the time in business and scale in terms of turnover and number of employees is presented. As an indication of IT adoption behavior, company usage of cloud-based data storage is also charted. It can be seen that the majority were well established companies (greater than 10 years in business) with a broad range of annual turnovers (mostly less than 5 million euro per annum). Most companies were small in terms of staff numbers with the majority employing less than 50 people. Over 70% of the companies were yet to use cloud-based data storage indicating low adoption of new technology in line with the findings of Marasini et al. (2008) regarding IT adoption in SMEs. Results of the complexity self-assessment show that 54% of respondents identified their production system as "somewhat complex", 27% "not very complex" with 19% reporting "very complex" production processes.

Table 3 *Characteristics of Survey Respondents*

Time in Business		**Annual Turnover**		**Total Employees**		**Data Stored on Cloud**	
< 1 year	1%	< 1	32%	< 10	26%	None	71%
1-5 years	4%	1-5 million	38%	11-50	48%	Some	19%
6-10 years	6%	6-10 million	12%	51-100	13%	About Half	5%
> 10 years	83%	11-50 million	14%	101-150	6%	Most	4%
		> 50 million	4%	151-200	2%	All	1%
No response	6%	No response	2%	>200	5%		

Figure 1 presents the breakdown of survey responses to the questions on complexity determinants. For the three variability questions (batch size, demand and production), the offered response options of high, medium and low are taken as a proxy for "very", "somewhat", or "not very" complex. The SKU related questions were open-ended to allow respondents enter their approximated values. The intention at the outset of the survey was that responses would be categorized into three groupings

corresponding to the three complexity levels and then included in an overall complexity calculation. However on review of the survey responses (Table 3), it became apparent that a very large range of values existed (i.e. 1 to 300,000 for total SKUs and 1 to 15,000 for active SKUs) and that the relative impact of the number of SKUs on complexity was dependent on many factors. These factors include the distinction between individual products (and the consequent implications for machine changeover and processing times), the relative number and complexity of products across companies (i.e. a greater number of simple products versus fewer but more complicated products) and the Pareto breakdown of sales of active SKUs.

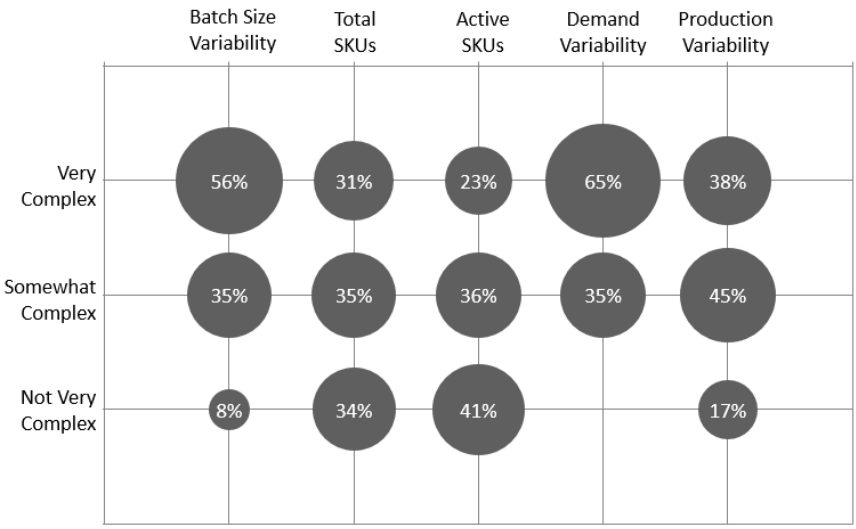


Figure 1 *Survey results for selected complexity determinants*

For these reasons, the SKU results have been omitted from the calculated average complexity level as presented in Table 4. This table shows that when based on calculated complexity relating to individual responses to the three variability questions (batch size, demand and production) a far greater number of respondent SMEs (174/328 – 53%) can be categorized as "very complex" versus 19% of SMEs that based on self-assessment would define themselves as "very complex".

Table 4: *Responses received to SKU related questions*

Complexity Level	Total SKUs	No.	%	Active SKUs	No.	%
Very	300,000	1	31%	15,000	1	22%
	10,000-20,000	4		6,000-8,000	3	
	1,000-9,000	21		1,000-5,000	14	
Somewhat	300-900	11	35%	200-500	19	38%
	25-273	19		25-150	12	
Not Very	0-20	29	34%	0-20	33	40%

Across all respondents, it was found that the majority (approximately 60%) of companies collect data in some relatively balanced combination of manual and automated processes. Interestingly, when respondents were grouped by their self-assessed levels of complexity (Figure 2), this position was far less pronounced in the case of companies perceived to be "very complex" than was the case in the other two cases (40% versus circa 65%). A greater percentage of these companies had mostly automated data collection (30% versus circa 15%) as would possibly be expected in complex environments. However, those with mostly manual data collection also accounted for 30% of the "very complex" cohort.

Table 5: *Calculated average complexity versus perceived complexity*

Complexity Level	Batch Size Variability	Production Variability	Demand Variability	Total Responses	Calculated Complexity	Self-Assessed Complexity
Very	62	42	70	174	**53%**	**19%**
Somewhat	39	49	38	126	**38%**	**54%**
Not Very	9	19	0	28	**9%**	**27%**
Total Responses	110	110	108	328		

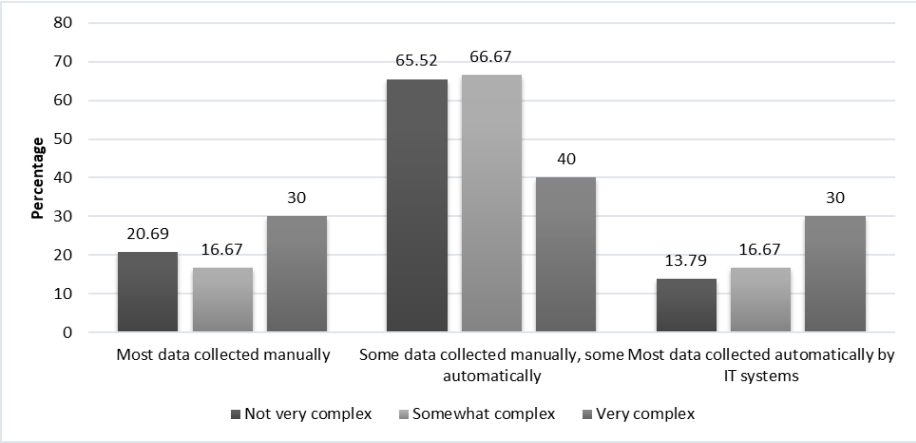


Figure 2 *Data Collection Practices presented against Self-Assessed System Complexity*

Of those surveyed only 21.4% of respondents were aware of DES. Figure 3 illustrates that there is a correlation between perceived complexity and awareness of DES with 35% in the "very complex" category being aware of DES as opposed to 10% and 22% of the "not very complex" and "somewhat complex" categories respectively.

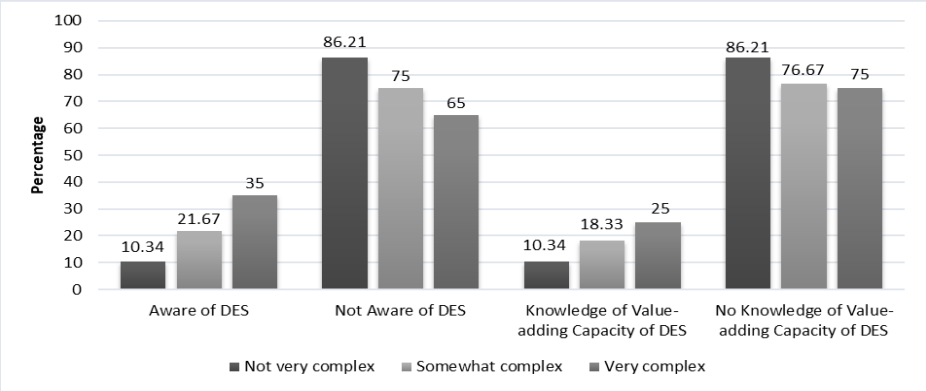


Figure 3 *Self-assessed 'awareness of DES' and 'knowledge of value adding capacity of DES'*

This correlation could be due to stakeholder motivation to find appropriate analytical methods or due to being more likely targets for vendors of such services. What speaks to the applicability of DES is that a similar trend holds true for respondent ability to identify where DES would add value to their business. As shown, the greater the complexity the more likely that the benefit of DES could be seen. As a percentage of those that were aware of DES in the first place, 70% of those in very complex environments could see how it would add value to their business.

5 DISCUSSION AND CONCLUSIONS

When questioned directly, a large majority of SMEs identify their business operations as medium or high complexity systems. This in itself would suggest that DES is an appropriate analytical method to apply in these settings as it is particularly suited to modelling complex systems (Robinson, 2005). However, when queried about specific aspects of their business, SMEs show higher levels of complexity with over half of respondents falling into the highest variability category. It is interesting to note that the SMEs surveyed tend to consistently underestimate their overall system complexity when compared to their own assessment of the systems individual components. As evidence of this, the assessment of the combined individual system components places 53% of the SMEs studied in the highest complexity category versus the self-assessment of only 19%. On the opposite side, the study found that the lowest category (Not Very), only 9% were calculated as being in this category when based on the systems individual components versus 27% when based on the respondents self-assessments. This aligns with Park and Okudan Kremer (2015) where they note that it is "very hard for general manufacturing companies to practically identify their current complexity levels at which they operate". Given that SMEs may not fully understand their complexity levels and thus underestimate the usefulness of DES to them, there is an even greater need to ensure DES platforms are presented in an SME usable format and that awareness and training is raised and provided by the SME community. The need for an appropriate pitch is echoed by Löfving et al. (2014) where they suggest that most manufacturing strategy frameworks are prescriptive and developed for larger companies while SMEs require more descriptive frameworks.

Variability is a key driver of uncertainty and one that is difficult to capture in the static spreadsheet analyses that are typically found across manufacturing SMEs. In DES, sources of variability (i.e. stochastic and dynamic system parameters) can be represented and the cumulative effect of these factors can be predicted. From an analysis of the results it is also clear that the levels of complexity experienced by the SMEs studied remain high in both the internal and external categories (Gabriel, (2008); Efthymiou et al., (2012); Blome et al., (2014)). Therefore, the appropriateness of DES to manufacturing SMEs may be even greater as the systems concerned are more complex than the stakeholders perceive. Furthermore, the highest level of variability was reported for customer demand. As an external complexity determinant, this is largely outside the control of the SME but may have an influence on internal complexity, as can be seen in the results for batch size variability. This relationship highlights the potential importance for companies to understand and predict how changes to in-house processes can influence overall performance. Indeed other authors also allude to this being the case (e.g. see Bozarth et al. (2009)) and future work would include performing a deeper analysis to determine the link between external and internal complexity through comparing individual cases. Based on these findings it would suggest that the systems under review are 'sufficiently complex' to warrant the use of DES for particular scenarios.

REFERENCES

Aelker J, Bauernhansl T and Ehm H (2013) Managing Complexity in Supply Chains: A Discussion of Current Approaches on the Example of the Semiconductor Industry. *Procedia CIRP* 7. Elsevier B.V.: 79–84.

Anderson P (1999) Perspective: Complexity Theory and Organization Science. *Organization Science* 10(3).

Beer S (1959) *Cybernetics and Management*. Wiley.

Blome C, Schoenherr T and Eckstein D (2014) The impact of knowledge transfer and complexity on supply chain flexibility: A knowledge-based view. *International Journal of Production Economics* 147(PART B). Elsevier: 307–316. DOI: 10.1016/j.ijpe.2013.02.028.

Bozarth CC, Warsing DP, Flynn BB, et al. (2009) The impact of supply chain complexity on manufacturing plant performance. *Journal of Operations Management* 27(1): 78–93.

Calinescu a, Efstathiou J, Schirn J, et al. (1998) Applying and assessing two methods for measuring complexity in manufacturing. *Journal of the Operational Research Society* 49(7): 723–733. DOI: 10.1057/palgrave.jors.2600554.

Chance F, Robinson J and Fowler J (1996) Supporting manufacturing with simulation: model design,

development, and deployment. *Proceedings of the 1996 Winter Simulation Conference*. DOI: 10.1109/WSC.1996.873268.

Choi TY, Dooley KJ and Rungtusanatham M (2001) Supply networks and complex adaptive systems: control versus emergence. *Journal of Operations Management* 19(3): 351–366.

Cicirelli F, Furfaro A and Nigro L (2011) Modelling and simulation of complex manufacturing systems using statechart-based actors. *Simulation Modelling Practice and Theory* 19(2). Elsevier B.V.: 685–703.

De Toni AF De, Nardini A, Nonino F, et al. (2001) Complexity measures in manufacturing systems. Bourgine P., Kepes F. SM (ed.) *Proceedings of the European Conference on Complex Systems - Towards a science of complex systems*: 1–20.

Deshmukh a. V, Talavage JJ and Barash MM (1992) Characteristics of part mix complexity measure for manufacturing systems. *Proceedings of the 1992 IEEE International Conference on Systems, Man, and Cybernetics*: 1384–1389. DOI: 10.1109/ICSMC.1992.271590.

Deshmukh A V, Talavage JJ and Barash MM (1998) Complexity in Manufacturing Systems Part 1 : Analysis of Static Complexity. *IIE Transactions* 30(7): 645–655. DOI: 10.1023/A:1007542328011.

Efthymiou K, Pagoropoulos a., Papakostas N, et al. (2012) Manufacturing systems complexity review: Challenges and outlook. *Procedia CIRP* 3(1): 644–649. DOI: 10.1016/j.procir.2012.07.110.

ElMaraghy W, ElMaraghy H, Tomiyama T, et al. (2012a) Complexity in engineering design and manufacturing. *CIRP Annals - Manufacturing Technology* 61(2). CIRP: 793–814. DOI: 10.1016/j.cirp.2012.05.001.

ElMaraghy W, ElMaraghy H, Tomiyama T, et al. (2012b) Complexity in engineering design and manufacturing. *CIRP Annals - Manufacturing Technology* 61(2). CIRP: 793–814. DOI: 10.1016/j.cirp.2012.05.001.

Fowler JW and Rose O (2004) Grand Challenges in Modeling and Simulation of Complex Manufacturing Systems. In: *Simulation*, 1 September 2004, pp. 469–476. DOI: 10.1177/0037549704044324.

Fredendall LD and Gabriel TJ (2003) Manufacturing Complexity : A Quantitative Measure. In: *Proceedings of POMS 14th Annual Conference*, Savannah, Georgia, 2003.

Frizelle G and Suhov Y (2008) The measurement of complexity in production and other commercial systems. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2008, pp. 2649–2668. DOI: 10.1098/rspa.2007.0275.

Frizelle G and Woodcock E (1995) Measuring complexity as an aid to developing operational strategy. *International Journal of Operations & Production Management* 15(5): 26–39. DOI: 10.1108/01443579510083640.

Gabriel (2008) Measuring the manufacturing complexity created by system design. In: *Proceedings of the 38th Annual Meeting of the Southeast Region of the Decision Sciences Institute (SEDSI)*, Orlando, Florida, USA, 2008.

Garzon AL, Arokiam A and Greig N (2012) Complexity of scheduling in SMEs specialising in High Variety Low value marketing. In: *KTP Associates Conference*, University of Brighton, 2012.

Götzfried M (2013) *Managing Complexity Induced by Product Variety in Manufacturing Companies. Complexity Evaluation and Integration in Decision-Making*. University of St. Gallen.

Ingemansson A, Bolmsjö GS and Harlin U (2002) A Survey of the Use of the Discrete-Event Simulation in Manufacturing Industry. In: *Proceedings of the 10th International Manufacturing Conference in China (IMCC2002)*, 2002.

Isik F (2010) An Entropy-Based Approach for Measuring Complexity in Supply Chains. *International Journal of Production Research* 12: 3681–3696.

Johansson M, Jahansson B, Leong S, et al. (2008) A real world pilot implementation of the core manufacturing simulation data model. In: *Proceedings of the Simulation Interoperability Standards Organization (SISO) Spring 2008 SIW Workshop*, 2008.

Lang M, Deflorin P, Dietl H, et al. (2014) The Impact of Complexity on Knowledge Transfer in Manufacturing Networks. *Production and Operations Management* 23(11): 1886–1898. DOI: 10.1111/poms.12193.

Lewin AY (1999) Application of Complexity Theory to Organization Science. *Organization Science* 10(3).

Löfving M, Säfsten K and Winroth M (2014) Manufacturing strategy frameworks suitable for SMEs. *Journal of Manufacturing Technology Management* 25(1): 7–26. DOI: 10.1108/JMTM-08-2012-0081.

Meyer MH and Curley KF (1995) The impact of knowledge and technology complexity on information systems development. *Expert Systems with Applications* 8(1): 111–134. DOI: 10.1016/0957-4174(94)E0003-D.

Miller JH and Page SE (2017) *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press.

Modrak V and Marton D (2012) Modelling and complexity assessment of assembly supply chain systems. *Procedia Engineering* 48: 428–435. DOI: 10.1016/j.proeng.2012.09.536.

Modrak V and Marton D (2013) Structural complexity of assembly supply chains: A theoretical framework. *Procedia CIRP* 7. Elsevier B.V.: 43–48. DOI: 10.1016/j.procir.2013.05.008.

Mönch L (2007) Simulation-based benchmarking of production control schemes for complex manufacturing systems. *Control Engineering Practice* 15(11): 1381–1393. DOI: 10.1016/j.conengprac.2006.05.010.

Park K and Okudan Kremer GE (2015) Assessment of static complexity in design and manufacturing of A product family and its impact on manufacturing performance. *International Journal of Production Economics*. Elsevier. DOI: 10.1016/j.ijpe.2015.07.036.

Pickering A (2004) The science of the unknowable: Stafford Beer's cybernetic informatics. *Kybernetes* 33(3/4): 499–521.

Pradhan S and Damodaran P (2009) Performance characterization of complex manufacturing systems with general distributions and job failures. *European Journal of Operational Research* 197(2). Elsevier B.V.: 588–598. DOI: 10.1016/j.ejor.2008.07.013.

Robinson S (2005) Discrete-event simulation: from the pioneers to the present, what next? *Journal of the Operational Research Society* 56(6): 619–629. DOI: 10.1057/palgrave.jors.2601864.

Serdarasan S (2013) A review of supply chain complexity drivers. *Computers and Industrial Engineering* 66(3). Elsevier Ltd: 533–540. DOI: 10.1016/j.cie.2012.12.008.

Simon HA (1962) The Architecture of Complexity. *Proceedings of the American Philosophical Society* 106(6): 467–482.

Zhang Z (2011) Modeling complexity of cellular manufacturing systems. *Applied Mathematical Modelling* 35(9). Elsevier Inc.: 4189–4195. DOI: 10.1016/j.apm.2011.02.044.

AUTHOR BIOGRAPHIES

DR JAMES BYRNE is an Assistant Professor in Operations and Analytics in DCU Business School, Ireland. In the past, he has worked with SAP Research (UK) as a Senior Research Scientist in the Technology Infrastructure domain, the Enterprise Research Centre in the University of Limerick. He has been co-PI and has lead work packages on a number large EU funded collaborative research projects including OPTIMIS (FP7), CACTOS (FP7) and RECAP (H2020).

DR PAUL LISTON is a Lecturer in Data Analytics at Athlone Institute of Technology. Paul previously held the position of Research Fellow at Dublin City University and undertook the roles of Global Commodity Manager and Supply Chain Consultant with Dell Computers. Prior to this, Paul worked with the Irish Centre for Manufacturing Research in the area of process improvement with companies including Analog Devices, Bombardier, Boston Scientific, Seagate, HP, Intel and Pfizer.

PROF PJ BYRNE is a Professor of Operations Management in Dublin City University Business School. Prior to joining DCU he worked in the University of Limerick from 2004 - 2007. He has an extensive track record in industrial based research and was a founding member of the Modelling and Simulation Research Group in DCUBS. He has been a PI for a number of international research projects in the areas of DES and its application in manufacturing, services and healthcare.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.030

DECISION SUPPORT SYSTEM WITH SIMULATION-BASED OPTIMIZATION FOR HEALTHCARE CAPACITY PLANNING

Canan Gunes Corlu
John Maleyeff
Chenshu Yang
Tianhuai Ma
Yanting Shen

Metropolitan College
Boston University
1010 Commonwealth Avenue
Boston, MA, 02215 USA

ABSTRACT

Capacity management of hospital staff and other resources is an important challenge faced by a healthcare administrator. Because of the variation in service times and the inability to inventory services, capacity buffers are required to ensure reasonable waiting times for patients. The nonlinear relationship between resource utilization and patient wait times makes it difficult to determine the optimal capacity buffer, called the knee. This work concerns the development of a decision support system using Python to determine optimal capacity buffers using a Monte Carlo simulation and knee optimization model that allows for flexibility in specifying uncertain arrival patterns and service times. Key factors relating to the system's size, amount of service time variation, and arrival patterns are shown to affect optimal buffer sizes. The system shows users their current status and where changes need to be made to the service times or the number of servers to achieve optimal results.

Keywords:

simulation, optimization, healthcare, capacity planning

1 INTRODUCTION

When asked why he engaged in research on the mathematics of queues, John D.C. Little, the famed operations research pioneer and MIT professor, stated that one day he discovered "queues are everywhere." Indeed, all healthcare professionals are well aware of queues - usually in the form of patients waiting in a hospital's emergency department (ED), their physician's office, or at a pharmacy. In fact, there is potential for a queue associated with every resource a healthcare system employs to serve patients. In an MRI facility, for example, there may be a queue for parking, a queue for the elevator, a queue for checking in, a queue for the device, a queue for the technician, a queue for image interpretation, and a queue for payment. Other queues are hidden from obvious view, such as physicians waiting for blood test results, patients waiting for a call from their physician, or others waiting on hold for a call center representative.

When planning capacity, healthcare administrators may allocate resources to processes at levels somewhat higher than the demand forecast in order to provide effective customer service. Although *Little's Law* (which states that the average waiting time is equal to the ratio of the number of customers in queue and the service rate) is known to many practitioners, they do not always appreciate the non-linear relationship between server utilization (the average percentage of time spent serving customers) and customer wait times. Figure 1 shows a generic example that is applicable to any queuing system. As the server utilization increases, waiting time will increase in a pattern commonly known as a *hockey stick*.

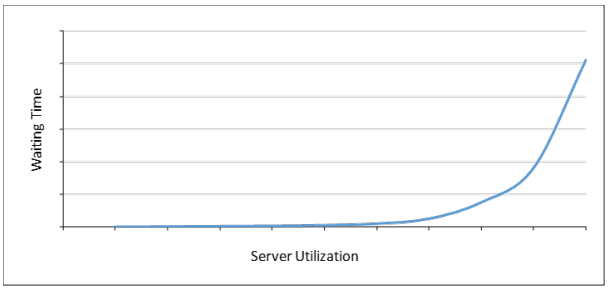


Figure 1: Server Utilization versus Waiting Time "Hockey-Stick."

The hockey-stick phenomenon has great implications for planners who wish to most effectively utilize their resources. Ideally, a capacity plan will effectively balance the needs of the planner (i.e., by maximizing server utilization) and the needs of the patient (by minimizing waiting times). In some fields, most notable computer science, the optimal system configuration takes place at a threshold referred to as the *knee* of the curve. Visually, the knee would be positioned at the point just before the slope of the hockey-stick curve significantly increases. The knee represents the optimal server utilization and capacity buffer (i.e., a 92% knee corresponds to a 8% capacity buffer). The goal of this paper is to propose a decision support system (DSS) for use by capacity planners to identify the location of the knee. The targeted application would be any multitude of queues found in healthcare settings. Queuing theory has been used to analyze queuing systems, but the restrictive assumptions associated with well-known queuing models preclude its effective use in many real-world settings.

This paper is organized as follows. Literature pertaining to healthcare capacity planning and queuing theory is reviewed, along with the challenges of embedding a Monte Carlo simulation (MCS) into a DSS. The methodology is then described, including the approach to finding the optimal capacity buffer (i.e., the knee). Then, the simulation model is described and used to evaluate a wide range of queues that could represent most queuing systems encountered by a healthcare administrator. Important results are discussed focusing on buffer recommendations for systems with various sizes and patterns of variation. Finally, the development and use of the DSS is described.

2 LITERATURE SEARCH

Congested systems (i.e., those with over-utilized servers and long customer wait times) have a negative impact on profit and customer satisfaction in many industries, including restaurants (Jain and Ali 2016), call centers (Sze 1984), and hospitals (Camacho et al. 2006). Medical professionals experience stress that leads to lower efficiency in congested systems (Sze 1984). Patients' expectations also increase in proportion to the time spent waiting for a service (Grossman 1972). And, health conditions of patients deteriorate as waiting times increase (Schulz 2017).

In healthcare, capacity buffering has been employed to hedge against congestion when the number of patients increases (Terwiesch et al. 2011; Towers 2014). The balancing of factors when determining capacity buffers in light of the high costs of healthcare delivery has been discussed (Bittencourt et al. 2018). Creating capacity buffers can be achieved by adding more resources or by reducing service times. Service times can be decreased by incorporating new technologies or removing non-value-added activities from the service process (Nicolaou 2016).

The use of queuing theory in healthcare capacity planning is relatively recent (Vass and Szabo 2015; Patel 2015; Gonzalez-Horta et al. 2011). Queuing theory has been applied frequently in the emergency department (ED) where both patient arrivals and treatment times are subject to uncertainty (Laskowski et al. 2009; Wang et al. 2013). In fact, two of the top three challenges faced by ED managers are shortage of inpatient beds and long patient flow times (Statista 2016). MCS is used when queues are complex (Lee and Elcan 1996). For example, it has helped healthcare managers make better capital decisions (Kennedy 2009). The simulation approach is robust because it allows a user to change many parameters (Zilm et al. 2003) including setting up work shifts (Kang and Park 2015). Physical resources are important to consider, such as beds, because they are often the bottleneck resource in the ED (Schiff 2011).

A DSS can assist healthcare administrators by providing a means to implement consistent and reliable recommendations (Marcial et al. 2018). Because a DSS may be active (i.e., it makes the decision) or passive (i.e., it evaluates a decision), it can impact how administrators make decisions (Berner and La Lande 2007). The DSS should include pertinent information, a user interface, a mathematical or empirical model, and optimization criteria (Spooner 2016). It is especially valuable when it provides a mechanism to treat seemingly-different but similar problems using a structured approach that avoids having to *reinvent the wheel* (Koutsoukis and Mitra 2003). It works best when the DSS is developed by an integrated team of developers, practitioners, and users (Yasnoff and Miller 2003).

Models embedded in a DSS can take many forms, from qualitative rules-based checklists to artificial intelligence methodologies (Reyna et al. 2015). Stochastic (i.e., probability-based) decision models have been employed (Maleyeff et al. 2004), including those that address designing a clinical testing system (Benneyan and Kaminsky 1996) and those that evaluate images with signal detection models (Lynn and Barrett 2014). Kadri et al. (2014) use a model that simulates transition between various states of the ED. Embedding a MCS into a DSS can be difficult because simulations tend to be case specific and therefore cannot always be relied upon to apply in more general settings (Hertz et al. 2014). When using this approach, the system developer needs to be cognizant that output will vary randomly and therefore the DSS must include a proper balance of accuracy and execution time (Fanti et al. 2015). Some developers address output uncertainty indirectly by requiring users of the DSS to input the number of iterations (Yu et al. 2019). The contribution represented by the application presented here seeks to determine the optimal capacity buffer based on specific parameters associated with a healthcare queuing system.

3 METHODOLOGY

The queuing system assumed here is robust with the following structure: (1) customers wait in a single queue to be served by multiple parallel servers, (2) the population of customers is infinite, (3) customers arrive according to an assumed deterministic or random pattern, (4) there is no limit to the queue size, (5) the queue employs a first-come first-served discipline, and (6) service time follows an assumed random pattern. The assumptions regarding customer arrivals enables the system to assume that appointments are made (deterministic, or scheduled arrivals) or that they arrive according to a Poisson process (random arrivals). The Poisson assumption is valid because customers mainly arrive independently of one another.

For the analysis below, service time distributions are right-skewed because most atypical patients require longer service times than typical patients. Treatment times in healthcare often follow similar patterns that can be effectively modeled with the gamma distribution (Millhiser and Veral 2019). The gamma distribution is right skewed and flexible based on the coefficient of variation (CV) of the service times. Figure 2 shows gamma distributions for various values of the CV. Its skewness is directly proportional to the CV (when the CV = 1.0, it is known as exponential distribution). In practice, the service time CV needs to be determined from the data collected at the facilities being modeled. The simulation described below can be easily adapted to other service time distributions.

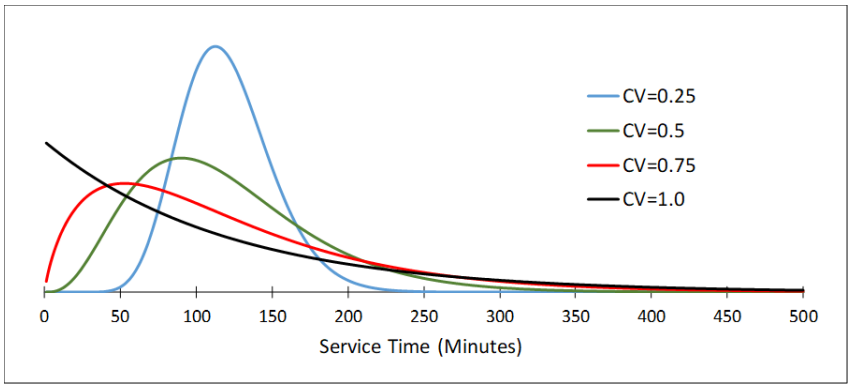


Figure 2: Gamma Distributions (Average Service Time is 120 Minutes).

The performance of any queuing system is determined by the server utilization (ρ), which is the ratio of the customer arrival rate (λ) to the system's service rate. The system's service rate is the product of the number of servers (s) and the service rate for each server (μ). Typically, performance is measured based on the time spent by customers in the system, the customer's wait time, the number of customers in the system, and the length of the queue.

3.1 Simulation Model

The MCS and knee-optimization algorithm were developed in Python, a programming language that is offered free of charge by the Python Software Foundation. The simulation logic mimics the queuing system by routing patients through a facility while keeping track of waiting times, total time spent in the system, and queue sizes. The inputs are the value of s, the CV of the gamma service time distribution, and the arrival pattern (scheduled or random). The MCS standardizes the queuing system to generate performance statistics based on the value of ρ. The model is applicable in a wide-range of applications without the need for re-programming the Python code.

The results obtained from the MCS were validated by using the analytical formulas of the M/M/s model (s parallel servers with exponential service times and Poisson arrivals). The MCS is run for a specified number of iterations, and a specified number of initial and terminal iterations dropped from consideration when calculating performance statistics. When employing a MCS, users need to be aware that output of the simulation will vary randomly. Therefore, the simulation is run for a specified number of macro-replications so that a statistical confidence interval (CI) can be calculated for each key result. The length of the CI is inversely proportional to the number of iterations and macro-replications, with run times taken into consideration for practical reasons.

3.2 Optimization Model

The knee (optimal server utilization) is employed by computer scientists to control network congestion. A popular approach was developed by Kleinrock (2018) who used a power function that identifies the level of server utilization that maximizes its *good* (i.e., server utilization) as compared to its *bad* (customer time in system). At various points of the hockey-stick function, power is calculated as the ratio of good to bad, using the following equation (where W_s is the average time a customer spends in the system):

$$P(G) = \frac{\rho}{\mu W_s}$$

In the power function, the numerator is equal to the server utilization and the denominator is equal to the normalized average time a customer spends in the system. It is normalized to have a minimum value of 1.0 for any combination of inputs and will fall in the range of 1.0 to 2.0 when the power is maximized. Therefore, it provides an unbiased scaling for power function calculation. The MCS finds the knee (optimal server utilization) by systematically changing the value of ρ from 40% to 95% (in increments of 5%) and simulating the system repeatedly over this range. This approach generates the information required to quantify the hockey-stick graph.

4 NOTEWORTHY OBSERVATIONS

A factorial experimental design was used to explore how the optimal server utilization (i.e., the knee) changed based on various levels of key variables. The MCS was run for 88 combinations of factors, corresponding to: (a) two levels of arrivals (scheduled or random); (b) four levels of service time CV (0.25, 0.5, 0.75, or 1); and (c) 11 levels for the number of servers. When determining the knee for each condition, 10 trials with 10000 iterations each were used to generate results for every level of server utilization. The 10000 iterations at each trial are extracted from the middle of 12000 runs.

4.1 Congestion and Instability

Although it is obvious that systems with higher server utilization exhibit longer wait times, the MCS showed that when the system is congested: (a) the variation of wait times increases as a percentage of the average wait time, and (b) wait times behaved erratically. Figure 3 shows the 95% CI for the

average wait time when 15 servers are employed, 6 customers per hour arrive randomly, and the service time CV is 50%. When $\rho = 0.80$ the average wait time was 8.6 minutes and its standard error was 0.29 minutes (a ratio of 3.4%). When $\rho = 0.95$ the average wait time was 94.7 minutes and its standard error was 8.93 minutes (a ratio of 9.4%).

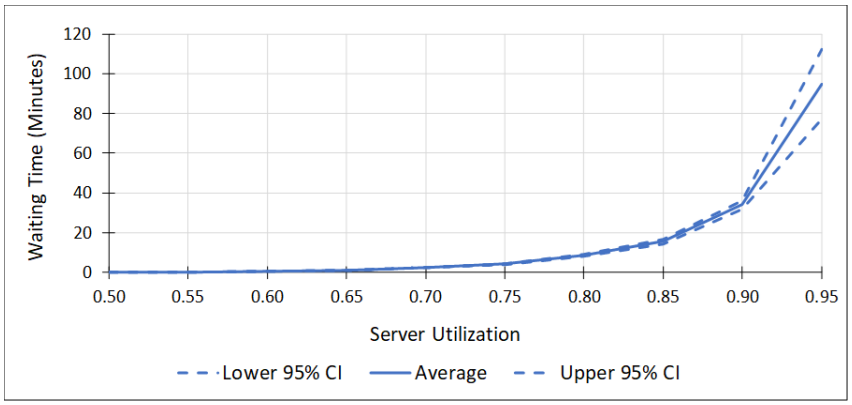


Figure 3: Example Confidence Intervals (s=15, CV=0.5, 6/hour random arrivals).

Although not shown here, the variation of the time spent in a congested system (e.g., $\rho = 0.95$) exhibited a great deal of instability due to the significant autocorrelation among patient wait times. The impact of autocorrelation across patients was much less evident when $\rho = 0.80$. There are implications for managers who would typically be unaware of the instability resulting from congested queuing systems. They may resort to *tampering*, which occurs when changes are made to a system in response to random variation, often causing the system to operate less effectively and impacting worker morale (Deming 1986, p. 327).

4.2 Relieving Congestion

The most obvious approach to reducing wait times would be the addition of servers. Capacity buffers would increase by hiring more staff or by placing clinical personnel "on call," both of which have financial implications. A less costly approach is to reduce service times. As an example, consider a system with 15 servers, where the arrival rate is 6 patients per hour and the service time CV is 50%. In this case, an average service time of 142.5 minutes results in $\rho = 95\%$. As shown in Figure 3, the average wait time is 94.7 minutes. If the average service time can be reduced from 142.5 minutes to 127.5 minutes (now $\rho = 85\%$), the average wait time decreases to 15.5 minutes. This is a somewhat counter-intuitive result for many practitioners that is consistent with the hockey-stick function. That is, a 15 minute reduction of average service times will reduce wait times by an average of 79.2 minutes.

The ability to reduce service times is situation-dependant. In an ED, for example, patients occupy a bed while they wait for a physician, various treatments, or to be discharged. In many cases, a significant portion of this time is not value-added. Consider the process used to test blood when ordered by a physician. The process, including drawing the blood, testing the blood, and evaluating the results, may take as little as 15-20 minutes. But, the total processing time can be several hours due to waiting for the technician to draw the blood, moving the blood to the lab, waiting for previous tests to be completed, setting up the testing equipment, entering the patient's information, waiting for a volunteer to transport the blood to the lab, and waiting for the physician to complete other activities before viewing the results. At one hospital, the service time was reduced by creating a process whereby the lab was notified when the blood testing order was written (rather than when the blood arrived) so that they could plan an efficient testing sequence.

4.3 Optimal Capacity Buffering

Table 1 shows optimal server utilization (i.e., the knee) for the 88 conditions evaluated with the MCS. As expected, the value of the knee is inversely proportional to the levels of service time variation. That is, the knee is higher for scheduled arrivals (which exhibit less variation than random arrivals), and the knee is higher for lower service time CV values.

		Queuing System Assumptions							
		Random Arrivals				Scheduled Arrivals			
	CV	25%	50%	75%	100%	25%	50%	75%	100%
Servers	1	0.60	0.55	0.50	0.50	0.85	0.75	0.65	0.55
	2	0.65	0.65	0.60	0.60	0.90	0.80	0.70	0.65
	3	0.70	0.70	0.70	0.60	0.90	0.85	0.80	0.75
	4	0.70	0.75	0.70	0.65	0.90	0.85	0.80	0.75
	5	0.75	0.75	0.70	0.65	0.90	0.85	0.80	0.75
	10	0.80	0.80	0.80	0.75	0.95	0.90	0.85	0.85
	15	0.85	0.80	0.85	0.80	0.95	0.90	0.90	0.90
	25	0.85	0.85	0.85	0.85	0.95	0.95	0.90	0.90
	50	0.90	0.90	0.90	0.90	0.95	0.95	0.95	0.90
	75	0.90	0.90	0.90	0.95	0.95	0.95	0.95	0.95
	100	0.95	0.95	0.90	0.95	0.95	0.95	0.95	0.95

Table 1: Optimal Knee Values.

Figure 4 illustrates how the Power function is used to determine the knee for 4 of the 88 experimental combinations. In this case, the knee is 55% (s = 1), 75% (s = 5), 80% (s = 15), and 90% (s = 75).

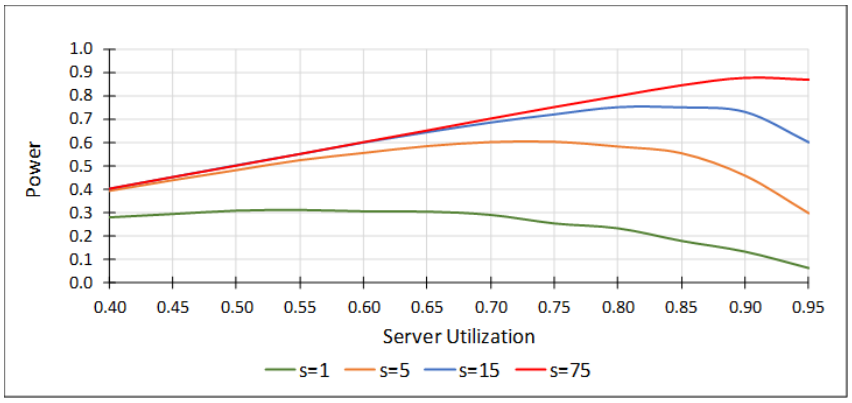


Figure 4: Utilization vs. Power Function (Random arrivals, CV=0.5).

An analysis of variance showed that each of the three independent variables (CV, s, and the pattern of arrivals), as well as their two-way interactions, affected the value of the knee (in all cases p-values were 0.001 or less). These relationships are summarized in Figure 5. This result illustrates the value in systems where patients arrive based on appointments, although it is not always possible. It also shows that optimal capacity buffers are differ depending on the amount of service time variation.

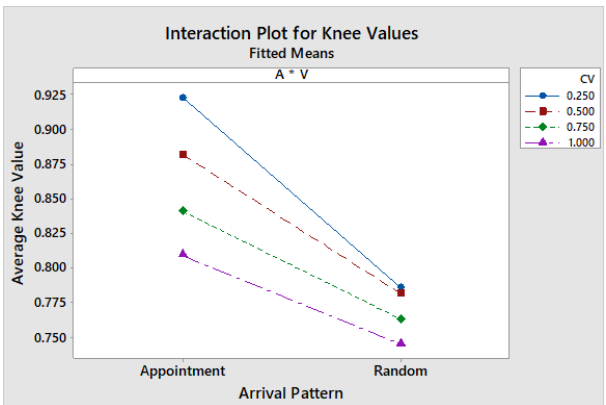


Figure 5: Knee Values versus Arrival Patterns.

5 DECISION SUPPORT SYSTEM

A DSS has been developed using the Python code. The DSS provides a comprehensive tool that evaluates the current capacity planning parameters and shows the optimal server utilization (i.e., the capacity buffer) based on the input parameters provided by the user. Figure 6 shows the user interface, where the user enters the average service rate, average service time, service time CV, the number of servers, and a choice of random or scheduled arrivals. The DSS begins by simulating the queuing system based on the user inputs, then it finds the knee using the approach described earlier. The program then calculates the changes required to move from the current ρ to the optimal ρ, either by changing the average service time or by changing the number of servers.

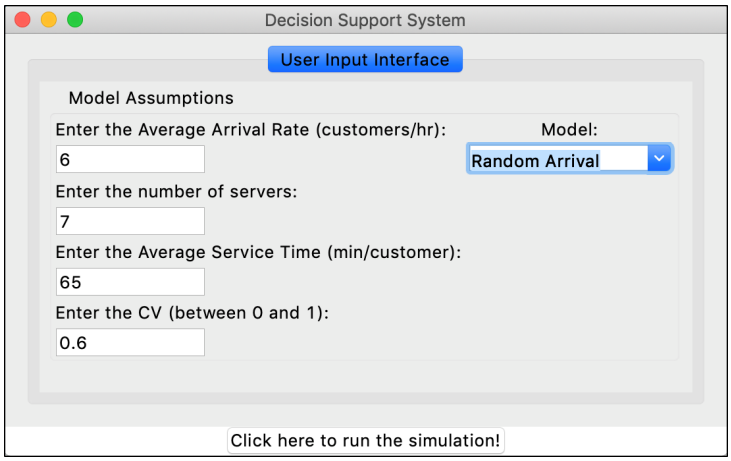


Figure 6: DSS User Interface.

Both tabular and graphical reports are provided as output. Figure 7(a) shows results with the current server utilization compared to results with a range of utilization values based on service time changes. Server utilization is 92.9% based on an average service time of 65 minutes, which results in an average time in system (W_s) of 135.8 minutes, an average wait time (W_q) of 70.8 minutes, an average patient population (L_s) of 14.1, and an average queue length (L_q) of 7.1 patients. At the optimal knee (75% server utilization) the average service time would be 52.5 minutes, which would result in an average time in system of 60.9 minutes, an average queue time of 8.4 minutes, and average patient population of 6.6, and an average queue length of 0.8 patients. The output shows these results graphically, using the hockey-stick format. Here, we see the current average service (65 minutes) with its corresponding wait time. The knee is located at the average service time of 52.5 minutes. That is, a 12.5 minute decrease in average service time would decrease the average wait for customers by about 62.4 minutes.

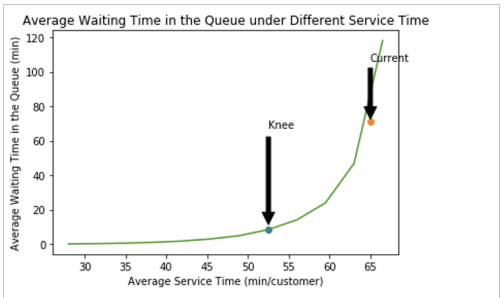
Figure 7(b) is an alternative analysis that focuses on changing the number of servers to optimize the configuration. The tabular output shows the results for the current number of servers ($s = 7$) along with various alternatives ($s = 8$ and $s = 9$). The graphical output shows that the optimal number of servers is 8.7 (calculated based on the knee of 75%). This value is rounded to create the recommendation of 9 servers. The results show that adding two servers reduces the average customer wait time by 66.7 minutes. The decision maker is not required to choose this number of servers. As the graphical output shows, a suitable alternative may be to employ 8 servers, which decreases the average customer wait time by 56.7 minutes.

```
Option 1 : Change the Service Time

Arrivals / Hour = 6
Servers = 7
  Util.      ST      Ws      Wq      Ls      Lq    Power
   0.40    28.0   28.12    0.14    3.32    0.01   0.3983
   0.45    31.5   31.80    0.30    3.68    0.03   0.4458
   0.50    35.0   35.53    0.57    4.05    0.06   0.4925
   0.55    38.5   39.51    1.02    4.44    0.10   0.5360
   0.60    42.0   43.79    1.79    4.88    0.18   0.5754
   0.65    45.5   48.57    3.03    5.35    0.30   0.6089
   0.70    49.0   53.88    4.92    5.88    0.49   0.6366
   0.75    52.5   60.91    8.43    6.60    0.85   0.6465
   0.80    56.0   70.12   14.09    7.53    1.41   0.6389
   0.85    59.5   83.43   23.92    8.86    2.40   0.6062
   0.90    63.0  109.89   46.85   11.52    4.70   0.5160
   0.95    66.5  184.53  118.07   18.95   11.83   0.3424

Knee:
  Util.      ST      Ws      Wq      Ls      Lq    Power
   0.75    52.5   60.91    8.43    6.60    0.85   0.6465

Current:
  Util.      ST      Ws      Wq      Ls      Lq    Power
 0.9286      65  135.79   70.79   14.09    7.08   0.4445
```

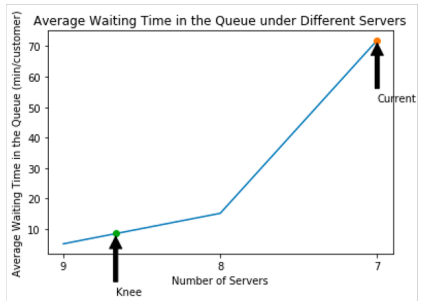



```
Suggestion: Decrease the average service time by 12.5 min/customer.
```

(a)

```
Option 2 : Change the Number of Servers

Arrivals / Hour = 6
Average Service Time / Minute = 65
Optimal Utilization = 0.75
 Servers   Util.      Ws      Wq      Ls      Lq    Power
       7   0.929  136.86   71.89   14.21    7.20   0.4410
       8   0.813   80.19   15.16    8.53    1.52   0.6586
       9   0.722   70.21    5.16    7.51    0.51   0.6686
```




```
Optimal servers = 8.7

Suggestion: Change the number of servers from 7 to 9.
```

(b)

Figure 7: **(a)** Output for Service Time Adjustment. **(b)** Output for Number of Services Adjustment.

6 CONCLUSION

The results of this study show that capacity planning in the presence of uncertainty cannot be done using simple rules of thumb or values that remain constant across situations. In a hospital, planning resource allocations in an ED would require different buffering levels than planning capacity in a small medical office, a large medical practice, or a hospital's call center.

A DSS is provided that is useful for capacity planners to determine suitable configurations for resource-limited facilities. The DSS incorporates simulation and optimization models that make robust real-world assumptions. Unlike simulations of specific facilities, the embedded simulation mimics queues in a general fashion making it applicable to a variety of situations without the need for re-programming.

REFERENCES

Benneyan, J. C., and F. C. Kaminsky. 1996. "Statistical and economic models for analysis and optimal design of laboratory screening policies for cervical cancer". *Annals of Operations Research* 67 (1): 235–285.

Berner, E. S., and T. J. La Lande. 2007. "Overview of clinical decision support systems". In *Clinical Decision Support Systems*, 3–22. Springer.

Bittencourt, O., V. Verter, and M. Yalovsky. 2018. "Hospital capacity management based on the queueing theory". *International Journal of Productivity and Performance Management* 67 (2): 224–238.

Camacho, F., R. Anderson, A. Safrit, A. S. Jones, and P. Hoffmann. 2006. "The relationship between patient's perceived waiting time and office-based practice satisfaction". *North Carolina Medical Journal* 67 (6): 409–413.

Deming, W. E. 1986. *Out of the Crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study.

Fanti, M. P., G. Iacobellis, W. Ukovich, V. Boschian, G. Georgoulas, and C. Stylios. 2015. "A simulation based Decision Support System for logistics management". *Journal of Computational Science* 10:86–96.

Gonzalez-Horta, F. A., R. A. Enriquez-Caldera, J. M. Ramirez-Cortes, J. Martínez-Carballido, and E. Buenfil-Alpuche. 2011. "Mathematical model for the optimal utilization percentile in M/M/1 systems: A contribution about knees in performance curves". *arXiv:1106.2380*.

Grossman, M. 1972. "On the concept of health capital and the demand for health". *Journal of Political Economy* 80 (2): 223–255.

Hertz, P., S. Cavalieri, G. R. Finke, A. Duchi, and P. Schönsleben. 2014. "A simulation-based decision support system for industrial field service network planning". *Simulation* 90 (1): 69–84.

Jain, P., and R. Ali. 2016. "A case study of take away restaurant using simulation modelling approach". *Journal of Service Science Research* 8 (2): 207–221.

Kadri, F., S. Chaabane, and C. Tahon. 2014. "A simulation-based decision support system to prevent and predict strain situations in emergency department systems". *Simulation Modelling Practice and Theory* 42:32–52.

Kang, S. W., and H. S. Park. 2015. "Emergency department visit volume variability". *Clinical and Experimental Emergency Medicine* 2 (3): 150.

Kennedy, M. H. 2009. "Simulation modeling for the health care manager". *The Health Care Manager* 28 (3): 246–252.

Kleinrock, L. 2018. "Internet congestion control using the power metric: Keep the pipe just full, but no fuller". *Ad Hoc Networks* 80:142–157.

Koutsoukis, N.-S., and G. Mitra. 2003. *Decision Modelling and Information Systems: The Information Value Chain*. Springer Science & Business Media.

Laskowski, M., R. D. McLeod, M. R. Friesen, B. W. Podaima, and A. S. Alfa. 2009. "Models of emergency departments for reducing patient waiting times". *PloS One* 4 (7): e6127.

Lee, Y., and A. Elcan. 1996. "Simulation modeling for process reengineering in the telecommunications industry". *Interfaces* 26 (3): 1–9.

Lynn, S. K., and L. F. Barrett. 2014. ""Utilizing" signal detection theory". *Psychological Science* 25 (9): 1663–1673.

Maleyeff, J., L. B. Newell, and F. C. Kaminsky. 2004. "Probability modeling applied to CAD systems for mammography". *International Journal of Health Care Quality Assurance* 17 (3): 125–134.

Marcial, L. H., J. E. Richardson, B. Lasater, B. Middleton, J. A. Osheroff, K. Kawamoto, J. S. Ancker, D. van Leeuwen, E. A. Lomotan, S. Al-Showk et al. 2018. "The imperative for patient-centered clinical decision support". *eGEMs* 6 (1).

Millhiser, W. P., and E. A. Veral. 2019. "A decision support system for real-time scheduling of multiple patient classes in outpatient services". *Health care management science* 22 (1): 180–195.

Nicolaou, L. M. 2016. "Quality improvement and leadership capacity development through lean methodology". *The American Journal of Accountable Care* 9 (16): 38–43.

Patel, N. M. 2015. "Half-latency rule for finding the knee of the latency curve". *ACM SIGMETRICS Performance Evaluation Review* 43 (2): 28–29.

Reyna, V. F., W. L. Nelson, P. K. Han, and M. P. Pignone. 2015. "Decision making and cancer.". *American Psychologist* 70 (2): 105.

Schiff, G. D. 2011. "System dynamics and dysfunctionalities: levers for overcoming emergency department overcrowding". *Academic Emergency Medicine* 18 (12): 1255–1261.

Schulz, M. 2017. "The intertwined relationship between patient education, hospital waiting times and hospital utilization". *Health Services Management Research* 30 (4): 213–218.

Spooner, S. A. 2016. "Mathematical foundations of decision support systems". In *Clinical Decision Support Systems*, 19–43. Springer.

Statista 2016. "Leading challenges for emergency departments in the US as of 2016". https://www.statista.com/statistics/631285/emergency-department-challenges-according-to-survey/. [Online; accessed 28-January-2020].

Sze, D. Y. 1984. "OR practice – a queueing model for telephone operator staffing". *Operations Research* 32 (2): 229–249.

Terwiesch, C., K. Diwas, and J. M. Kahn. 2011. "Working with capacity limitations: operations management in critical care". *Critical Care* 15 (4): 308.

Towers, T. J. 2014. "Pressure and performance: buffering capacity and the cyclical impact of accreditation inspections on risk-adjusted mortality". *Journal of Healthcare Management* 59 (5): 323–337.

Vass, H., and Z. K. Szabo. 2015. "Application of queuing model to patient flow in emergency department. Case study". *Procedia Economics and Finance* 32:479–487.

Wang, J., J. Li, and P. K. Howard. 2013. "A system model of work flow in the patient room of hospital emergency department". *Health Care Management Science* 16 (4): 341–351.

Yasnoff, W. A., and P. L. Miller. 2003. "Decision support and Expert Systems in public health". In *Public Health Informatics and Information Systems*, 494–512. Springer.

Yu, H.-H., K.-H. Chang, H.-W. Hsu, and R. Cuckler. 2019. "A Monte Carlo simulation-based decision support system for reliability analysis of Taiwan's power system: Framework and empirical study". *Energy* 178:252–262.

Zilm, F., K. Culp, and B. Dorney. 2003. "Virtual ambulatory care: computer simulation applications". *The Journal of Ambulatory Care Management* 26 (1): 7–21.

AUTHOR BIOGRAPHIES

CANAN G. CORLU is an Assistant Professor in the Administrative Sciences Department of Metropolitan College at Boston University. She holds a PhD in Operations Management from Carnegie Mellon University. Her email address is canan@bu.edu.

JOHN MALEYEFF is an Associate Professor of the Practice in the Administrative Sciences Department of Metropolitan College at Boston University. He holds a PhD in Industrial Engineering & Operations Research from the University of Massachusetts.

CHENSHU (CHARLOTTE) YANG holds an MS degree in Applied Business Analytics from Boston University.

TIANHUAI (JOSEPHINE) MA holds an MS degree in Applied Business Analytics from Boston University.

YANTING (EMMA) SHEN holds an MS degree in Financial Management from Boston University.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.031

SOME FORMULATION ISSUES IN CONSTRUCTING METAMODELS

Dr. Russell R. Barton

Durham University and The Pennsylvania State University
413 Business Building, University Park PA 16802
rbarton@psu.edu

ABSTRACT

Metamodels approximate the mapping of simulation input parameters to simulation outputs, providing a fast-running approximation, and (sometimes) insight on the nature of the response. Sometimes the obvious input variables and model structure give inferior metamodel fidelity. This paper illustrates three cases where the choice has significant impact on the fidelity and usefulness of the metamodel.

Keywords: Formulation, Model Parameters, Metamodels

1 INTRODUCTION

The simulation community has used metamodels to study the behaviour of computer simulations for more than fifty years (Burdick and Naylor, 1966) although the term was not employed until ten years later. The term for such 'models of a model' was used first by Blanning (1975), then by Kleijnen (1975), who popularized the term and made many significant contributions. Other disciplines refer to such models as response surface or surrogate models. For recent reviews, see Barton (2015) and Klejnen (2017).

There are numerous works devoted to model form and experiment design technique – again see the references above. But on some occasions, the usual approaches fail. In the sections below, three such cases are presented. Section 2 presents a scenario involving prediction of job completion time quantiles with a response function having nonhomogeneous variance. While the form of the Box-Cox variance stabilizing transformation is clear, employing the transformation increases nonlinearity and consequently model complexity. Section 3 focuses on metamodels for network simulations, where routing probabilities are the independent variable in the metamodel. Because each routing probability vector must sum to one, the design space is not full-dimensional. But the design space of the natural reformulation (leave one element of the vector out) has poor spatial structure as the number of components increases. Two alternative formulations have better structure. Section 4 presents more detail on an issue raised in Barton (2005). When seeking a collection of metamodels that are invertible, the choice of responses affects invertibility. The paper concludes with some remarks on awareness of model formulation issues when constructing simulation metamodels.

2 WHEN A BOX-COX TRANSFORMATION CAN FAIL

Pedrielli and Barton (2019) examined metamodels for quantiles of job completion time distributions as a function of jobs awaiting processing at each workstation upon release of a new job. Quantiles can be viewed as a guarantee of completion time. In this setting, the new job completion time increases as a function of the number of jobs waiting, and the independent processing times of jobs implies that the variance of completion time also increases as a function of the number of jobs waiting.

There are a number of transformations that can be used for variance stabilizing purposes and to improve the analysis. The family of power transformations proposed by Box and Cox (1964) are of the form:

$$y^{\lambda} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases},$$

where λ typically takes on values of -1 (reciprocal), 0 (log), 1/2 (square root), and 2 (square). Statistical software can estimate the value of λ for the Box-Cox transformation by the method of maximum likelihood. More simply, a variance stabilizing transformation can be selected via a plot of log standard deviation of response vs. log mean, assuming that there are multiple replications for each setting for the independent variables. See Chapter 3 of Montgomery (2017) for a detailed discussion of variance-stabilizing transformations.

Barton (2015) illustrated the advantage of a Box-Cox transformation for a simple queueing simulation with response as average waiting time as a function of mean service time. A log transformation of waiting time reduced heteroscedasticity of the response and also reduced the nonlinearity of the model, improving the R^2 of a fitted quadratic metamodel from 82.7% to 88.9%.

The response function for the job completion time quantile has a different structure that does not share this synergistic characteristic. The nonlinearity in the response arises primarily from the heteroscedastic variance; the mean completion time is a linear function of the jobs awaiting processing ahead of the new job when processing times are exponential, and approaches linearity for other distributions as the number of queued jobs increases. As a result, the impact of the quantile nonlinearity can be small relative to the linear relationship of the mean.

Figure 1 shows residual plots for quadratic regression for untransformed and transformed responses.

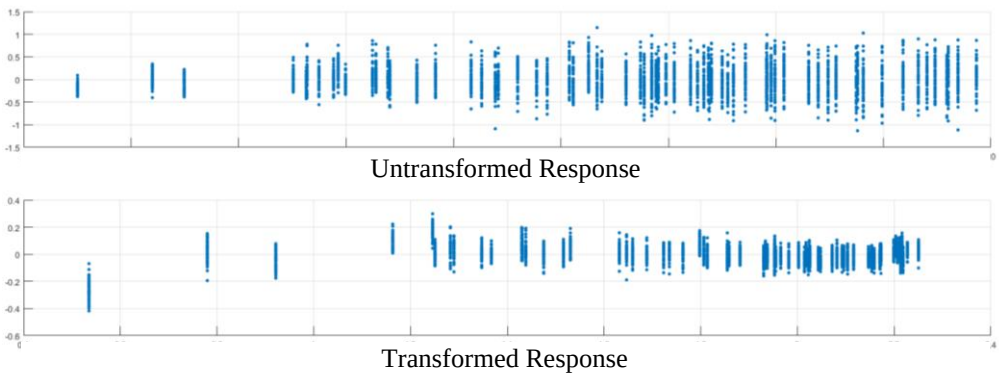


Figure 1 *Residuals for Quadratic Metamodel of Completion Time Quantile*

The results show the transform-induced nonlinearity at the lower end of the scale for the residuals of the transformed response metamodel – beyond what a quadratic function could capture. While the transformation reduced heteroscedasticity of the response, the predictive ability (over randomly selected points in the design space) of the quadratic metamodel was reduced from that of the metamodel of the untransformed completion time quantiles.

3 MODELING ROUTING PROBABILITIES

Barton (2005) developed metamodels for a network design simulation. For these metamodels, routing probabilities were the independent variables. For the general case, a model (and the corresponding metamodel) would have one (outgoing) routing probability vector for each node. A probability vector for a particular node, say i^{th}, might be represented by the vector whose elements were the set $\{P_{ij}\}$, where P_{ij} is the probability of routing to node j upon leaving node i. We will consider the vector for a single node, and so drop the i subscript in what follows. In Barton's network design simulation, there was a single routing node, and probabilities $\{P_j;\ j = 1, 2, 3\}$ were parameterized as $\{P_1, P_2/(1-P_1)\}$.

The reduction in cardinality from three to two was necessary to provide a full-dimensional space for the metamodel independent variables, since $P_1+P_2+P_3 = 1$, implying that the space of values would lie on a two-dimensional hyperplane in 3-space. This is illustrated in Figure 2. This is a characteristic of any routing probability vector, so for metamodeling it is natural to reduce the dimension by one. The simplest and natural strategy is to drop one element of the probability vector from the metamodel, say the last, which results in a full-dimensional design space of *Pdim* – 1, where *Pdim* is the cardinality of $\{P_j\}$. The metamodel variables then are $\{P_1, P_2, \ldots, P_{Pdim-1}\}$.

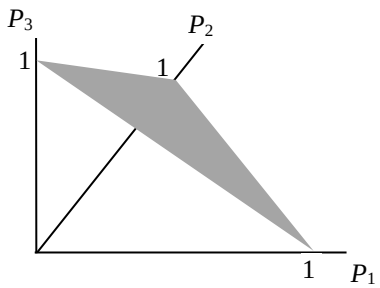


Figure 2 *Design Space (shaded) Falls on 2-D Hyperplane for Pdim = 3*

When *Pdim* is small, the parameterization 'All-But-1' works quite satisfactorily for metamodeling. For example, when *Pdim* = 3, one has the shaded space shown in Figure 3 for the design region. Note that the shaded region covers one half of the two-dimensional hypercube (e.g., square) $[0,1]^2$.

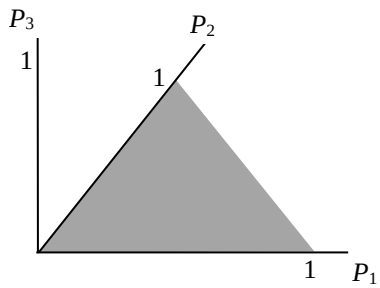


Figure 3 *Design Space (shaded) for All-But-1 with Pdim = 3*

While the geometry of this structure appears satisfactory, in high dimensions the corresponding shaded region captures an increasingly small (flattened) fraction of the *Pdim*-1 dimensional hypercube. This results in an ill-conditioned design space, leading to problems in fitting and using the resulting metamodel. This is shown in the computational examples below.

An alternative formulation is possible: embedding the shaded space in Figure 2 (a regular simplex) in *Pdim*-1 space, and representing each probability vector by its Cartesian coordinates in *Pdim*-1 space. This is an easy transformation: given the Cartesian coordinates of the simplex vertices, the Cartesian representation of the probability vector is just the convex combination of the coordinates of the simplex vertices, weighted by $\{P_j\}$. Vertices for a regular simplex with edge length = 1 were defined by Spendley, Hext and Himsworth (1962) for their optimization algorithm. One vertex is the zero vector, each of the other vertices has a representation $(q, q, \ldots, q, p, q, \ldots, q)$ where

$$p = \frac{1}{Pdim\sqrt{2}}\left((Pdim-1)+\sqrt{Pdim}\right) \text{ and } q = \frac{1}{Pdim\sqrt{2}}\left(\sqrt{Pdim}-1\right).$$

and the position of *p* varies from 1 to *Pdim*-1. For *Pdim* = 3, the corresponding simplex vertices are shown in Figure 4.

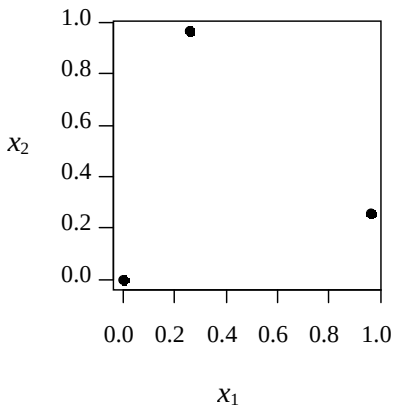


Figure 4 *Design Space within Vertices of Regular Simplex with Pdim = 3*

An alternative representation permits the dimension of parameter space to remain at *Pdim*, but relaxes the requirement that the probability vector sum to one. We call this a 'PseudoProbability' vector. The metamodel operates with this vector, while the simulation model uses the normalized version. The result is a response surface that extends the values found on the simplex along rays emanating from the origin. For values close to the origin, spatial variations occur more rapidly. In the computational comparison below, each *Pdim*-dimensional random vector in the design is multiplied by a (different) sample from U(0.5, 1.5).

3.1 A Computational Comparison

Consider a fitting design that is random, with design vectors computed as deviations from the vector with all values = 1/*Pdim*. Suppose that the deviations for each probability are uniform +/- .2/*Pdim* about the 1/*Pdim* nominal values. In the evaluations below, 1000 design points are generated. It is important to remember that we are generating vectors to use in an experiment design, and the (random) generation process is not the same as generating 1000 'empirical' probability vectors, each based on 1000 samples from the multinomial (*n*, *Pdim*, (1/*Pdim*, ..., 1/*Pdim*) distribution. To illustrate the contrast, for small *n*, empirical values more distant than +/- .2/*Pdim* from 1/*Pdim* are likely for some components of the probability vector. For large *n*, values as far away as +/- ε*Pdim* are increasingly unlikely for any fixed ε.

To compare the quality of the different parameterizations, we use a common measure of design quality, the covariance matrix of the design points (e.g. (*X*-*Xbar*)'(*X*-*Xbar*)), where *X* is the *numSamples* x *Pdim* or *Pdim*-1 matrix of probability vectors used to fit the metamodel and *Xbar* is a vector of average values elementwise for the probability vectors in the design. This is equivalent to *X'X* for a design matrix with each probability element centred about its mean across the design.

Figure 5 shows the eigenvalues of the covariance matrices for each parameterization for *Pdim* = 3 (on the left) and *Pdim* = 30 (on the right). The relative magnitudes are important, and indicate the balance of the design across the spatial dimensions of the probability vector. For a good design all would be approximately equal. The figure shows that using the full probability vector results in a zero eigenvalue, both for *Pdim* = 3 and 30. This indicates that the design falls in a *Pdim*-1 dimensional subspace, problematic for metamodels incorporating the full probability vector. For the *Pdim* = 3 case, all three alternative representations perform reasonably well. For *Pdim* = 30, only the Cartesian

representation is satisfactory: both the All-But-1 and the PseudoProbability representations each produce one eigenvalue that is near zero.

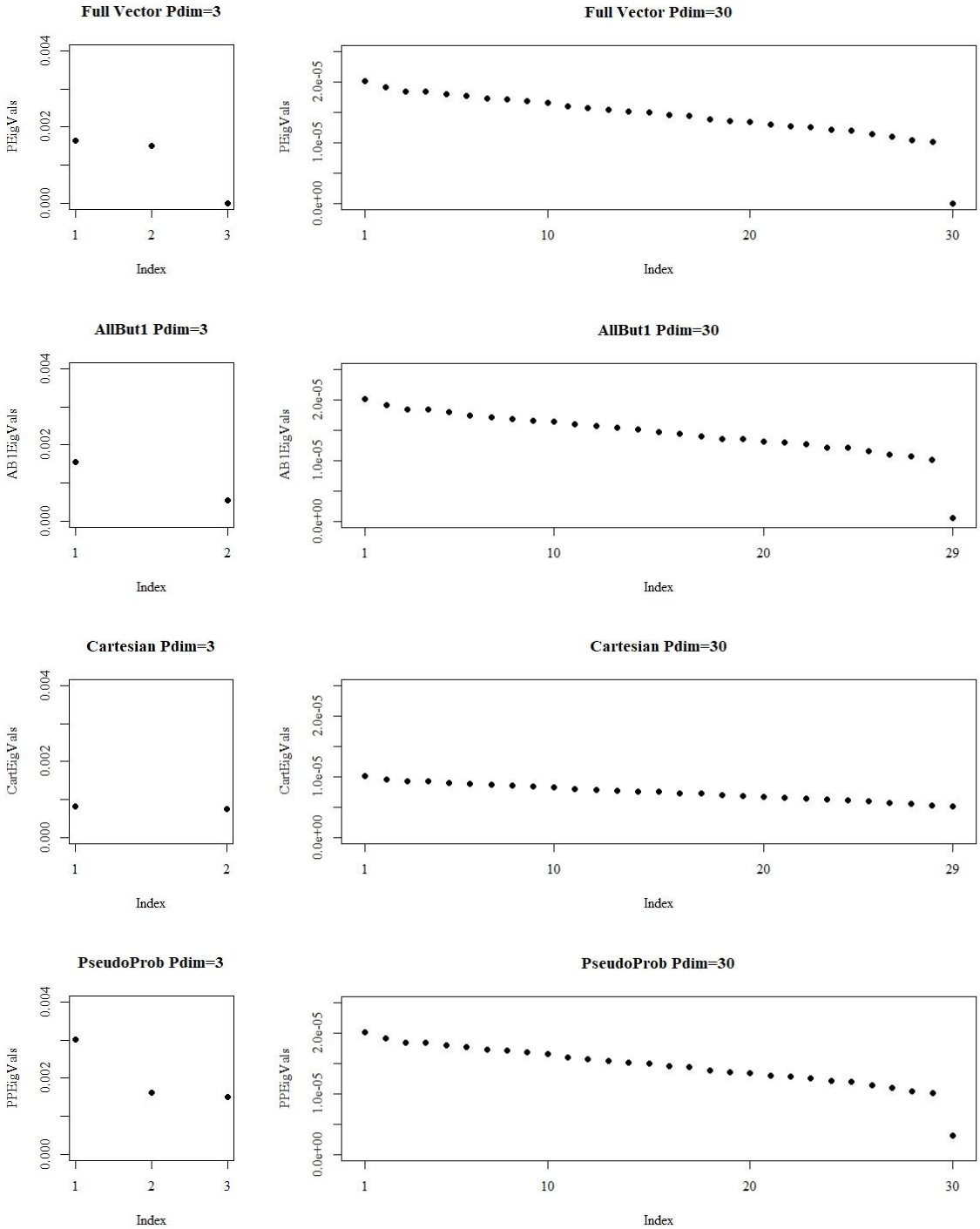


Figure 5 *Eigenvalue Plots for Each Probability Parameterization, Pdim = 3, 30*

Simulations of routing networks in manufacturing settings may only have a few possible routes from each machining node. But simulations of communications networks and large service operations may have tens or hundreds of routing possibilities at a particular node or point of service. As a consequence, it is important to use the better representation, i.e., the Cartesian representation, when metamodeling with large probability vectors.

4 MODELING FOR INVERTIBILITY

In customer-driven design of systems or products, one has performance targets in mind and would like to determine values for product or system parameters that meet such targets. Simulation models predict performance given design parameter values; meeting a target is done iteratively through an optimization search procedure, typically by optimizing a regression, neural network or other type of approximation metamodel of the computer model. Barton (2005, 2006) describes situations that allow inversion of the mapping. Consider a set of {X, Y} data, where X and Y are each n x k matrices, with a row of X corresponding to the k design parameter settings for a particular run, and the same row of Y corresponding to k output measures from one or more simulation runs. While a metamodel can be constructed for each column of Y as a function of X, it may also be possible to construct a metamodel for each column of X as a function of Y.

But the selection of the columns of Y, the coordinate functions of the mapping, can affect invertibility. Each coordinate function should be monotonic in the independent (x) variables. This means, for example, decomposing a total cost function into elements that vary monotonically with the elements of x. Total cost functions typically have a minimum which is sought, but often they are a sum of monotonic elements. Once can view this as how to parameterize the independent variables for the metamodels from y-space to x-space. The natural cost representation for the network example in Barton (2005) resulted in a mapping that was not invertible, and an alternative was proposed that led to an invertible mapping. Here we construct a simpler example (with analytic rather than simulation maps) to illustrate the issue.

Imagine making an investment decision on two projects, say 1, and 2, affecting different parts of a manufacturing operation. The level of investment for each project can range from 20 units to 65 units independently. Increasing investment reduces operations cost, expressed as a monthly quantity, but increasing investment brings on added debt service costs. Diminishing returns in delay cost savings are observed for both projects, and debt costs increase linearly in cost to first order, but with a second-order term increasing in the total debt. Figure 6 shows a grid of points in x-space, covering the range of investment choices (left side), and the corresponding images in y-space, when y_1 = monthly total cost for project 1 and y_2 = monthly total cost for project 2 (right side).

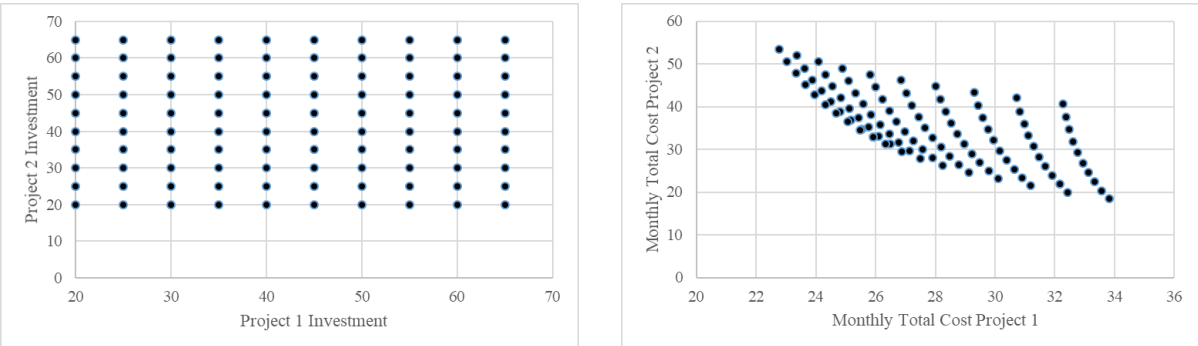


Figure 6 *Domain of Investment Decision Space and Corresponding Image with Project Costs*

The lower left corner of the cost plot dots in fact overlap. The mapping is not invertible in this area. This is because the total cost function for either project is not invertible. The cost decreases with increasing investment until the diminishing returns are overcome by increasing debt service costs. Instead, one can construct the coordinate functions to be total operations cost for projects 1 and 2, and total debt service costs for both projects. Figure 7 again shows the grid of points in x-space, covering the range of investment choices, but the corresponding images in y-space, are based on y_1 = monthly total operations cost for projects 1 and 2 and y_2 = monthly total debt service cost for projects 1 and 2.

Because both coordinate functions are monotonic, the mapping is invertible. Note that it is easy to determine total monthly costs from the contour lines on the figure. In this simple two-dimensional

example, the inverse map can be determined visually. In more complex settings, a fitted inverse metamodel would allow one to choose an investment decision to match any feasible cost profile.

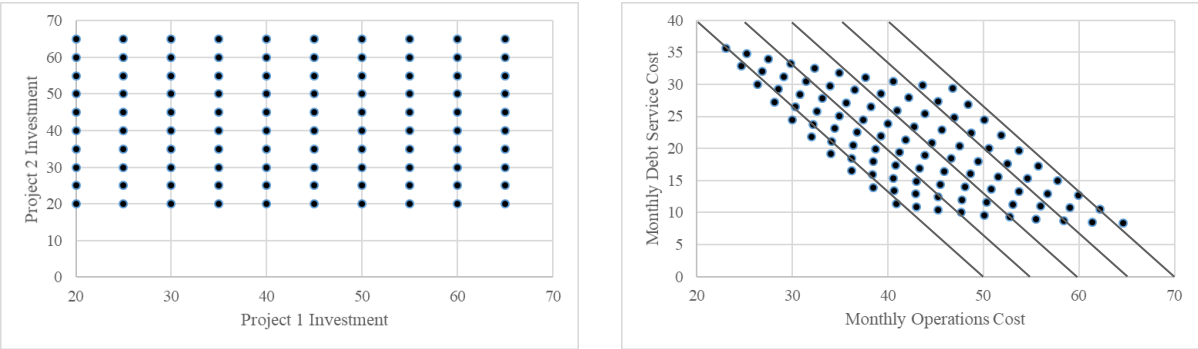


Figure 7 *Investment Decision Space and Corresponding Image, using Operations and Debt Costs*

5 SUMMARY

Researchers and practitioners interested in metamodeling typically focus on the choice of model type and on the experiment design for fitting the metamodel. In some cases, the success of the effort depends on how one chooses to represent the independent variables and dependent variables. First, when facing heterogeneous variance, which often simplifies the response as well. The example here shows that sometimes the response is already a simple function, and the VST can destroy that property. Second, metamodeling with routing probabilities is particularly prone to mis-parameterization; for more than a few dimensions, probability vectors should be cast in terms of their Cartesian coordinates on the probability simplex. Finally, when one is building a set of metamodels for invertible maps, a good choice of function decomposition can provide invertibility.

ACKNOWLEDGMENTS

I am grateful to my collaborators on metamodeling projects that led me to these findings. The parameterization issues in fitting metamodels of job completion quantiles arose during research with Giulia Pedrielli at Arizona State University. The parameterization issues in fitting metamodels using routing probabilities arose during research with Eunhye Song at The Pennsylvania State University and Henry Lam at Columbia University. Finally, I thank the reviewers for helpful suggestions to improve the clarity and quality of this paper.

REFERENCES

Barton, R R (2005). Issues in development of simultaneous forward-inverse metamodels. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M E Kuhl, N M Steiger, F B Armstrong, and J A Joines, 209-217. Piscataway, N.J.: Institute of Electronic and Electrical Engineers.

Barton, R R (2006). A two-phase maxi-min algorithm for forward-inverse experiment design. In *Proceedings of the 2006 Winter Simulation Conference*, ed. L F Perrone, F P Wieland, J Liu, B G Lawson, D M Nicol, and R M Fujimoto, 376-381. Piscataway, N.J.: Institute of Electronic and Electrical Engineers.

Barton, R R (2015). Introductory tutorial: simulation metamodeling. in *Proceedings of the 2015 Winter Simulation Conference*, ed. L Yilmaz, W K V Chan, I Moon, T M K Roeder, C Macal, and M D Rossetti, 1765-1779. Piscataway, N.J.: Institute of Electronic and Electrical Engineers.

Blanning, R W (1975). The construction and implementation of metamodels. *Simulation* **24**: 177-184.

Burdick, D S and Naylor, T H (1966). design of computer simulation experiments for industrial systems. *Communications of the ACM* **9**: 329–339.

Jacoby, J E and Harrison, S (1962). Multi-variable experimentation and simulation models. *Naval Research Logistics Quarterly* **9**: 121–136.

Kleijnen, J P C (1975). A comment on Blanning's metamodel for sensitivity analysis: the regression metamodel in simulation. *Interfaces*, **5**: 21–23.

Kleijnen, J P C (2017). Regression and kriging metamodels with their experimental designs in simulation: a review. *European Journal of Operational Research* **256**: 1-16.

Montgomery, D C (2017). *Design and Analysis of Experiments*. 9th ed., John Wiley & Sons, Hoboken, NJ.

Pedrielli, G and Barton, R R (2019). Metamodel-based quantile estimation for hedging control of manufacturing systems. In *Proceedings of the 2019 Winter Simulation Conference*, ed. N Mustafee, K-HG Bae, S Lazarova-Molnar, M Rabe, C Szabo, P Haas, and Y-J Son. Piscataway, N.J.: Institute of Electronic and Electrical Engineers.

Spendley, W, Hext, G R, and Himsworth, F R (1962). Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics* **4**: 441-461.

AUTHOR BIOGRAPHY

RUSSELL R. BARTON is Distinguished Professor of Supply Chain and Information systems and Professor of Industrial Engineering at the Pennsylvania State University. He is currently a Visiting Professor at Durham University Business School. He received a BS (Hons) in Electrical Engineering from Princeton University and MS and PhD in Operations Research from Cornell University. He is a fellow of IISE and a Certified Analytics Professional. His research interests include applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is rbarton@psu.edu.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.032

DAIRY SUPPLY CHAIN IN WEST JAVA: MODELLING USING AGENT-BASED SIMULATION AND REPORTING USING THE STRESS GUIDELINES

Dr. Bhakti Stephan Onggo

CORMSIS – Southampton Business School
University of Southampton, Southampton, UK
b.s.s.onggo@soton.ac.uk

Dr. Dhanan Sarwo Utomo

Centre for Sustainable Road Freight
Heriot Watt University, Edinburgh, UK
d.utomo@hw.ac.uk

ABSTRACT

Agent-based simulation (ABS) is one of the preferred methods to model supply chains, especially when we are interested in estimating the impact of the interactions between individuals on system level outcomes. This paper presents a case study in which ABS is used to model the dairy supply chain in West Java. The characteristics of the supply chain in this case study, such as, the existence of forage as a common resource and the dominance of smallholder farmers, are different from most dairy supply chains in higher income countries. It is more common in low-middle income countries. ABS is used because we are interested in analysing the behaviour of the farmers, in particular, their buying and selling decision rules, and their impact on cow population and milk production. This paper demonstrates how the STRESS guidelines can be used to report the simulation study.

Keywords: Dairy Supply Chain, Agriculture, Agent-Based Simulation, Simulation Reporting Tool

1 INTRODUCTION

Like most dairy supply chains, a dairy supply chain in Indonesia is typically formed by many tiers comprising farmers (producers), cooperatives (collector and handler), milk processing industries (manufactures), retailers and consumers as shown in Figure 1. Most farmers are smallholders with low production levels. Our survey on 153 farmer households in 19 villages in the West Java shows that 98% of them are smallholders (own fewer than eight cows) and 85% of them own less than 600 m^2 of land.

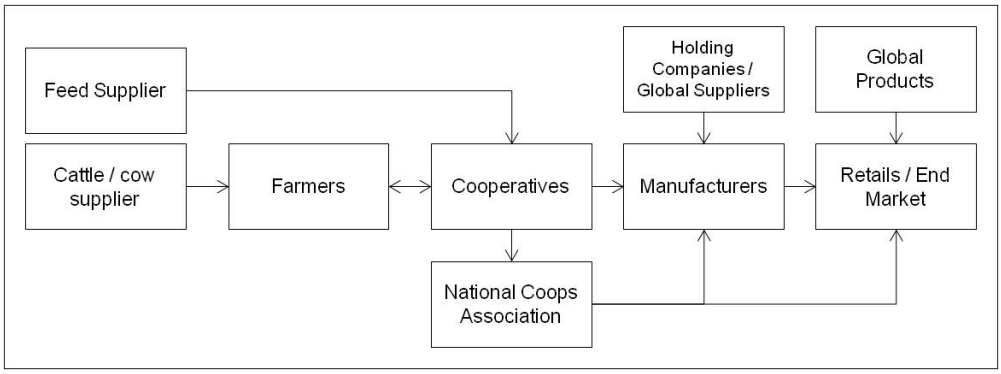


Figure 1 *A typical dairy supply chain in Indonesia (Daud et al., 2015)*

Milk is highly perishable so it must be transported efficiently and refrigerated at all times. This makes it prohibitively expensive for the smallholder farmers. Therefore, the role of a farmers' cooperative is important in transporting the milk from the farmers to the milk processors. It is also cheaper for the milk processing industries to buy milk in large quantities from cooperative than in

smaller quantities from farmers. We are interested in designing policy interventions to help smallholder farmers and dairy supply chain in West Java. Hence, the first objective of our research is to model the dyadic interaction between smallholder farmers and the cooperative using agent-based simulation (ABS). ABS is chosen because it is arguably the best tool to model the interactions between agents (or decision making entities) and to estimate their impact on system level behaviours (Onggo, 2016; Macal, 2016).

This paper reports an ABS project using the STRESS (Strengthening the Reporting of Empirical Simulation Studies) guidelines (Monks et al. 2019, available online since 2018). Hence, the second objective of this paper is to demonstrate the application of STRESS guidelines to ABS reporting. Monks et al. (2019) introduced the STRESS guidelines to address the reproducibility issue in simulation studies (and simulation projects in general). The guidelines aim to improve how we report a simulation study and hence, it should lead to better reproducibility of simulation studies. The guidelines provide three checklists, STRESS-ABS, STRESS-DES and STRESS-SD, for agent-based simulation, discrete-event simulation and system dynamics, respectively. Since the guidelines is relatively new, the number of examples reported in the literature is limited. As far as we are aware, there is only one example that has been reported, i.e. Taylor et al. (2018). Hence, this motivates us to evaluate this new ABS reporting tool.

2 DAIRY SUPPLY CHAIN IN WEST JAVA

The dairy supply chain in the case study area (i.e. Pangalengan, West Java) is one of the biggest in Indonesia. We consider it representative of other dairy supply chains in the country. Hence, some of our findings can be generalized to other dairy supply chains in Indonesia.

The population density of Java island is high (approximately 1121 people/km^2) so the price of land is expensive. Hence, most farmers own relatively small area of land that is only sufficient to build a pen for their cattle. The pens are usually located next to the farmers' houses in the middle of residential areas (Figure 2, left). The farmers cannot herd their cattle through the residential areas as it may cause conflict with the residents. Therefore, they must gather forage from outside of their villages for their cattle. Depending on their wealth, they usually transport the forage using carts or motorcycles (Figure 2, right). The forage grows along the road and river banks. We can view the forage as a common resource for all these farmers. Consequently, when the forage availability is low (e.g. prolonged drought), the competition between farmers to obtain forage becomes more intense. The government is interested in cow population and volume of production. Hence, our model will be used to estimate these outputs.

Figure 2 *A cattle pen in the middle of a residential area (left). A farmer is transporting forage using a cart (right).*

Based on the World Bank definition, Utomo et al. (2018) reviewed research into agri-food supply chain and found that most research took place in high income (58%) and middle income (37%) countries. This economic development categorization is important because actors from different economic development levels may behave differently as shown in our case study. Hence, more research is needed to improve agri-food supply chains (including dairy) in low middle income countries (LMIC). This is important because most farmers in LMIC are smallholders and are more vulnerable to external events such as change in prices (supplies, competing imported products, retail) and extreme weather. At the same time, their role cannot be easily replaced by large farming companies for reasons such as employment, local economy and social stability. Foraging in farming is also an interesting field to study as it is also relevant to high income and middle income countries where resources are shared (e.g. fish in international waters or disputed maritime borders, water for irrigation system) and it may result in disputes or conflicts (e.g. the cod wars, the Nile river dam row).

3 REPORTING THE MODEL USING STRESS GUIDELINES

Although the details differ, the three checklists provided by the STRESS guidelines are organised into the same six sections: objectives, logic, data, experimentation, implementation and code access. In what follows, we describe our model and the experimentation using STRESS-ABS checklist. We develop the model in close collaboration with subject matter experts (lecturers and graduates from the Animal Husbandry Department at a local university and a farmer).

3.1 Objectives

This is where we explain the background and rationale for the model, the model outputs and questions to be answered using the model (Table 1). In ABS, we are typically interested in the system level outputs that that emerge from the interactions between agents. The outputs can be qualitative (e.g. patterns) or quantitative.

Table 1 *Objectives*

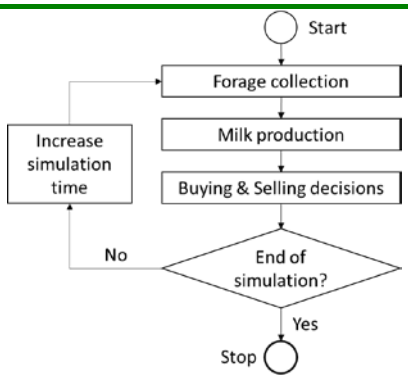
Section/Subsection	Item	Recommendation
1. Objectives		
Purpose of the model	1.1	The purpose of the model is to estimate the impacts of farmers' behaviours on the dynamics of milk production, and cow population in West Java Indonesia.
Model Outputs	1.2	The outputs of the model are the daily milk production volume (in litres) and the number of cows.
Experimentation Aims	1.3	The experimentation aim is to demonstrate how the hypothesized farmer's buying and selling decision rules affect the daily milk production volume and the number of cows.

3.2 Logic

In this section, we provide the model detail using suitable conceptual model representation (see Onggo 2010). If the experimentation involves scenarios that use multiple model, then we need to provide the detail for the models. In our case, we simply want to know the effect of the hypothesized buying and selling decision rules on model outputs. Hence, we do not compare scenarios.

Table 2 *Logic*

Section/Subsection	Item	Recommendation
2. Logic		
Base model overview diagram	2.1	The main sequence of the simulation model is shown in the flowchart below.



Base model logic 2.2 *Forage collection*

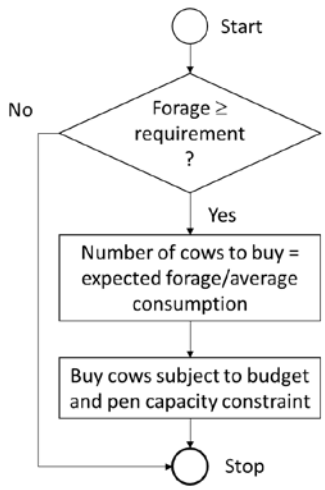
Farmers choose a reachable cell with the highest amount of forage and collect the forage (Martin et al 2016). The amount of forage collected is constrained by the capacity of the vehicles (cart, 40 kg; motorcycle, 60 kg; truck, 600 kg) and time spent for collecting forage (source: expert opinion).

Milk production (source: expert opinion):

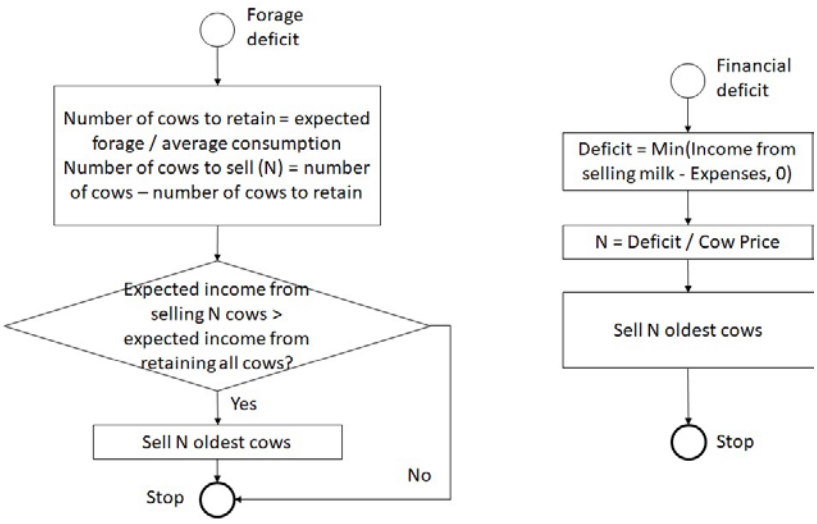
$$Qm_i = \begin{cases} MaxProd_i * ProdEff(NumPreg_i) * \overline{Forage_i} & , PregPeriod < 7\ month \\ 0 & , PregPeriod > 7\ month \end{cases}$$

Qm_i = the quantity of milk produced by cow *i* in a day.
$PregPeriod$ = pregnancy duration
$MaxProd_i$ = maximum milk production of cow *i* in a day.
$ProdEff$ = the production efficiency of cow *i* which is a function of the number of pregnancies ($NumPreg_i$). A cow achieves its maximum milk production after the second pregnancy ($ProdEff(2) = 100\%$) and the production efficiency then decreases linearly.
$\overline{Forage_i}$ = the average forage fulfilment of cow *i* (between 0 and 1, where 1 means that the given cow always obtains sufficient forage).

Buying decision (Gross et al. 2006):



Selling decisions (Boone et al. 2011, Gross et al. 2016):



Scenario Logic	2.3	Not applicable
Algorithm	2.4	Not applicable
Components	2.5	2.5.1 Environment

2D Grid containing 306 patches (each patch represents an area of one kilometre square). There are three types of patch:

- used patch: area occupied by building, houses, roads, etc
- unused patch: area that can be used to build new pens
- forage patch: area that are overgrown with forage

2.5.2 Agents

Agent Patch

Each time step, the amount of forage grows at rate (source: expert opinion and Bahar (2014)):

$$\frac{dF}{dt} = Min\big((F_{max} - F_t - Fc_t), (F_t - Fc_t) * (1 + G)\big)$$

F_t = initial forage level at day *t*. Fc_t= the amount of forage taken by the farmers on day *t*. F_{max} = the maximum amount of forage (kg) per kilometre square (uniformly distributed between 270 and 734 tonnes per km^2). *G* = the forage growth rate (1.1% per day).

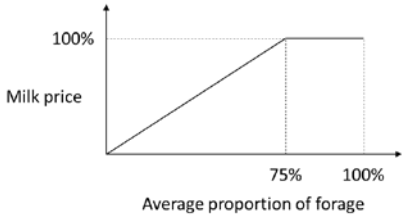
Agent farmer

Forage collection, buying and selling decisions: see item 2.2

Note: cows are treated as agent farmer's property (not as an agent)

Agent cooperative

Cooperative will always buy milk from farmers and the price is a linear function of milk quality and, in turn, is a linear function of the amount of forage fed to the cows.



2.5.3 Interaction Topology

Agent farmers collect forage from agent patches
Agent farmers interact indirectly with each other via environment (i.e. forage)
Agent farmers sells milk to agent cooperative

2.5.4 Entry / Exit

No new agents are created. Agent farmer leaves when s/he has no money left or cow left.

3.3 Data

The principle of Garbage-In-Garbage-Out suggests that the quality of data determines the quality of the simulation outcome, especially in empirical simulation studies. Data collection has been identified as one of the main issues in simulation projects (Onggo and Hill 2014) and modellers spend up to 40% of their project time dealing with data issues (Onggo et al. 2013). Hence, it is important to document the data and the data collection process.

Table 3 *Data*

Section/Subsection	Item	Recommendation
3. Data		
Data sources	3.1	Secondary data from BPS (2017), KBPS (2016), expert opinion Primary data from: • Interview with the stakeholders and expert was done to face validate the base model and to pilot test the survey instrument, • Close-ended and scenario-based questionnaire survey. The respondents are 153 farmer households in 19 villages in West Java. The data was collected in August 2016.
Pre-processing	3.2	Standard descriptive statistics and distribution fitting; The questionnaire used traditional measurement units. Therefore, we needed to convert them into international units.
Input parameters	3.3	Number of farmers, proportion of patches (item 2.5.1), run length, farmers' characteristics (e.g. age, number of cows), farmers' retirement age, prices (e.g. calf, cows, milk). The details are as follows: • Farmer agents' age in years: Triangular (22,74, 38) • The number of family labour per households: Binomial (0.92) • Number of cows own by each household (heads): Poisson (4.1) • Number of bulls own by each household (heads): Poisson (0.81) • The peak milk production of a cow (litre/day): Normal (20.81, 19.35) • Service per conception a cow (times): Poisson (2.38) • Average cow selling price (millions IDR/head): 13.1 • Average bull selling price (millions IDR/head): 16.4 • Average heifer buying price (millions IDR/head): 9.6

		• Minimum milk price (IDR/litre): 3350 • Maximum milk price (IDR/litre): 5200 • Additional fodder price (IDR/Kg): 2400
Assumptions	3.4	Farmers' home are spread randomly across the grid, retirement age is the same for all farmers, farmers work maximum of eight hours per day.

3.4 Experimentation

The settings in which the model is used to generate outputs are described in this section. This includes the warm up period for the non-terminating simulation, initial system state condition, run length and estimation approach (e.g. replications, batch means).

Table 4 *Experimentation*

Section/Subsection	Item	Recommendation
4. Experimentation		
Initialisation	4.1	Most input parameters (item 3.3) are set using the user interface (Figure 3); farmers' characteristics are sampled using the distributions obtained based on the primary and secondary data (item 3.1).
Run length	4.2	5 years
Estimation approach	4.3	The model is stochastic. Each scenario uses 25 simulation replications.

3.5 Implementation

This section provides information about the execution platform which is important due to the lack of backward compatibility in some software tools. This information is essential if the performance measures such as computation speed and memory requirement are needed (e.g. when we propose a faster algorithm than the existing one).

Table 5 *Implementation*

Section/Subsection	Item	Recommendation
5. Implementation		
Software or programming language	5.1	NetLogo 5
Random sampling	5.2	Built-in functions from NetLogo
Model execution	5.3	ABS model is using fixed time steps. NetLogo randomise the sequence of agents activation in each step to avoid bias.
System Specification	5.4	Not relevant, i.e. we do not measure computation speed

3.6 Code access

Open Science initiative aims to make research accessible to wider audience, especially publicly funded research. Taylor et al. (2017) discuss how Open Science principles applicable to simulation. In ABS community, the formation of CoMSES network (https://www.comses.net/) with the OpenABM platform is aligned with the Open Science principles. Hence, we expect that more simulation models will be accessible. Our code is not ready for public access at the time of writing. When it is ready, we will upload it to the OpenABM platform.

4 DISCUSSION

4.1 Simulation Results

The model is implemented using NetLogo (Wilensky 1999). The user interface is shown in Figure 3. The farmers are shown in red circles. The cooperative is shown in yellow square. The three patches, i.e. used, unused and forage, are shown in black, brown and green, respectively. Darker shades of green indicate higher amount of forage.

The simulation outputs, i.e. number of cows and daily milk production, are shown in Figure 4. Qualitatively, the simulation outputs (black lines) show the same downward trend observed in the real-world (dark grey lines). Quantitatively, the simulation outputs do not perform well as shown that the real-word data fall outside the 95% prediction intervals (shown in the dotted grey lines). For our purpose, since we are more interested in the medium-term planning (once every five years), the ability to reproduce the patterns is sufficient. This model would allow us to demonstrate the qualitative impact of interventions (such as subsidy) that can change farmers' behaviours on milk production and number of cows. However, if we want to use the model for short-term planning in which the accuracy of the estimates are important, we need to collect more data at the individual level. Indeed, further research (Utomo 2018) shows that the behaviours of farmers in West Java are different from the hypothesized behaviours used in the base model.

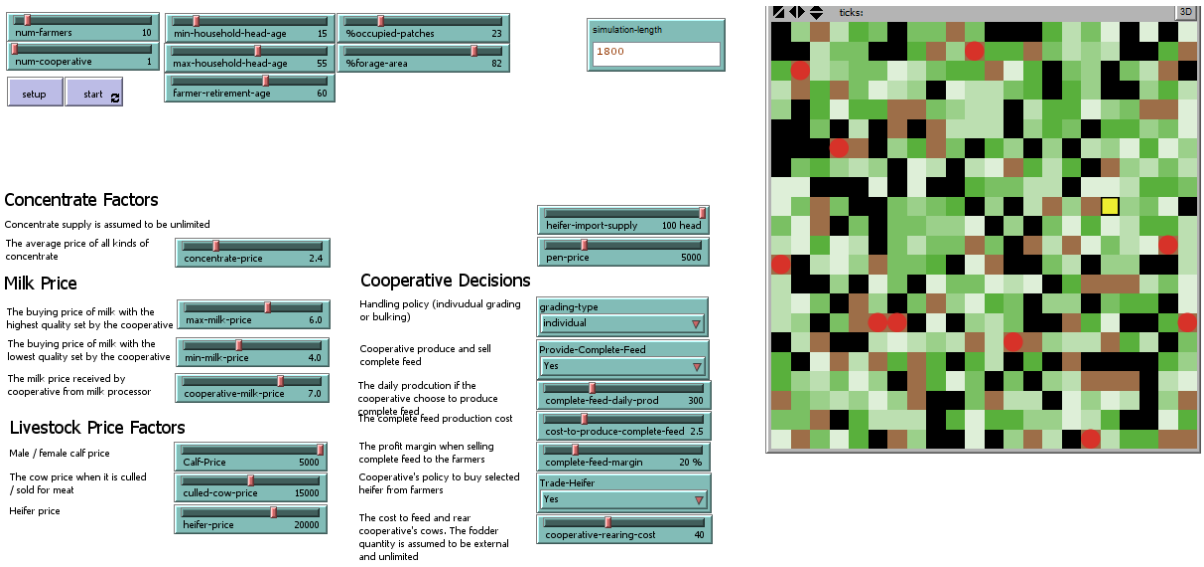


Figure 3 *Model user interface in NetLogo*

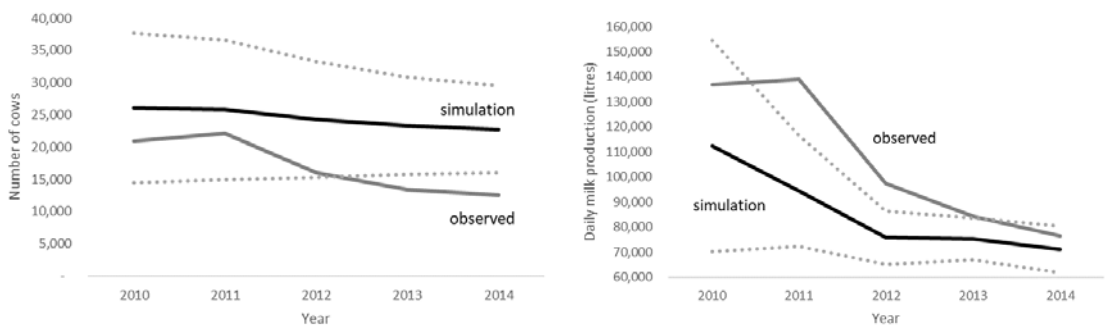


Figure 4 *Cow population (left) and daily milk production (right)*

This kind of ABS model is useful when we want to evaluate policy or intervention to change farmers' behaviours. The reason is that we can simulate the impact of a target behavior in comparison

to an existing behaviour on certain system level performance. The quantification of the estimated impact can complement a behavioural operations experiment which aims to find an effective intervention that can change existing behaviours.

4.2 Personal reflection on STRESS guidelines

Upon reflection, the fact that the STRESS guidelines are deliberately not prescriptive is useful. It makes the guidelines easier to use since we can concentrate on the check lists instead of how to write or format the document. The drawback of not being prescriptive is that we need to decide the right scope and level of detail that should be written in the document. The document should help us understand the computer simulation model or code but it should not explain all functions or variables. Hence, we still have to judge the scope (e.g. which functions and variables to include) and level of detail (e.g. block diagram, flow chart or detailed pseudocode). A too detailed report takes a lot of time to produce and may hinder the clarity of the report by focusing on the unnecessary details. There are details that are better documented in the computer model or code.

Related to the above is the issue of confidentiality. Detailed models such as the production lines in a factory can be confidential. Hence, it is not always practical to produce the report. However, we argue that it may be even more important to document such a model because the model is likely to be a high value asset. The issue of confidentiality can be mitigated by using a good security system that manages access to the report.

Another advantage is that the guidelines provides us with the list of most likely items that a reader needs to reproduce a simulation model and experiments using the model. This should help us minimise the number of missing important information.

5 CONCLUSION

This paper has applied STRESS guidelines to report a simulation study of a dairy supply chain in West Java. The case study shows that a dairy supply chain in LMIC may have different characteristics in comparison to dairy supply chain in high income countries, for example, the majority of farmers are smallholders and the presence of foraging behaviour. Hence, more research is needed to understand and improve the supply chain in LMIC. We have also implemented an ABS model to estimate the effect of farmers' behaviours on cow population and milk production. This model can be useful in estimating the impact of behavioural changes on population level performance such as milk production.

On reflection, the fact that the STRESS guidelines are deliberately not prescriptive is useful but at the same time, we need to decide the right scope and level of detail. The guidelines are especially useful in providing a list of commonly required information to understand a model. Like similar initiatives, as the guidelines become widely used and tested, we will know more about what needs to be improved.

REFERENCES

Bahar S (2014). Produktivitas hijauan pakan untuk produksi sapi potong di Sulawesi Selatan. JITV, 19. Available: https://docplayer.info/47163408-Produktivitas-hijauan-pakan-untuk-produksi-sapi-potong-di-sulawesi-selatan.html [Acessed 23 August 2019]

Boone R, Galvin K, Burnsilver S, Thornton P, Ojima D and Jawson J (2011). Using coupled simulation models to link pastoral decision making and ecosystem services. *Ecology and Society*, 16.

BPS (2017). Agricultural and Animal Husbandry Census [Online]. Available: https://www.bps.go.id/subject/24/peternakan.html#subjekViewTab3 [Accessed 22 December 2017]

Daud A, Putro U and Basri M (2015). Risks in milk supply chain: A preliminary analysis on smallholder dairy production. *Livestock Research for Rural Development*, **27**:137.

Gross J E, McAllister R R J, Abel N, Smith D M S and Maru Y (2006). Australian rangelands as complex adaptive systems: A conceptual model and preliminary results. *Environmental Modelling & Software*, **21**:1264-1272.

KPBS 2016. Data Populasi dan Penghasilan Anggota KPBS. Unpublished Dataset.

Macal C M (2016). Everything you need to know about agent-based modelling and simulation. *Journal of Simulation*, **10(2)**: 144–156.

Martin R, Linstädter A, Frank K & Müller B (2016). Livelihood security in face of drought – Assessing the vulnerability of pastoral households. *Environmental Modelling & Software*, **75**:414-423.

Monks T, Currie C, Onggo B S S, Robinson S, Kunc M, and Taylor S J E (2019). Strengthening The Reporting of Empirical Simulation Studies: Introducing the STRESS Guidelines. *Journal of Simulation*, **13(1)**: 55–67.

Onggo B S S (2010). Methods for Conceptual Model Representation. In Robinson S, Brooks R, Kotiadis K. and van der Zee D-J (eds). *Conceptual Modelling for Discrete-Event Simulation*. Taylor and Francis: Boca Raton, FL, pp 337–354.

Onggo B S S, Hill J, and Brooks R J (2013). A Pilot Survey on Data Identification and Collection in Simulation Projects. In: *Proceedings of the 27th European Simulation and Modeling*, Onggo B S S and Kavicka A (eds). Eurosis-ETI: Ostend, Belgium, pp 77–84.

Onggo B S S and Hill J (2014). Data Identification and Data Collection Methods in Simulation: A Case Study at ORH Ltd. *Journal of Simulation* **8(3)**:195–205.

Onggo B S S (2016). An introduction to agent-based simulation as a decision-support tool. In: Anagnostou A, Hoad K and Kunc M (eds). *Proceedings of the 8th Operation Research Society Simulation Workshop*. The Operational Research Society: Birmingham, UK, pp 1–12.

Taylor S J E, Anagnostou A, Fabiyi A, Currie C, Monks T, Barbera R, and Becker B (2017). Open Science: Approaches and Benefits for Modeling & Simulation. In *Proceedings of the 2017 Winter Simulation Conference*, Chan W.K.V. et al. (eds). IEEE: Piscataway, New Jersey, 535–549.

Taylor S J E, Anagnostou A, Monks T, Currie C, Onggo B S S, Kunc M and Robinson S L (2018). Applying the stress guidelines for reproducibility in modeling & simulation: Application to a disease modeling case study. In *Proceedings of the Winter Simulation Conference*. IEEE: Piscataway, New Jersey, pp 739–748.

Utomo D S, Onggo B S S, Eldridge S (2018). Applications of Agent-Based Simulation in the Agri-Food Supply Chains. *European Journal of Operational Research*, **269(3)**: 794-805

Utomo D S (2018). Improving and Comparing Data Collection Methodologies for Decision Rule Calibration in Agent Based Simulation. A Case Study of Dairy Supply Chain in Indonesia. PhD thesis. Department of Management Science, Lancaster University.

Wilensky U (1999). *NetLogo: Center for Connected Learning Computer-Based Modeling*. Northwestern University: Evanston, IL.

AUTHOR BIOGRAPHIES

BHAKTI STEPHAN ONGGO is an Associate Professor of Business Analytics at Southampton Business School, the University of Southampton. He is a member of the Centre of Operational Research, Management Sciences and Information Systems (CORMSIS). His research interests lie in the areas of simulation modelling methodology (symbiotic simulation, hybrid modelling, agent-based simulation, discrete-event simulation) with applications in operations and supply chain management as well as health and social care. He is the Associate Editor for the Journal of Simulation. His email address is b.s.s.onggo@soton.ac.uk. His website is https://bsonggo.wordpress.com

DHANAN SARWO UTOMO is a Research Associate in Centre for Sustainable Road Freight, Heriot-Watt University. He received his MS degree from Institut Teknologi Bandung, and earned his PhD from Lancaster University. His research interests are in computer simulation and modelling areas i.e., system dynamics and agent-based simulation. He is also interested in the application of computer simulation and modelling e.g., in agriculture, public sector, environment, and logistics. His email address is d.utomo@hw.ac.uk.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.033

USING MICROWORLDS FOR RESILIENCE MANAGEMENT OF FOOD SYSTEMS

Dr Hugo Herrera

University of Bergen, Department of Geography,
System Dynamics Group
Fosswinckelsgt. 6 5020 Bergen
hugojhdl@gmail.com

Dr Birgit Kopainsky

University of Bergen, Department of Geography,
System Dynamics Group
Fosswinckelsgt. 6 5020 Bergen
birgit.kopainsky@uib.no

ABSTRACT

Resilience management of food systems requires building understanding of the underlying drivers of systems adaptive capacity. We argue that system dynamics models and simulations can be used to inform discussions aiming to such resilience by providing interactive environments to experiment with different policies. We make this argument based on our experience analysing the key structures conditioning the response of beef farming systems in France to climate change effects. The analysis illustrate how small and aggregated system dynamics models can foster understanding about mechanisms driving resilience by facilitating discussions about its drivers, opportunities for intervention and their trade-offs.

Keywords: System dynamics, Microworlds, Resilience, Food Systems

1 INTRODUCTION

Food systems are socio-ecological systems (SES) in which a variety of stakeholders interact through a wide range of activities such as production, packaging, selling and consumption of food (Ericksen, 2008). The objectives for food systems include long-term sustainability of food security and social and environmental outcomes (Ericksen, 2008). A prerequisite for long-term sustainability is the capacity of a system to maintain its functionality without compromising its ability to do so in the future. There is an increased awareness of the vulnerabilities of food systems to changes in the environment like those introduced by climate change (e.g. water scarcity, weather variability) (Campbell et al, 2016; Tendall et al, 2015). This is particularly the case in Europe, where food systems are experiencing changes in the technological, demographic, environmental and economic landscape (Knickel et al, 2016, Saifi and Drake, 2008).

Socio-ecological resilience is essentially understood as a system ability to maintain its functionality even when it is being affected by a disturbance (Folke et al, 2010; Holling, 1996). While sustainability provides a framework for long-term planning, resilience focuses on adaptive mechanisms that will support a system functionality in the medium and long-term future. The emphasis on adaptive mechanisms to unpredictable changes has made resilience a compelling forward-looking approach to adaptation (Pizzo, 2015) attracting the attention of researchers and policymakers.

While resilience is a characteristic of the system, resilience management is the active modification of a system with the explicit aim to improve its capacity to absorb and adapt to change (Nettier et al, 2017; Fath et al, 2015; Walker et al, 2002). These capacities depend on the way the system has been organised and, therefore, resilience management is interested in understanding such organisation and identifying more effective ways for structuring the system.

According to Walker et al (2002: 14), the aims of resilience management are a) to prevent a system from transitioning into undesirable configurations in the face of external shocks and b) to cultivate the conditions that facilitate system adaptability following a massive change. Undesirable

configurations can be operationalised as the set of conditions and relations within the system that prevent it from providing its main outcomes. For instance, in the case of food systems, undesirable configurations are those that prevent them from providing: enough, affordable and good quality food; sustainable revenues for farmers; or enough jobs in their communities.

The second aim proposed by Walker et al (2002: 14) suggests that a resilience management process is not a normative process, but a structured and systematic framework that allows stakeholders to adapt to challenges in the environment (Nettier et al, 2017; Holling & Gunderson, 2002). In this paper we work towards this aim by focusing on means to analyse and communicate the mechanisms driving resilience of food systems (Sieber et al, 2018; Biesbroek, et al., 2014).

Namely, we illustrate how small models can inform our understanding of the system structure and the mechanisms driving the system responses to external shocks. Next, we present the analysis and insights gained from using a small system dynamics model to explore resilience of food systems in France. The paper proceeds as follows, first, we start by describing what are microworlds for resilience management. Next, we briefly describe the case study, the model that was prepared for its analysis and a summary of the results the model produced. The paper concludes with and a short discussion describing the kind o benefits that can be gained from using small models for resilience management.

2 MICROWORLDS FOR RESILIENCE MANAGEMENT

Building resilience through resilience management requires to understand social, economic and environmental aspects of food systems (De Bruijn et al, 2014; Berkes, 2009). With this purpose, systems should be studied as a whole and the processes and subsystems within the system viewed as interdependent (Bruijn et al, 2017; Walker et al, 2002). Elements of the system traditionally considered in isolation are often part of complex structures linking them and conditioning the system outcomes (Spielman et al., 2009).

The complexity of food systems arises from the large number of interactions between many actors (e.g. farmers, retailers, workers, local governments, national governments, etc.) (Schut et al, 2014) and food systems interdependencies with socio-technical and socio-ecological systems (Giller et al, 2008; Schut et al, 2014; Olsson et al, 2014). This complex network of interactions and interdependences introduces time delays between causes and effects, reducing decidability and making it difficult for decision-makers to anticipate systems responses to shocks and changes (Davidson, 2010; Axelrod and Cohen, 2000).

One alternative for dealing with such complex systems is to use analytical constructs that help decisionmakers and stakeholders to make sense of the real world and operationalise resilience concept. Morecroft (1992) introduced the concept of ‘microworlds for policymaking debate’. Microworlds are system dynamics (SD) models that act as transitional objects stakeholders can use to explore scenarios, experiment and debate. This ‘microworlds’ help to foster understanding among stakeholders by capturing stakeholders knowledge into diagrams and enriching that knowledge with the insights from computer simulations (Morecroft, 1992).

SD models are helpful constructs for enhancing understanding about resilience because they aggregate detail while focusing on dynamic complexity, focusing on the main relationships between aggregated parts of the system and how these interactions evolve over time makes it easy to identify points for intervention. Having a simulation model also offers an opportunity for exploring the system response to different disturbance. While resilience overall is a wide, and vague, concept (Herrera, 2017; Tendall et al, 2015), it can be operationalised through the system outcomes and their behaviour when affected by an externa disturbance (Bruijn et al, 2017; Walker et al, 2004). Whereas each outcome is likely to exhibit its own particular response, for simplicity, these responses can be cluster in three big groups:

a) *building robustness:* creating the overall conditions that allow the system to withstand the shocks from the environment without significant changes in its outcomes’ behaviour. Building robustness often requires building infrastructure or building capabilities providing a first response to extreme weather conditions (e.g. creating food banks).

b) increasing stakeholders adapting capacity: fostering stakeholders’ ability for managing the system and respond to changes in the environment. These strategies often focus on making

critical thresholds and tipping points more difficult to reach by making the access and distribution of key resources more flexible across different sectors within the system (Walker et al, 2004). For instance, decentralised governance, stakeholders' networks and opportunities for innovation are often seen as critical strategies for adaptation (Biggs et al, 2012).

c) preparing for managing transformation: preparing for massive changes that otherwise could results on effects at large and catastrophic scales. If the system transforms to create a new fundamental new system, stakeholders' might be deprived of fundamental services provided by the previous system. Preparing for transformation requires to explore cross-scale interactions with other systems for building redundancy.

SD models are also great tools for identifying 'control' ('slow') variables (Ludwig et al, 1978; Holling, 1986; Carpenter and Turner, 2000). Control variables, often described as slow because they need time to change (be depleted or accumulated), differentiate from other variables because they shape how other variables, particularly outcomes, respond to external drivers. For example, soil organic matter is a control variable because, as described by Walker et al (2012), it shapes how crop production responds to variations in rainfall (external driver) during growing season.

By focusing on how control variables change it is possible to understand why systems respond differently to shocks from external variables (changes in the environment). Shocks to the system introduce might have an impact on control variables but won't produce significant change in the outcomes if the resources remain within certain thresholds. As the shocks increase in magnitude or frequency, control variables move closer to the threshold and fluctuations in the outcomes became more pronounced. These fluctuations are the results of the internal dynamics of the system and the action of feedback loops within the system (Walker et al, 2012; Carpenter and Brock, 2008; Scheffer et al, 2009). Once key resources reach their thresholds, the strength of these feedback loops shifts and the system moves towards a different, and potentially undesired, configuration exhibiting different behaviours.

In the reminded of the paper we illustrate how a small SD model can be used to facilitate policymaking debate on resilience of beef farming systems in France. The model is an aggregated conceptual system dynamics model built using historical data, case studies described in the literature (e.g. Lien et al, 2007 study in Norway and Eakin and Wehbe, 2009 studies in Latin America) and insights from the research conducted in the SURE-Farm project. SURE-Farm is a research and innovation project funded by the European Union's Horizon 2020 programme and involves 16 universities and research institutes from 11 European countries. Its full title is "Towards SUstainable and REsilient EU FARMing systems".

While the model structure is generic to many livestock and mixed crops-livestock systems, in this paper we focused on beef cattle systems in France. Having a simplified model is advantageous in this case because allow us to expand on the explanation of the analysis performed rather than spend long time describing the model itself.

3 A MICROWORLD FOR MANAGING RESILIENCE OF BEEF FARMING SYSTEMS

Farming is one of France's most important industries. The country is self-sufficient in food supplies, from cereal crops, to beef, pork and poultry to fruit and vegetables. Beef production is of economic importance since meat production is the biggest agri-food business in the country (Eurostat, 2016).

Figure 1 portrays the basic structure of the beef farming system. This structure is a simplified diagram compared to the actual system dynamics model. In Figure 1 the 'livestock units (LU)' stock represents the total amount of cattle held by the farmers in the region. The amount of livestock units is depleted by the 'slaughtering LU' outflow and replenished and increased by the 'Replacement and additional LU' inflow. The revenues from beef production are used to paid dividends to the shareholders of the farm whom, if the return on their equity is higher than the return offered in the market, decide the replenish and increase the LU (R1-Profits driving growth on livestock in Figure 2).

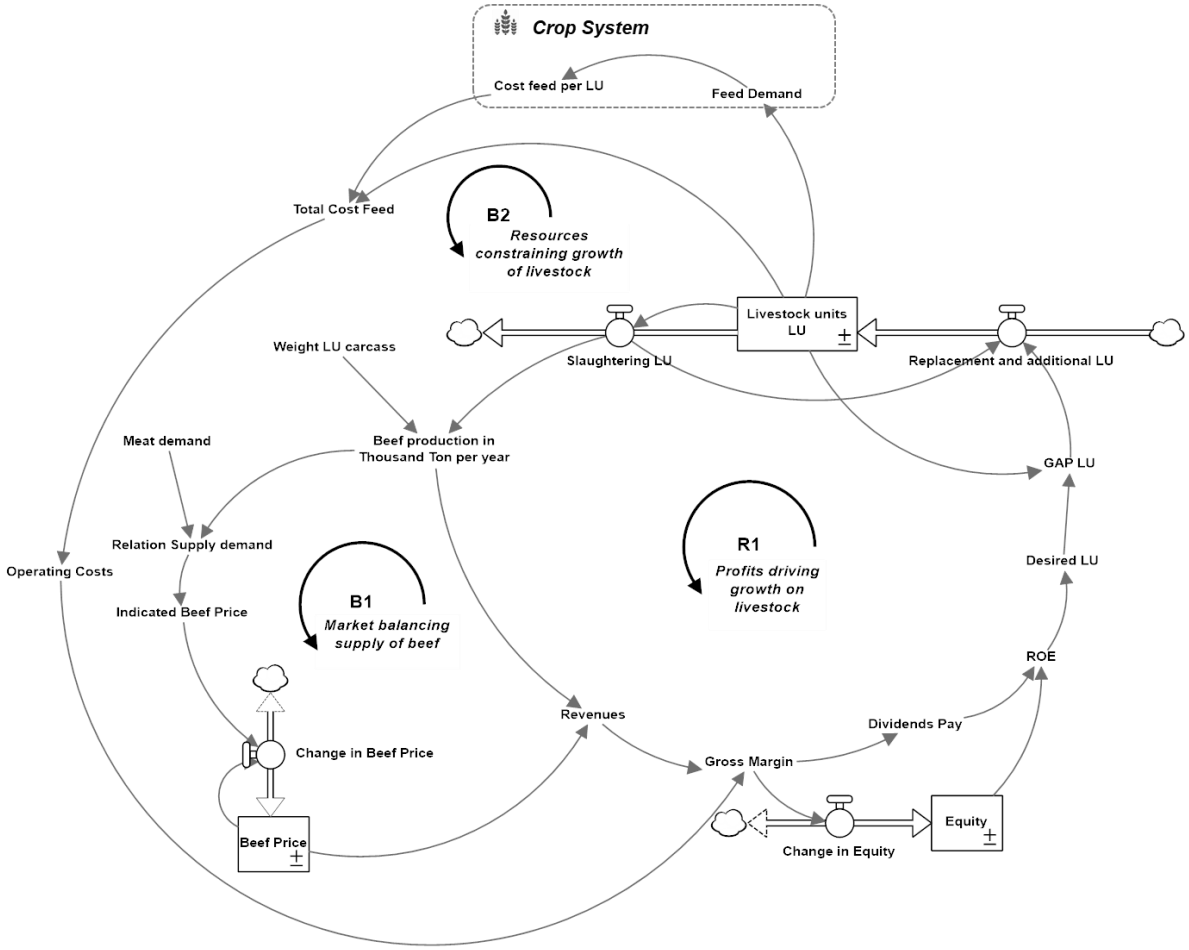


Figure 1 *Illustrative structure beef farming system*

The 'Replacement and additional LU' also depends on the selling price for the meat. The meat selling price is assumed to be elastic to the relation between supply and demand. Hence, the size of the beef farming system is constrained by the rate at which the markets (local and international) change (see B1 Market balancing supply of beef in Figure 1). Similarly, the size of the beef farming system is constrained by the resources available for production (e.g. land, feed, and workforce). In the diagram this constrained is represented by the cost of acquiring such resources with the costs increasing as they become scarcer (see B2 Resources constraining growth of livestock in Figure 1).

3.1.Introducing the impact of climate change

The change in weather patterns is a visible global phenomena (IPCC, 2013). There is substantial evidence showing that climate conditions are quickly changing around the globe as result of a sustained increase in the global mean temperature (Wheeler and von Braun, 2013). Food systems are intrinsically sensitive to changes in the weather and highly vulnerable to climate change (Campbell et al, 2016; Ericksen, 2008).

Climate change is likely to have a direct effect on the environmental drivers in ways that will diminish food system capability to support food production. For instance, the increase of droughts and weather variability is expected to result in a net reduction in crop yields (Schmidhuber and Tubiello, 2007; Wheeler and von Braun, 2013). In livestock systems, the effects of climate change are both direct and indirect. Directly, the increase in mean temperatures resulting from climate change will have an impact on the animals themselves. Indirectly, the climate change effects on production of crops and grasslands will increase operating costs making it difficult for animals to gain weight (Silanikove and Koluman, 2015).

To introduce these indirect effects, additional structure is needed in the model. Figure 2 portrays the basic structure of a crop system producing animal feed as well as crops for human consumption.

As before, the structure shown in Figure 2 is a simplified diagram compared to the actual model, but illustrates the main dynamics analysed.

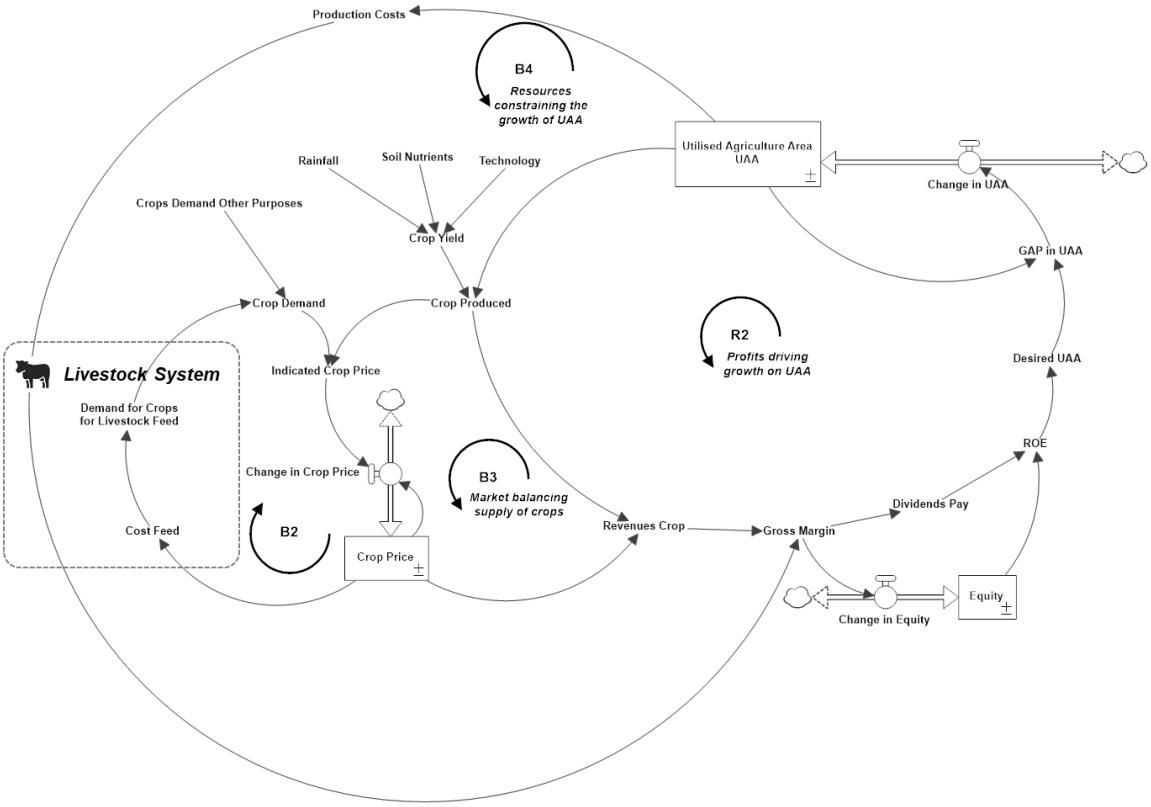


Figure 2 *Illustrative structure crop production system*

The amount of crop produced depends on the amount of utilised agricultural area (UAA) and the average yield per UAA. Like in the case of beef production, highest crop production results on higher revenues and incentive investment on more UAA and more and larger farms (see R2-Profits driving growth on UAA in Figure 2). Like in the beef farming, the investment on UAA is constrained by the demand (see B3-Market balancing supply of crops in Figure 2) and the resources available (see B4-Resources constraining the growth of UAA in Figure 2). As the UAA increases, the quality of the land used is likely to be less suitable for crop production due to deficiencies in soil nutrients, landscape and water access. However, the model assumes these deficiencies could be resolved by spending more on fertilisers, irrigation and other operating costs and we opted for keeping yields constant while making operational costs proportional to the UAA.

4 ANALYSIS AND RESULTS

The outputs of the model operationalise farming systems' performance through the outcomes and functions they provide to the region. For our analysis we used a selection of indicators suggested by the SURE-Farm resilience framework (Meuwissen et al, 2019). In particular, we looked at the system capacity to: i) deliver healthy and affordable food products and ii) ensure its economic viability. In this case, we used the proxy variables 'beef production' and 'price per carcass' as approximations of those outcomes. Using the model, we analysed the resilience of the French beef cattle systems to climate change by simulating the behaviour of those variables under more challenging weather conditions. Many studies regarding the potential impact of climate change in agriculture show that changes in weather conditions and increase on pests could reduce crops yields in Europe. For analysis purposes, we considered the system response to a single shock in weather conditions that will temporarily reduce crops' yield for a three years period. The simulation results for these variables are presented in Figure 3.

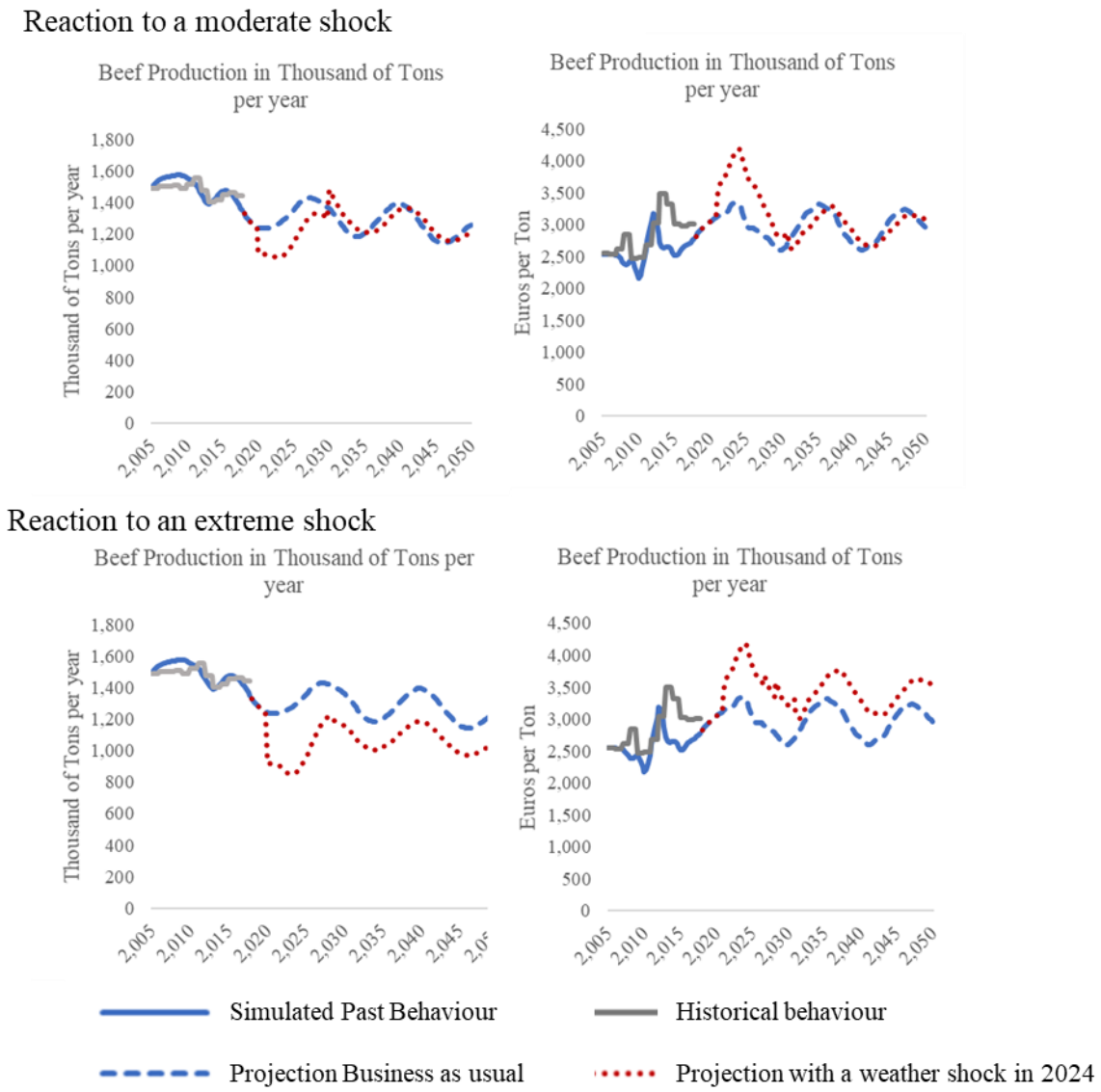


Figure 3 *Beef production and beef price response to a) a moderate shock and b) an extreme shock on the weather conditions for three years.*

The behaviour shown in Figures 4 illustrate the difference between adaptation and transformation. If the variation on the weather is moderate, the simulated results shown in Figure 4a suggests the system will underperform for a relatively short period of time but the system is eventually able to bounce back. This type of response is described in the resilience literature as 'adaptation' (Walker et al, 2004).

The mechanisms driving adaptation can be better understood by looking at the underlying structure of the system (Sieber et al, 2018; Biesbroek et al, 2014). Climate change effects reduce crops productivity increasing the production costs of the beef cattle industry. In the short term, the unbalance between crop supply and demand increases crops prices and encourages an increase on UAA (see Figure 4a). More UAA results on higher crop production what in turn reduces production costs for livestock farmers. The temporal reduction in the production costs gives the farmers an opportunity to recover after years receiving lower margins see Figures 4a.

Alternatively, if the weather variations are too intense and or occur often, the simulated results suggest that the system will move towards a new equilibrium state (Figure 4b). This type of response is known in the resilience literature as 'transformation' and the mechanisms driving this response in the system can also be understood by looking at the system structure. In these cases, the adjustments

described above in price are not enough to balance the system because the equity needed during the periods of poor performance makes it economically unattractive to remain in the business (see for example Figure 4b). Because the burning of climate change is too big to bear with, farming systems are likely to respond by shrinking to more efficient sizes and could potentially disappear.

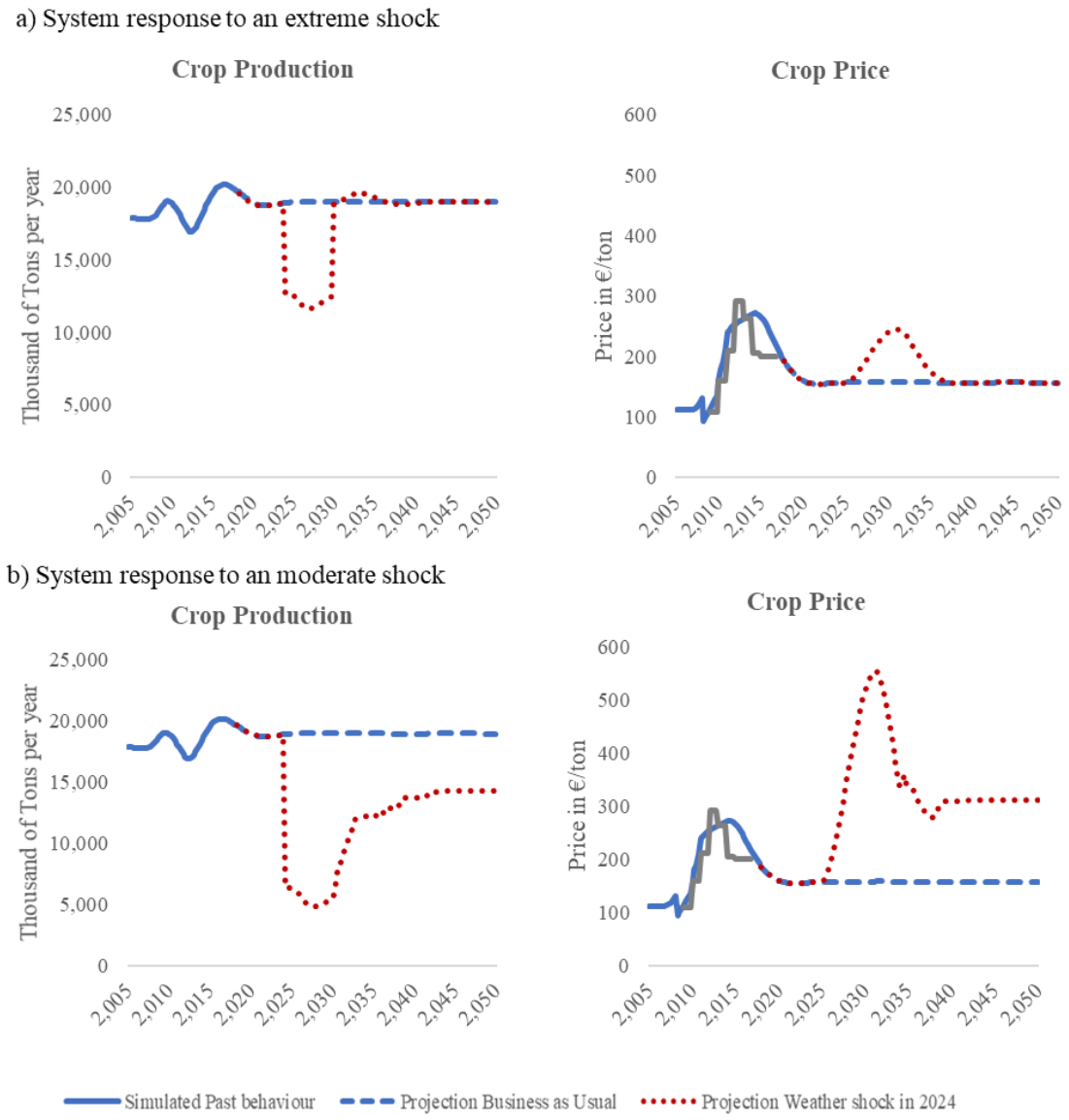


Figure 4 *Crop system response to a a) moderate shock and b) extreme shock on the weather conditions in 2024*

4.1 Exploring trade-offs: Imports vs local production

As described in Section 3, climate change has a direct effect on crops yields through water scarcity and the increase of pests and an indirect effect on the demand for crops through the reduction of forage available in the grasslands. Both reduction of yields and increase dependency on crops result on higher costs than otherwise for livestock producers, however, they have opposite effects for crop producers who seen a reduction in their throughput but an increase in their price. The long-term response of crop farmers is then governed by the elasticity of the markets and heavily influenced by market openness to foreign crops and logistic constraints.

Figure 5 shows the response of the crop production and the beef production to a shock in weather conditions when a) local crops cannot be easily substitute by foreign crops (e.g. because quotas or tariffs are in place) and b) local crops are easily replaced by foreign ones. As shown in the figure, openness to markets increases resilience of beef production but reduces resilience of crop production.

Moreover, in the long-term, beef production becomes more dependent on foreign supplies potentially decreasing its resilience to market disturbances.

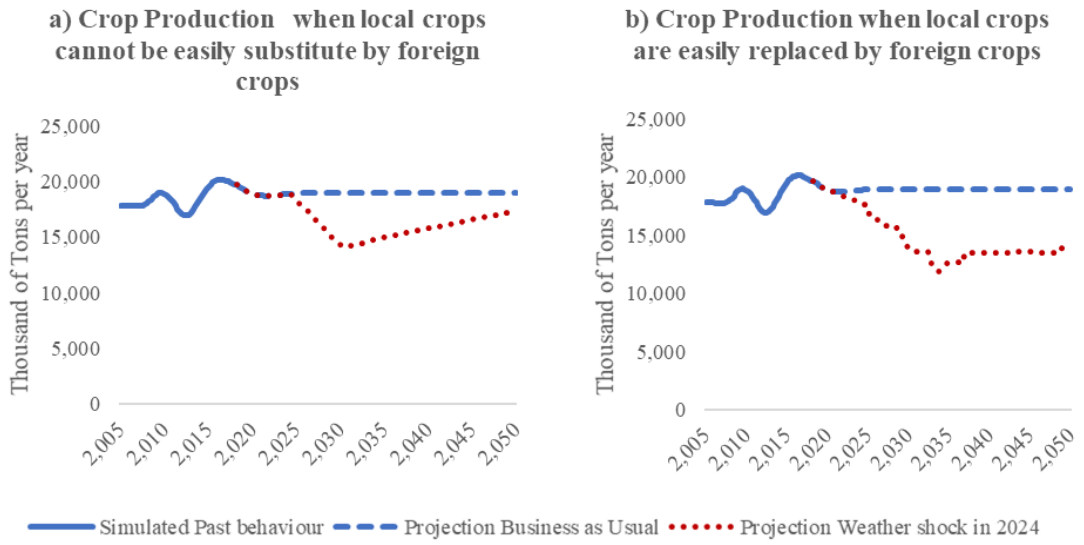


Figure 5 Crop system response to an increase on weather variability when a) local crops cannot be easily substitute by foreign crops and b) local crops are easily replaced by foreign ones

5 CONCLUSIONS

Fostering security of supplies, stability of price and financial viability of European farms in times of climate change is vital for the economic wellbeing of rural regions. Hence, although complex and challenging, resilience management of farming systems is a pressing task that needs to be undertaken. In this paper we make a case for supporting the resilience management process with small SD models that help policymakers to make sense of the problem at hand. The results in this paper show that are at least three clear benefits from taking this approach. First, small models allow us to aggregate complex systems into their main dynamics helps to understand what are the underlying mechanisms driving systems responses. The diagrams and simulation results presented in this paper illustrate how theoretical and empirical knowledge can be translated into mathematical tools that facilitate a discussion about resilience and its drivers.

Second, the simplicity and transparency of the models used also ease the analysis and discussion of potential points for intervention and strategies that can enhance resilience. Whereas the simulation results produced by the model are not meant to offer an accurate prediction of future developments, the analysis we presented shows how simulation results can be used to explore the complex mechanisms fostering resilience.

Third, having this kind of microworlds where we can experiment with different strategies might help stakeholders to understand some trade-offs between different types of resilience. As shown in this paper, looking at the simulation results and model structures makes evident some trade-offs between resilience to climate change and resilience to fluctuations in commodity markets. The results also highlight the presence of conflicting objectives between crop farming and livestock farming.

REFERENCES

Berkes F (2009). Evolution of co-management: role of knowledge generation, bridging organizations and social learning. *Journal of environmental management* **90(5)**: 1692-1702.

Biesbroek GR, Termeer CJ, Klostermann JE and Kabat P (2014). Rethinking barriers to adaptation: Mechanism-based explanation of impasses in the governance of an innovative adaptation measure. *Global Environmental Change* **26**: 108-118.

Biggs R, Schluter M, Biggs D, Bohensky EL, Burnsilver S, Cundill G, … West, PC (2012). Toward Principles for Enhancing the Resilience of Ecosystem Services. *Annual Review of Environmental Resources* **37**: 421–448.

Campbell BM, Vermeulen SJ, Aggarwal PK, Corner-Dolloff C, Girvetz E, Loboguerrero AM,… Wollenberg E (2016). Reducing risks to food security from climate change. *Global Food Security* **11**: 34–43.

Carpenter SR, and MG Turner (Eds) (2000). Special issue on interactions of fast and slow variables in ecosystems **3(6)**: 495–595.

Carpenter SR, Brock WA (2008). Adaptive capacity and traps. *Ecology and Society* **13 (40)**.

Davidson DJ (2010). The applicability of the concept of resilience to social systems: some sources of optimism and nagging doubts. *Society and Natural Resources* **23(12)**: 1135-1149.

De Bruijn KM, Diermanse FLM and Beckers, JVL (2014). An advanced method for flood risk analysis in river deltas, applied to societal flood fatality risk in the Netherlands. *Natural Hazards and Earth System Sciences* **14(10)**: 2767-2781.

De Bruijn K, Buurman J, Mens M, Dahm R, Klijn F (2017). Resilience in practice: Five principles to enable societies to cope with extreme weather events. *Environmental Science & Policy* **70**: 21-30.

Eakin HC and Wehbe MB (2009). Linking local vulnerability to system sustainability in a resilience framework: two cases from Latin America. *Climatic change* **93(3-4)**: 355-377.

Ericksen P (2008). What Is the Vulnerability of a Food System to Global Environmental Change? *Ecology and Society* **13(2)**: 18.

Fath BD, Dean C and Katzmair H (2015). Navigating the adaptive cycle: an approach to managing the resilience of social systems. *Ecology and Society* **20(2)**.

Folke C, Carpenter SR, Walker B, Scheffer M and Chapin T (2010). Resilience Thinking: Integrating Resilience, Adaptability and Transformability, *Ecology and Society* **15(4)**.

Herrera H (2017). From metaphor to practice: Operationalizing the analysis of resilience using system dynamics modelling. *Systems Research and Behavioral Science* **34(4)**: 444-462.

Holling C (1986). The resilience of terrestrial ecosystems: local surprise and global change. Pages 292–317 in W. C. Clark and R. E. Munn, editors. *Sustainable development of the biosphere*. Cambridge University Press, Cambridge, UK.

Holling CS, Gunderson LH (2002). *Panarchy: Understanding transformations in human and natural systems*.

Intergovernmental Panel on Climate Change (IPCC) (2013). Working Group I Contribution to the IPCC Fifth Assessment Report, Climate Change 2013: The Physical Science Basis.

Knickel K, Redman M, Darnhofer I, Ashkenazy A, Chebach TC, Šūmane S, ... and Strauss A (2018). Between aspirations and reality: Making farming, food systems and rural areas more resilient, sustainable and equitable. *Journal of Rural Studies* **59:** 197-210.

Lien G, Hardaker JB and Flaten O (2007). Risk and economic sustainability of crop farming systems. *Agricultural systems* **94(2):** 541-552.

Ludwig D, Jones D, Holling C (1978). Qualitative analysis of insect outbreak systems: the spruce budworm and forest. *Journal of Animal Ecology* **47**:315–332

Meuwissen, M. P., Feindt, P. H., Spiegel, A., Termeer, C. J., Mathijs, E., de Mey, Y., ... & Vigani, M. (2019). A framework to assess the resilience of farming systems. *Agricultural Systems*, 176, 102656.

Nettier B, Dobremez L, Lavorel S and Brunschwig G. (2017). Resilience as a framework for analyzing the adaptation of mountain summer pasture systems to climate change. *Ecology and Society* **22(4)**.

Saifi B and Drake L (2008). A coevolutionary model for promoting agricultural sustainability. *Ecological Economics* **65(1)**: 24-34.

Sieber IM, Biesbroek R and de Block D (2018). Mechanism-based explanations of impasses in the governance of ecosystem-based adaptation. *Regional environmental change* **18(8)**: 2379-2390.

Silanikove N and Koluman N (2015). Impact of climate change on the dairy industry in temperate zones: predications on the overall negative impact and on the positive role of dairy goats in adaptation to earth warming. *Small Ruminant Research* **123(1)**: 27-34.

Schmidhuber J and Tubiello FN (2007). Global food security under climate change. *Proceedings of the National Academy of Sciences of the United States of America* **104(50)**: 19703–8.

Vermeulen SJ, Campbell, BM and Ingram J. (2012). Climate Change and Food Systems. *Annual Review of Environmental Resources* **37**: 195–222.

Walker B, Carpenter S, Anderies J, Abel N, Cumming G, Janssen M … and Pritchard R (2002). Resilience management in social-ecological systems: a working hypothesis for a participatory approach. *Conservation ecology* **6(1)**.

Walker B, Holling CS, Carpenter SR and Kinzig A (2004). Resilience, adaptability and transformability in social--ecological systems. *Ecology and Society* **9(2)**: 5.

Walker B, Carpenter S, Rockstrom J, Crépin AS and Peterson G (2012). Drivers," slow" variables," fast" variables, shocks, and resilience. *Ecology and Society* **17(3)**.

Wheeler T and Von Braun J (2013). Climate change impacts on global food security. *Science* **341(6145)**: 508-513.

AUTHOR BIOGRAPHIES

HUGO HERRERA received a MSc. from the European Master in System Dynamics in 2015 He completed his PhD System Dynamics in the University of Bergen in 2018. Hugo is passionate about system dynamics analysis and modelling techniques and applies them in a variety of contexts. Hugo is affiliated to The University of Bergen and is currently part of the SUREFarm project.

BIRGIT KOPAINSKY is a lecturer at The University of Bergen, Norway. In her research, Birgit explores the role that system dynamics analysis and modelling techniques play in facilitating transformation processes in social-ecological systems, such as the transition towards sustainable agri-food systems on local, national and international levels. Birgit works both in Europe and in several sub-Saharan African countries and is currently part of the SUREFarm project.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.034

A BOTTOM-UP SUPPLY-SIDE SIMULATION MODEL OF RESIDENTIAL AND COMMUNITY ENERGY SYSTEMS USING SYSTEM DYNAMICS

Bilal Bugaje

Faculty of Engineering
University of Nottingham, Nottingham, NG7 2RD
bilal.bugaje@nottingham.ac.uk

Peter Rutherford

Faculty of Engineering
University of Nottingham, Nottingham, NG7 2RD
peter.rutherford@nottingham.ac.uk

Mike Clifford

Faculty of Engineering
University of Nottingham, Nottingham, NG7 2RD
mike.clifford@nottingham.ac.uk

ABSTRACT

Increasing interest in distributed generation of energy, increasing number of participants in energy planning, and the interdisciplinary nature of energy planning, provide an opportunity for new tools to model energy systems from bottom-up. This paper presents bottom-up supply-side models of residential and community energy systems based on a case study using System Dynamics. The simulation models are constructed to facilitate Systems Thinking, which can enable interdisciplinary and transdisciplinary approaches, whilst performing analyses expected from similar models. The modelling process is presented followed by discussion on the outcome of the process. The paper concludes with future work on the model.

Keywords: System Dynamics, Energy System, Simulation, Systems Thinking

1 INTRODUCTION

Centralised electricity generation has been facing many challenges like constraints to efficiency and capacity, which can be resolved by Distributed Generation (DG) (Mendes et al 2011; Takahashi et al 2005). There is a trend in research and industry towards more distributed generation, especially from renewable sources (Sadeghi et al 2017). Global trends like liberalisation of energy markets, environmental protection and sustainable development have increased interest for energy planning at smaller geographic scales (Cormio et al 2003). Moreover, the main trends in energy planning include: growing interest in DG based on renewables; growing community awareness on environmental issues; increasing number of decision makers in energy planning (Mirakyan and De Guio, 2013). These trends may also explain the rise of the "prosumer" (Rathnayaka et al 2015), who is both producer and consumer of energy. Since the implementation of energy system depends on whether the user is a producer or consumer (Project SENSIBLE Partners, 2015), the prosumer presents a challenge to implementation of energy systems, and consequently energy planning. In distributed residential and community energy systems, any individual can be a planner.

Energy planning involves many areas of problem, activities and participants, which makes it unavoidably interdisciplinary, or even transdisciplinary. Modelling the operations of energy systems is essential in energy planning of different scales especially communities and cities (Huang et al 2015; Mirakyan and De Guio, 2013). Based on the review in (Huang et al 2015), it seems that at the heart of bottom-up energy planning is techno-economic assessments, techno-ecological assessments and what-if (scenario and sensitivity) analyses. Accordingly, it has been observed in (Manfren et al 2011) that there is a need for energy planning tools that can facilitate communication in an interdisciplinary or transdisciplinary manner whilst carrying out expected analyses. This would also benefit researchers taking an interdisciplinary or transdisciplinary approach.

This study aims to demonstrate creation and possible uses of System Dynamics (SD) models for bottom-up supply-side simulation models of residential energy systems and community energy systems. Some of the contributions of this study include the use of discrete feedback in SD to model energy systems, as well as the use of a diagrammatic transdisciplinary method for simulation.

The next section provides a background to the study by looking at relevant aspects of System Dynamics, followed by a literature review of similar models in light of System Dynamics. Then there is a brief section on methods followed by a section on results and discussion, in which the modelling process, outcomes and validity considerations are presented. Finally, there is a concluding section.

2 BACKGROUND

2.1 System Dynamics

System Dynamics (SD) is a method of modelling and simulation of complex systems – with features such as feedback, nonlinearity, delay – that can analyse dynamic behaviour over time based on the principles of system structures (Forrester, 1961; Sterman, 2000). SD provides a common means of representation and communication across several disciplines and beyond formal disciplines which makes it an interdisciplinary, as well as a transdisciplinary, method. This is achieved by using the generic language of systems as outlined in the system principles (Forrester, 1997); making SD a systems approach. In addition, SD meets the four minimum criteria (system hierarchy; means of communication; adaptation; emergent properties) for Systems Thinking according to (Checkland, 2012; Checkland, 1981).

All System Dynamics models have at least one of two general aims which can be achieved via a variety of architecture (which includes but not limited to Ordinary Differential Equations, Agent Based Modelling, Discrete Event Simulation, or a combination) (Rahmandad and Sterman, 2018; Sterman, 2000): improve understanding of a system by explaining its dynamics; virtually simulate and analyse possible configurations of the system. Some models have both aims. Improving understanding is applicable to continuous feedback, whereas virtual simulation and analysis is applicable to both continuous and discrete feedback. SD also has other benefits that are characteristic of the method which all SD models can benefit from. These include being a: method of structural realism; quantitatively rigorous soft method; diagrammatic simulation method; method based on systems language. These properties make SD a suitable method for interdisciplinary and transdisciplinary approaches to research or projects.

In SD, variables are mainly categorised into stocks or flows; others are auxiliary variables and constants. In stock and flow diagrams, links can be material links or information links represented as double arrows or single arrows respectively. The direction of material links indicate the movement of the same quantity between two variables, but information links simply indicate dependence. Stocks are accumulations, while flows are the rate of accumulation.

2.2 Literature Review

Energy models have been classified in (Van Beeck, 1999; Timmerman et al 2014; Hall and Buckley, 2016) along similar dimensions. However all their classifications are limited to software packages, excluding generic modelling methods like MATLAB and System Dynamics. This study is primarily about a modelling method, not specific problems of the modelled system, even though efficacy of the model will be demonstrated via examples of problems. The models of interest are bottom-up energy supply models of (the operation of) microgrids; which can be found in (Ravindra and Iyer, 2014; Hong et al 2017; Mukherjee et al 2017; Aguilar-Jiménez et al 2018; Kitson et al 2018; Dhundhara et al 2018; He et al 2018; Adefarati and Bansal, 2019; Nnaji et al 2019; Raji and Luta, 2019; Xu et al 2019; Astriani et al 2019; Bukar et al 2019; Castillo-Calzadilla et al 2019; Cornélusse et al 2019; Ge et al 2019; Griego et al 2019).

Given the varied purposes of the models listed, their capabilities will not be compared. It is sufficient that the tools or methods have been demonstrated to carry out what they aim to do in their respective studies. Therefore the use of SD in this study is not to improve on analytic capabilities where other methods fail, but to highlight its other beneficial features, in addition to its analytic

capabilities. To the knowledge of the authors, there has been no bottom-up supply-side simulation of an energy system using System Dynamics.

The main limitation of SD in this study is the inability to implement programming concepts that may be required for metaheuristic algorithms without a third party integration. SD can be understood as a modelling method that can present a model with three layers of complexity: the diagram layer; the equations layer; the programming layer/add-on. In the case of Energy Planning, different participants may be interested in the different layers, respectively: the public and prosumers; management; industry partners. Researchers may be interested in any of the layers. It has been observed that bottom-up energy models are characterised by the first two thermodynamic laws, economic and environmental constraints (Huang et al 2015). The thermodynamic laws can be implemented in the SD model of this study as energy conservation and energy conversion loss, whilst the constraints can be simulated via impact assessments. In addition, SD can offer more realism via the principle of causal connectedness; which aims to ensure that causal links in the model must be justified by evidence from the real system.

2.3 Project SENSIBLE

The case study to be modelled is Project SENSIBLE (Storage Enabled Sustainable Energy for Buildings and Communities). The aim of SENSIBLE "is to understand the economic benefits that energy storage can bring to households, communities, and commercial buildings" (https://www.projectsensible.eu/ accessed 1 July 2018). SENSIBLE explores the use of energy storage at residential and community levels implemented in real communities. One of the communities is in The Meadows, Nottingham, UK. The plan was to implement a Community Energy System and several Residential Energy Systems made up of power electronic and communication devices; a smart grid. The project has been well documented and available at the website (https://www.projectsensible.eu/downloads/ accessed 1 July 2018).

3 METHODS

Equivalent processes of system dynamics have been outlined in Table 1 based on (Randers, 1980; Richardson and Pugh III, 1981; Roberts et al 1983; Sterman, 2000). In addition, the corresponding parallel validity tests have been outlined in (Sterman, 2000; Shreckengost, 1984; Qudrat-Ullah and Seong, 2010). The main method of sourcing data is via archival research, and the data sources include project deliverables, quantitative data from operational devices, and specification sheets of devices.

Table 1 *Stages of System Dynamics process, tools, expected outcome and corresponding validity tests*

Order	Stage	Tool(s)	Expected Outcome	Validity Test
1	Problem articulation	Model Boundary Chart	Problem theme; Problem statement; Time scale and range; Key variables and concepts.	Boundary Adequacy
2	Construction of conceptual model	Stock and Flow Diagram	A conceptual model	Structure Verification
3	Formulation of simulation model	Stock and Flow Diagram	A simulation model	Dimensional Consistency
				Parameter Verification
				Extreme Conditions
				Behaviour Reproduction
				Sensitivity Analysis
4	Conclude Validity Test	-	Confidence, or means to establish confidence, in the model's structure and behaviour.	*Not Applicable*
5	Analyses	-	Answers to questions about the real system via the simulation system	*Beyond the scope of this research.*

4 RESULTS AND DISCUSSIONS

The discussion that follows is based on the stages of the System Dynamics process and their corresponding validity tests. The validity tests will be addressed in form of responses to the questions they pose, as outlined in (Sterman, 2000). For brevity, each question will not be dedicated a subsection. Moreover, the discussion that follows is not limited to the questions of the validity tests.

4.1 Problem Articulation

The purpose of the System Dynamics exercise is to create simulation models of residential and community energy systems from Project SENSIBLE to demonstrate scenario analysis, techno-economic impact analysis, and techno-ecological impact analysis. Therefore the focus is on the operational, economic and environmental dimensions of the system. The time resolution of the simulation model is minutes because the field devices measure in minutes, and the time range is the duration of a day in order to make system behaviour visually distinguishable in a time-series graph.
Table 2 is the model boundary chart showing variables from the real system that will be modelled (endogenous and exogenous) and those that will not be modelled (excluded). Forecasting, storage optimisation and energy market services are modules within SENSIBLE that rely on proprietary algorithms (which include machine learning and genetic algorithms) and stored data. The outcome of these modules are signals to the energy systems which have been represented as exogenous variables.

Table 2 *Model Boundary Chart*

Endogenous	Exogenous	Excluded
• Consumption • Battery State of Charge (SoC) • PV Consumption • PV Loss • Power Import • Power Export • PV Charging • PV Charging Loss • PV Discharging • PV Discharging Loss	• Load demand • PV production • Storage efficiency • Conversion efficiency • Signal to charge battery from grid • Signal to supply load from grid without battery discharge • Signal to sell to grid without battery charge • Signal to limit maximum battery SoC from grid • Power ratings	• Reactive power • Battery lifetime • Auto discharge

4.2 Conceptual Model

Data about structure of the system was obtained from project documents: SENSIBLE Deliverables (https://www.projectsensible.eu/downloads/ accessed 1 July 2018). The residential and community energy systems are made of devices with a variety of capabilities. Each system could be modelled as an electricity system or information system, depending on the level of abstraction; as an electric circuit or information network. The systems as information system is better suited for the purpose of this study. The chosen level of aggregation is to model power in Watts and energy in Watt-minutes because most of the monitoring devices measure power in Watts.

A conceptual model of a Residential Energy System is shown in Figure 1, created using Vensim. The conceptual model can facilitate Systems Thinking because it is a diagram that explicates the interdependence in a system with feedback (Richmond, 1993; Forrester, 1994; Sterman, 2000). In addition to causal dependency indicated by all arrows, other observable features of the material links (double arrows) are energy conservation and energy conversion loss. The simulation model is built on the conceptual model.

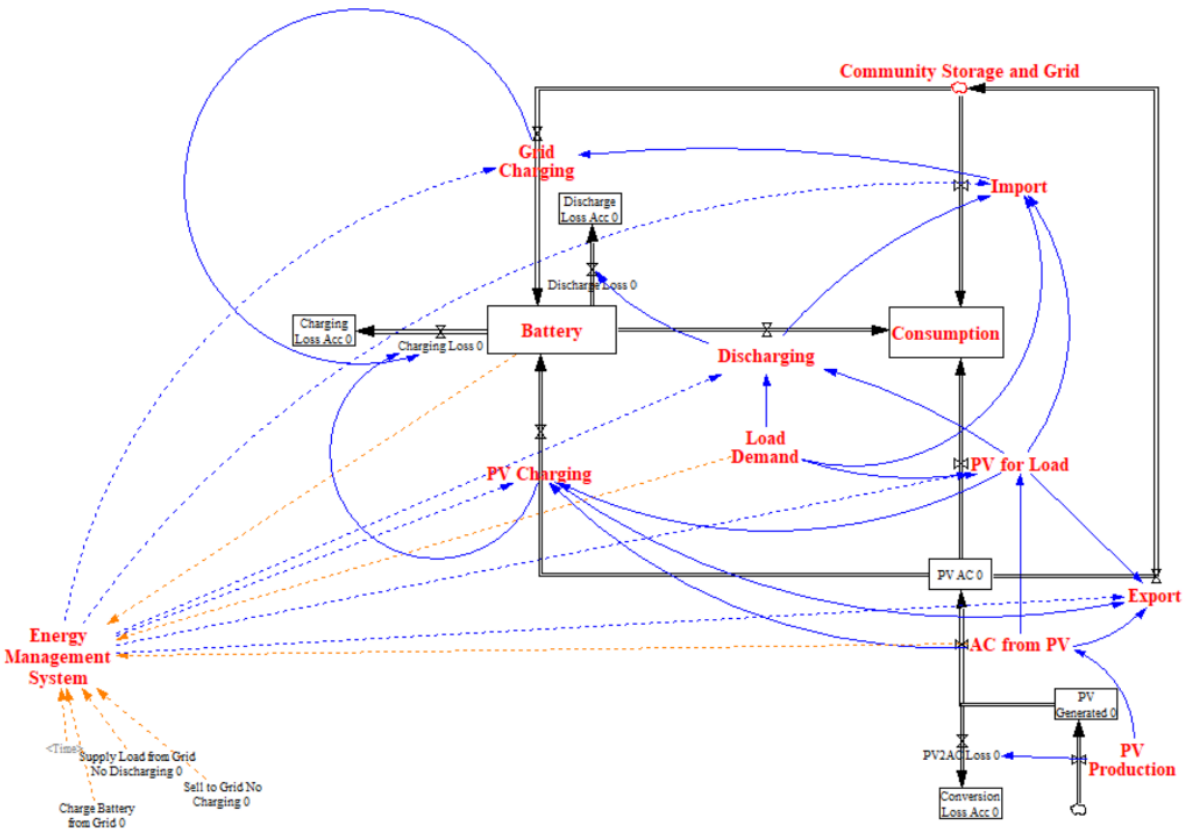


Figure 1 *Conceptual Model of a Residential Energy System showing causal dependencies among system elements using Stock and Flow Diagram. Dotted arrows indicate connections to and from the Energy Management System.*

4.3 Simulation Model

For moments where load demand exceeds the combined rated values of all power sources, the unserved load is monitored separately as a validation measure. The simulation models implement the following in fidelity with the real system:

- Nonlinear efficiency and loss of energy during energy conversion by the power inverters.
- Energy conservation such that energy can be accounted for from source to load and loss.
- Enforcement of causal connectedness such that decision making elements make decisions using only information realistically available to them.

All system rules have been derived from description of three use cases applicable to the Nottingham Demonstrator of SENSIBLE (https://www.projectsensible.eu/downloads/ accessed 1 July 2018). The discrete feedback of the system is determined by the system decisions. The model was checked using the integrated Units Checking tool in Vensim and there are no errors, which confirms dimensional consistency. All parameter values are obtained from the same project documents used for conceptual model, as well as device specification documents. Extreme values are handled within equations, for example, using rated values of devices. Only one normaliser variable in the model has no real world counterpart but it is used in order to balance the units. It is called "Energy to Power Normaliser", which has a value of 1 and a unit of 1/time. This is used in equations of the following variables where a power value is to be determined from energy values: Charge Power to Max; Discharge Power to Min; and Grid Charge Limit.

To run a simulation, a time series of electric load is provided as input. Data from real residences with four different composition of devices were obtained.

4.3.1 Validity Test - Behaviour Reproduction

The aim is to investigate if a model reproduces historical behaviour in the system. See Figure 2a for example. In general, for the same 24 hour period, the behaviour of the simulation models of the four residences matches data from the real system; for the variables of battery charging, battery discharging, power import, power export, and battery state of charge. The exception is in the case of import and export for the residence with Grid + Battery + PV, but more significant in the export. The discrepancy in export follows behaviour that does not match what has been documented.

Additionally, the error may be explained by one of three reasons: the real data is erroneous; or the documents did not capture properly the system's rules; or the rules were not properly applied in the simulation. The error is likely to be from the real data because other simulations resulted in reasonable match, including variables of the same residence (charging, discharging and battery state of charge), and the rules are from the same sources. Also, the use of optimisation signals (to export power or not), does not explain the error. At the moment, efforts are being made to obtain real data from other residences with the same configuration in order to verify.

4.4 Analyses

4.4.1 Scenario Analysis

The aim is to simulate and compare the dynamic behaviour of different compositions of the system. See Figure 2b for examples; it visually compares power import from simulations of Community Energy Systems with and without a community battery. During daylight hours, there is significant import from the grid in the scenario where there is no community battery because there is no stored energy that may have been obtained at night when energy is cheaper.

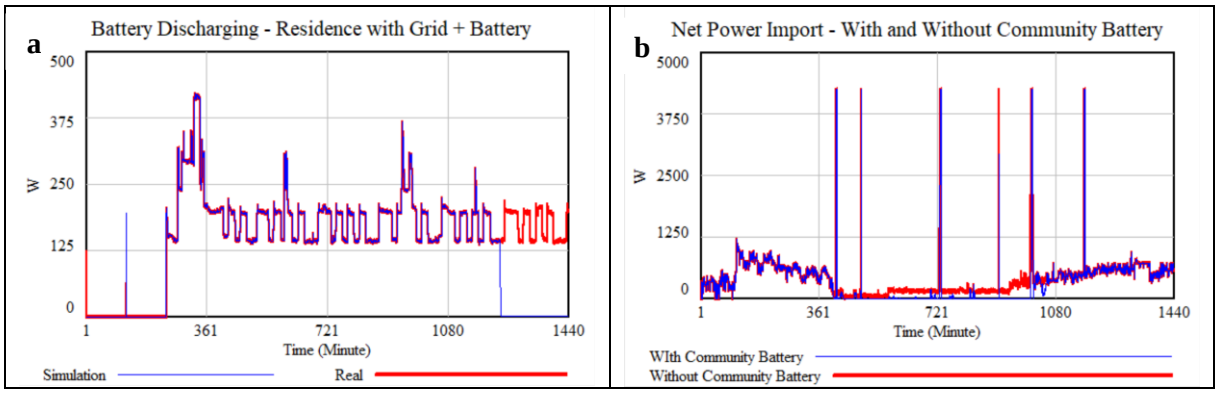


Figure 2 *a) Example of behaviour comparison between SD simulation and the real system for different residences. b) Example of Scenario Analysis comparing the same Community Energy System with and without community battery.*

4.4.2 Environmental Impact Analysis

Figure 3 estimates and compares CO2 (Carbon Dioxide) equivalent footprint and savings of different system compositions of the real system. The conversion factor is 0.28307 kg of CO2 saved for each kilowatt-hour (kWh) produced from a carbon free source, based on RenSmart (https://www.rensmart.com/Calculators/KWH-to-CO2 accessed 1 July 2019). Within a 24 hour period, a residence fitted with PV panels has an additional impact of about 400 grams of CO2 equivalent of imported power when it does not have a battery. However it appears there is no difference within the same 24 hour period with regards to CO2 savings from export; given the state of other variables. These values are influenced by the battery state of charge at time 0.

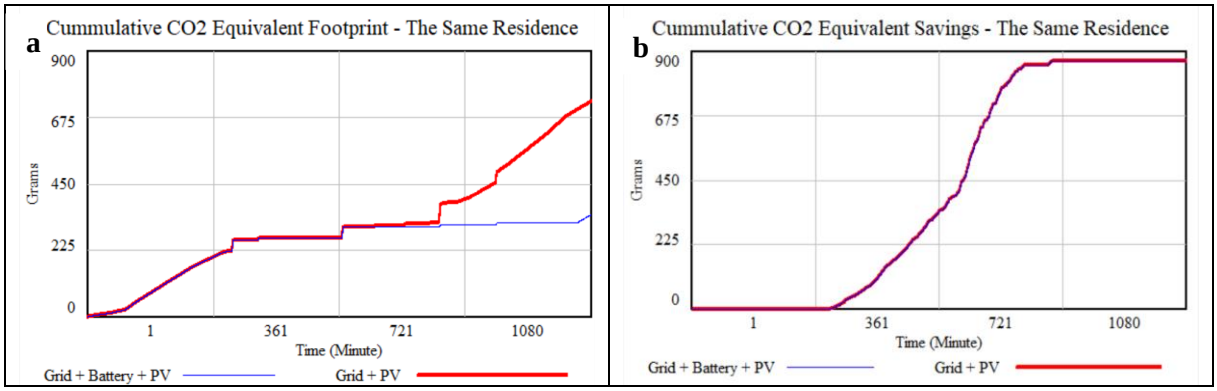


Figure 3 *Example of Environmental Impact Analysis: a) Comparing cumulative CO2 equivalent footprint from power import between the same residence with two different compositions; b) Comparing cumulative CO2 equivalent savings power export between the same residence with two different compositions.*

4.4.3 Economic Impact Analysis

Figure 4 estimates and compares monetary costs and savings of a residence with and without battery, when it is fitted with PV panels. The cost per kWh is based on Green Energy's TIDE tariff for weekdays (https://www.greenenergyuk.com/PressRelease.aspx?PRESS_RELEASE_ID=76 accessed 1 July 2019). Within a 24 hour period, a residence fitted with PV spends in excess of about 25 pence from imported power when it does not have a battery. On the other hand the residence earns about 10 pence in the same period from exporting excess power generated from PV. These values are influenced by the battery state of charge at time 0. Other system compositions of the real system that can be explored include: Community Energy System with and without battery; centralised community battery and individual residential batteries.

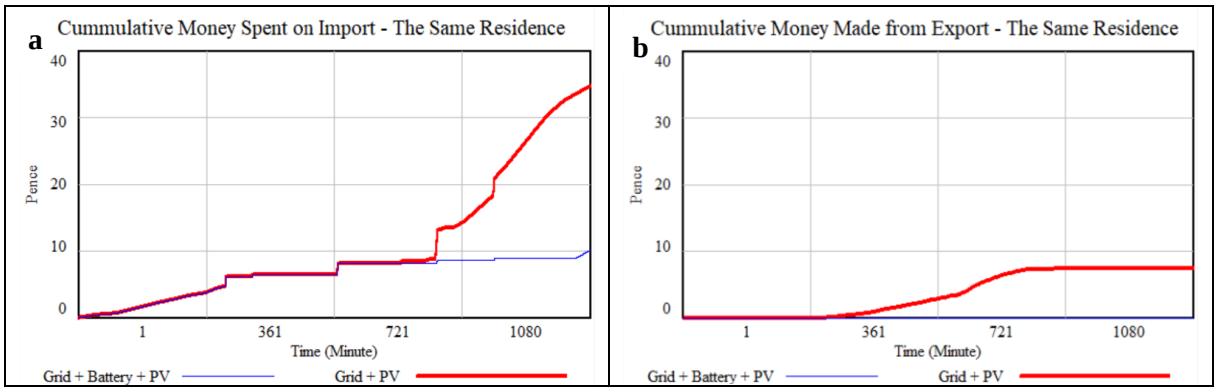


Figure 4 *Example of Economic Impact Analysis: a) Comparing cumulative money spent on power import between the same residence with two different compositions; b) Comparing cumulative money earned from power export between the same residence with two different compositions.*

5 CONCLUSION AND FURTHER WORK

System Dynamics has the capability to model supply-side of residential and community energy systems using a bottom-up approach. The simulation model facilitates systems thinking by being diagrammatic, modular and having multiple layers of detail. In addition, the model features energy conservation, nonlinear conversion efficiency and causal connectedness. Though the model considered the sophisticated optimisation of the real system as exogenous, it behaved like the real system, by mainly relying on the descriptions of use cases. Finally, the model could be used for a number of analyses, and examples have been presented.

Simulation models like the one presented in this paper could be a step towards allowing the public and prosumers to take part in transdisciplinary Energy Planning, as well as researchers in interdisciplinary and transdisciplinary research. In the future, the model could be expanded to include electric heating and Electric Vehicles (EV). In its current state, the model can be readily integrated with other System Dynamics models that have common variables. An issue worth exploring is the scalability of the community energy systems when it is made up of many residences beyond the few presented in this paper.

ACKNOWLEDGMENTS

Much thanks to Richard Davies, of FlexElec Laboratory at the University of Nottingham, for clarifying enquiries about the case study and providing data. Also thanks to Petroleum Technology Development Fund (PTDF) for support to one of the authors.

REFERENCES

Adefarati, T., & Bansal, R. C. (2019). Reliability, economic and environmental analysis of a microgrid system in the presence of renewable energy resources. *Applied Energy*, *236*(November 2018), 1089–1114. https://doi.org/10.1016/j.apenergy.2018.12.050

Aguilar-Jiménez, J. A., Velázquez, N., Acuña, A., Cota, R., González, E., González, L., … Islas, S. (2018). Techno-economic analysis of a hybrid PV-CSP system with thermal energy storage applied to isolated microgrids. *Solar Energy*, *174*(August), 55–65. https://doi.org/10.1016/j.solener.2018.08.078

Astriani, Y., Shafiullah, G. M., Anda, M., & Hilal, H. (2019). Techno-economic evaluation of utilizing a small-scale microgrid. *Energy Procedia*, *158*, 3131–3137. https://doi.org/10.1016/j.egypro.2019.01.1013

Bukar, A. L., Tan, C. W., & Lau, K. Y. (2019). Optimal sizing of an autonomous photovoltaic/wind/battery/diesel generator microgrid using grasshopper optimization algorithm. *Solar Energy*, *188*(June), 685–696. https://doi.org/10.1016/j.solener.2019.06.050

Castillo-Calzadilla, T., Macarulla, A. M., Kamara-Esteban, O., & Borges, C. E. (2019). A case study comparison between photovoltaic and fossil generation based on direct current hybrid microgrids to power a service building. *Journal of Cleaner Production*, *244*, 118870. https://doi.org/10.1016/j.jclepro.2019.118870

Checkland, P. (1981). Systems thinking, systems practice John Wiley & Sons. *New York*.

Checkland, P. (2012). Four Conditions for Serious Systems Thinking and Action. *System Research and Behavioral Science*, *29*(October), 465–469. https://doi.org/10.1002/sres.2158

Cormio, C., Dicorato, M., Minoia, A., & Trovato, M. (2003). A regional energy planning methodology including renewable energy sources and environmental constraints. *Renewable and Sustainable Energy Reviews*, *7*(2), 99–130. https://doi.org/10.1016/S1364-0321(03)00004-2

Cornélusse, B., Savelli, I., Paoletti, S., Giannitrapani, A., & Vicino, A. (2019). A community microgrid architecture with an internal local market. *Applied Energy*, *242*(October 2018), 547–560. https://doi.org/10.1016/j.apenergy.2019.03.109

Dhundhara, S., Verma, Y. P., & Williams, A. (2018). Techno-economic analysis of the lithium-ion and lead-acid battery in microgrid systems. *Energy Conversion and Management*, *177*(September), 122–142. https://doi.org/10.1016/j.enconman.2018.09.030

Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, Mass.: M.I.T. Press.

Forrester, J. W. (1994). System Dynamics, Systems Thinking, and Soft OR. *System Dynamics Review*, *10*(2), 1–14.

Forrester, J. W. (Editor). (1997). Road Maps 1-9: A Guide to Learning System Dynamics. System Dynamics in Education Project. System Dynamics Group, Sloan School of Management, Massachusetts Institute of Technology.

Ge, S., Sun, H., Liu, H., Li, J., Zhang, X., & Cao, Y. (2019). Reliability evaluation of multi-energy microgrids: Energy storage devices effects analysis. *Energy Procedia*, *158*, 4453–4458. https://doi.org/10.1016/j.egypro.2019.01.769

Griego, D., Schopfer, S., Henze, G., Fleisch, E., & Tiefenbeck, V. (2019). Aggregation effects for microgrid communities at varying sizes and prosumer-consumer ratios. *Energy Procedia*, *159*, 346–351. https://doi.org/10.1016/j.egypro.2019.01.004

Hall, L. M. H., & Buckley, A. R. (2016). A review of energy systems models in the UK: Prevalent usage and categorisation. *Applied Energy*, *169*, 607–628. https://doi.org/10.1016/j.apenergy.2016.02.044

He, L., Zhang, S., Chen, Y., Ren, L., & Li, J. (2018). Techno-economic potential of a renewable energy-based microgrid system for a sustainable large-scale residential community in Beijing, China. *Renewable and Sustainable Energy Reviews*, *93*(October 2016), 631–641. https://doi.org/10.1016/j.rser.2018.05.053

Hong, Y. Y., Chang, W. C., Chang, Y. R., Lee, Y. Der, & Ouyang, D. C. (2017). Optimal sizing of renewable energy generations in a community microgrid using Markov model. *Energy*, *135*, 68–74. https://doi.org/10.1016/j.energy.2017.06.098

Huang, Z., Yu, H., Peng, Z., & Zhao, M. (2015). Methods and tools for community energy planning: A review. *Renewable and Sustainable Energy Reviews*, *42*(4800), 1335–1348. https://doi.org/10.1016/j.rser.2014.11.042

Kitson, J., Williamson, S. J., Harper, P. W., McMahon, C. A., Rosenberg, G., Tierney, M. J., … Gautam, B. (2018). Modelling of an expandable, reconfigurable, renewable DC microgrid for off-grid communities. *Energy*, *160*, 142–153. https://doi.org/10.1016/j.energy.2018.06.219

Manfren, M., Caputo, P., & Costa, G. (2011). Paradigm shift in urban energy systems through distributed generation: Methods and models. *Applied Energy*, *88*(4), 1032–1048. https://doi.org/10.1016/j.apenergy.2010.10.018

Mendes, G., Ioakimidis, C., & Ferrão, P. (2011). On the planning and analysis of Integrated Community Energy Systems: A review and survey of available tools. *Renewable and Sustainable Energy Reviews*, *15*(9), 4836–4854. https://doi.org/10.1016/j.rser.2011.07.067

Mirakyan, A., & De Guio, R. (2013). Integrated energy planning in cities and territories: A review of methods and tools. *Renewable and Sustainable Energy Reviews*, *22*, 289–297. https://doi.org/10.1016/j.rser.2013.01.033

Mukherjee, U., Maroufmashat, A., Ranisau, J., Barbouti, M., Trainor, A., Juthani, N., … Fowler, M. (2017). Techno-economic, environmental, and safety assessment of hydrogen powered community microgrids; case study in Canada. *International Journal of Hydrogen Energy*, *42*(20), 14333–14349. https://doi.org/10.1016/j.ijhydene.2017.03.083

Nnaji, E. C., Adgidzi, D., Dioha, M. O., Ewim, D. R. E., & Huan, Z. (2019). Modelling and management of smart microgrid for rural electrification in sub-saharan Africa: The case of Nigeria. *Electricity Journal*, *32*(10). https://doi.org/10.1016/j.tej.2019.106672

Project SENSIBLE Partners. (2015). *Deliverable 1.1: Energy storage domain roles & classification*. Retrieved from https://www.projectsensible.eu/downloads/

Qudrat-Ullah, H., & Seong, B. S. (2010). How to do structural validity of a system dynamics type simulation model: The case of an energy policy model. *Energy Policy*, *38*(5), 2216–2224. https://doi.org/10.1016/j.enpol.2009.12.009

Rahmandad, H., & Sterman, J. (2018). System Dynamics or Agent-Based Models ? Wrong question! Seek the right level of aggregation . System Dynamics Society. Retrieved from https://www.systemdynamics.org/assets/docs/sdorabm.pdf

Raji, A. K., & Luta, D. N. (2019). Modeling and optimization of a community microgrid components. *Energy Procedia*, *156*(September 2018), 406–411. https://doi.org/10.1016/j.egypro.2018.11.103

Randers, J. (1980). Guidelines for model conceptualization. *Elements of the System Dynamics Method*, *117*, 139.

Rathnayaka, A. J. D., Potdar, V. M., Dillon, T., & Kuruppu, S. (2015). Electrical Power and Energy Systems Framework to manage multiple goals in community-based energy sharing network in smart grid. *International Journal of Electrical Power and Energy Systems*, *73*, 615–624. https://doi.org/10.1016/j.ijepes.2015.05.008

Ravindra, K., & Iyer, P. P. (2014). Decentralized demand-supply matching using community microgrids and consumer demand response: A scenario analysis. *Energy*, *76*, 32–41. https://doi.org/10.1016/j.energy.2014.02.043

Richardson, G. P., & Pugh III, A. I. (1981). *Introduction to system dynamics modeling with DYNAMO*. Productivity Press Inc.

Richmond, B. (1993). Systems thinking: critical thinking skills for the 1990s and beyond. *System Dynamics Review*, *9*(2), 113–133.

Roberts, N., Andersen, D., Deal, R., Garet, M., & Shaffer, W. D. (1983). Introduction to Computer Simulation: The System Dynamics Approach.

Sadeghi, H., Rashidinejad, M., & Abdollahi, A. (2017). A comprehensive sequential review study through the generation expansion planning. *Renewable and Sustainable Energy Reviews*, *67*(May 2016), 1369–1394. https://doi.org/10.1016/j.rser.2016.09.046

Shreckengost, R. C. (1984). DYNAMIC SIMULATION MODELS: HOW VALID ARE THEY?

Sterman, J. (2000). *Business dynamics : systems thinking and modeling for a complex world*. Boston: Irwin McGraw-Hill.

Takahashi, K, Baker, S, Kurdgelashvili, L. (2005). Policy Options to Support Distributed Resources, 1–94.

Timmerman, J., Vandevelde, L., & Van Eetvelde, G. (2014). Towards low carbon business park energy systems: Classification of techno-economic energy models. *Energy*, *75*, 68–80. https://doi.org/10.1016/j.energy.2014.05.092

Van Beeck, N. M. J. P. (1999). Classification of energy models. *FEW Research Memorandum*, *777*, 25.

Xu, D., Mumata, M., & Mogi, G. (2019). Economic Comparison of Microgrid Systems for Rural Electrification in Myanmar. *Energy Procedia*, *159*, 309–314. https://doi.org/10.1016/j.egypro.2019.01.010

AUTHOR BIOGRAPHIES

BILAL BUGAJE received a BEng (Hons) Electrical Electronic Engineering from the University of Bradford in 2008. He obtained an MSc Information Technology from the University of York in 2012. He is currently a PhD student in the Faculty of Engineering at the University of Nottingham since 2018.

PETER RUTHERFORD moved to the University of Nottingham's Department of Architecture and Built Environment in 2002 lecturing in areas including building / architectural science, design studio and environmental psychology. His research at Nottingham has looked at virtual acoustics, the relationship between performers and performance spaces, acoustics within mental health facilities and the impact of acoustics on native and non-native listeners within HE lecture facilities.
https://www.nottingham.ac.uk/engineering/departments/abe/people/peter.rutherford

MIKE CLIFFORD has been at the University of Nottingham since 1994, and is currently an associate professor. His past research include pollution transport studies, car exhaust emissions and building ventilation. Currently, his research areas include appropriate technology, composite materials, dynamical systems, fluid dynamics.
https://www.nottingham.ac.uk/Engineering/Departments/M3/People/mike.clifford

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.035

THE AGENCY ROLE OF SD MODELS IN MODEL-BUILDING GROUPS

Dr. Maria Kapsali

Manchester Metropolitan University
Manchester Metropolitan University Business School
Manchester, M15 6BH
UK
m.kapsali@mmu.ac.uk

Dr. Steffen Bayer

University of Southampton

Southampton Business School
Southampton SO17 1BJ
UK

s.c.bayer@soton.ac.uk

Prof. Sally Brailsford

University of Southampton

Southampton Business School
Southampton SO17 1BJ
UK

s.c.brailsford@soton.ac.uk

Dr. Tim Bolt

Saitama University

Faculty of Economics
255 Shimo-Okubo, Sakura-ku
Saitama 338-8570
Japan
tbolt@mail.saitama-u.ac.jp

ABSTRACT

This paper examines the role of a simulation model being developed within a group model building process. The paper argues that the model can be conceptualized as a being an agent in a model building group which influences, shapes and challenges the thinking of the group members. The paper presents two case studies of group model building projects and shows how the agency of the model in differently composed groups.

Keywords: Systems Dynamics; technological agency; case-based reasoning; model-based reasoning; group model building

1 INTRODUCTION

Simulation artefacts such as System Dynamics (SD) models are conceptual integrators and instrumental learning tools and as such are important in complex problem-solving. They are also found to perform a social role, e.g. as boundary objects. It is not clear however if SD modelling is more than just a tool, but rather plays the role of an agent when humans interact with it: "*Can an SD model be conceptualized as exerting agency in group-model building, by influencing the reasoning of the group members – and if so, how*?"

The setting is the group-model building process. We examine evidence of the influence of SD models on the reasoning of groups in two in-depth case studies.

In the rest of the article, we first explore this argument looking at two literatures: (i) the roles of technical artefacts in temporary groups and projects; and (ii) the SD group-model building. In the Section on 'Research Design' we explain how we investigated technological agency. Two in-depth case studies are offered in the 'Discussion of Case findings' Section. Then we conclude with the contributing points of this investigation.

2 LITERATURE REVIEW

2.1 The roles of technical artefacts in temporary groups and projects

Research in the field of temporary groups and projects consider the role of technological artefacts but has yielded diverse findings. One line of studies suggests that ICT artefacts cannot replace personal interaction (Sapsed and Salter, 2004) especially when there are too many interfaces (Hoegl et al., 2004) or when either interdependency is weak or knowledge creation and codification are complex (Brensen et al., 2004). A second line of studies advocates that technical artefacts '*enframe predictable and coordinated action*' (Kallinikos, 2005: 199) supporting predictable interactions, or partially reflect and embody hierarchy (Eriksson-Zetterquist et al., 2009) or materialize the temporal structuring of routines (Orlikowski, 2007).

A third line of studies examines the social roles of technological artefacts as boundary objects. These roles refer to mediating between the group and the parent organisation (Garrety et al., 2004); aligning objectives and handling conflict, negotiations, creativity, and contracts (Koskinen and Makinen, 2009; Jensen et al., 2006; Alderman et al., 2005); integrating and codifying knowledge and symbolic value (Swan et al., 2007); and supporting learning (Yakura, 2002) and sense-making (Papadimitriou and Pellegrin, 2007). On the other hand, they can also serve to intensify tensions, divisions and information asymmetry and to assert roles, norms, identity, power or status within group relations (Dodgson et al., 2007; Carlile, 2002) reifying cultural boundaries (Barrett and Oborn, 2010).

Previous studies in temporary groups and projects addressing the roles of technological artefacts do not consider their agency (Chongthammakun and Jackson, 2012) and see them merely as *carriers*, *representations* or *integrators* of knowledge, meaning, learning and organizing (Ewenstein and Whyte, 2009).

2.2 The process of group-model building

Simulation building is usually envisaged as a linear-rational process with a series of logical steps. However, recent studies have revealed a slightly different picture. Tako (2008) and Tako and Robinson (2010) found that SD projects in particular, deviate slightly from this stereotype in these ways: because SD models are systemic (i.e. built on relationships and interactions within a system) and represent a social system as a whole, they are suitable for policy and strategic problems that are quite complex. In such a situation, SD modellers are compelled to consider the broader aspects of a problem in a more holistic pattern. SD group-model building is not linear, and the cyclicality of action is distinctive. SD modellers jump iteratively from conceptual modelling to coding or data input and then go back to modelling, etc. Causal relationships, dynamic complexity and feedback are important conversations in SD groups who need both quantitative and qualitative data to address them. SD modellers perform 'black-box' validation: this means checking both the numerical outcomes and the patterns of the qualitative results to validate the model and are keener to experiment with scenarios. Users find SD slightly more representative of the problem and instructive, but complex.

The SD model building process is divided into two phases. First phase is the construction of a shared conceptual frame (or the model structure). In the construction phase, a group identifies links from diverse stories and these links are used to build the model structure and to define the data that needs to be collected. In this phase, people exercise two types of reasoning: case-based and model-based. Human sense-making and artefacts interact (Hernes and Maitlis, 2010: 31) and inductive (case-based) (Graham, Smith and Crapper, 2004) and abductive (model-based) reasoning (Develaki, 2017: 1003) may occur simultaneously or iteratively. Case-based reasoning happens when experiences of specific prior reasoning episodes are recalled and adapted to interpret a similar situation. Model-based reasoning happens when people correlate data in a model that can then be used to make sense of the "real world". In the reification phase, the SD consultants determine the equations describing the mathematical relationships, code the model in a simulation package or programming language and parameterise the model with data. Finally, they test different scenarios and validate the model.

The model building process can be problematic. Systems that are modelled are usually complex and problems are 'fuzzy' or ill-structured and the 'problem space' (Simon, 1978) is not well defined. Problem definition is the most important, intense, time-consuming and risky step in the construction

phase (Cochran et al., 1995). A major challenge in achieving group consensus around the problem definition is the existence of priors, which are conceptual preconceptions, assumptions or expectations (Urban, 1974). During testing scenarios, models constrain the ways in which they can be interpreted (Knuuttila and Voutilainen, 2003), and they can produce results that collide with priors. A further difficulty stems from the need to decide how detailed or abstract the model structure should be (Urban, 1974). A complex model will often challenge human understanding. Major hindrances include the lack of familiarity with modelling, and the inability to think in abstract terms and to prioritize information from multiple sources (Cochran et al., 1995).

3 RESEARCH DESIGN

We studied the agency patterns of simulation artefacts, and we define SD agency here as the ability to challenge the perceptions of a diverse group of people, create a reaction and change their individual stories and mode of reasoning, in the two modelling phases. Our definition is influenced by the definition of the material effects of technologies as "tangible resources that provide people with the ability to do old things in new ways and to do things they could not do before" (Leonardi and Barley 2008, 161). In this situation, we argue, the mode of reasoning is affected when the group members repeatedly confront the constraining and enabling features of the conceptual technology (technological agency; Jarrahi and Nelson, 2018). We therefore looked for evidence of how the SD model causes change in the type of reasoning (case vs. model) at any time during the group-model building project phases.

3.1 Data collection

We collected evidence from two group-modelling projects that both run over the course of one year, and we conducted ethnographic, secondary and in-depth interview methods. Firstly, by participating in the series of modelling workshops (eleven sessions in total), in which we observed how the groups worked. Secondly, by conducting interviews with the participants in both projects afterwards (an additional year), including clients, consultants and expert advisory members (seventeen interviews in total). We also collected their project documentation, including minutes of their meetings.

3.2 Data analysis

In examining the possible agency of models in group-model building, we adopted the approach of theory elaboration. Theory elaboration is conducted through qualitative analysis and depends on *comparison;* data from each case are used to refine concepts and then to compare the outcomes.

Our analysis was conducted in two steps. First, we used content analysis and comparative matrix analysis (Miles and Huberman, 1994) to identify how participants interacted with the SD model. Secondly, we mapped what happened in the phases to capture the roles and effects of the model.

3.3 Context

In theory, elaboration occurs simultaneously provided cases have been chosen: cases of similar phenomena with well-defined differences (Glaser and Strauss, 1967). Our comparative in-depth case design provides a strong foundation for elaborating theory (Bluhm et al., 2011) because the similarities between the cases allow for meaningful comparisons, and their differences provide a basis for discovering new themes. The groups we selected had two important similarities and one difference. First, both groups had similar heterogeneous consistency. The second similarity was that both groups had to tackle the same problem: to build a model that helps to develop interventions and policies to reduce hospital admissions due to the same type of health issue.

Group A consisted of twelve people from one locality, including a project manager organising the meetings and liaising with the client, a simulation consultant, a data manager and nine local group members. These included a commissioner at the local PCT, representatives of different NHS services (4 members) as well as of the local council (2 members) and from third-sector service providers (3 group members).The group met five times within a period of five months, during which observations were made by our team. Group B was charged with developing a simulation model for hospital admissions for use by PCTs across the country. Group B consisted of fourteen members, including a

project manager and a simulation consultant. Of the twelve other group members, seven of the members worked at the Department of Health (DoH) and the National Health Service (NHS) in decision-making roles (clinical, statistics and policy), two were PCT commissioners and three were senior academics who were nationally recognised experts in the specific health-care area, one was from the third sector and one from the National Audit Office. The group met six times across six months, which we observed.

The difference between the groups was their prior experience. In the local group (A), the members worked in different parts of the *local* health-care system, mainly occupying operational or managerial positions, and therefore had a 'grass-roots' view of the service. In group B, the members worked in research, in academia, policymaking and DoH; therefore they all had a 'helicopter' view of the problem and service.

4 DISCUSSION OF CASE FINDINGS

4.1 Group A: from case to model-based reasoning

The composition of group A shaped the model building process in three respects. First, most members were not familiar with modelling. Therefore, they did come to the project with expectations: '*we were more open-minded because we didn't know what to expect' (commissioner)*. Second, because members worked in different operational roles in different parts of the service, they had diverse perceptions of the problem based on their individual experiences. Some participants focused on public health measures and prevention while others cared mostly about service provision and treatment. The client's objective was the reduction in admissions. This was the first in a series of conflicts.

The members struggled with the fact that the problem was about reducing hospital admissions; eventually a more holistic point of view won them over, and they agreed to the "storyline" that preventive and treatment services are complementary: *'... I think the link in with the different services was helpful because it allowed them to understand where we could action things from the front' (manager 1); '...I can see in the grand scheme of things where we do fit in' (manager 2)*. In addition, the majority of members had a shorter-term perspective in mind than the one needed for the problem set to them: *'....what we're looking at there was a project that went well into the future' (manager 1)*.

During the construction process, the group members also disagreed about the definition of the concepts used: ' … *we did spend a lot of time talking about definitions of the stocks, key stocks in the model, and that was important because everybody had come in with different definitions about how best to define the groups of patients - there was a national definition which differed from the PCT definition and that was part of the debate ... there were two meetings, at least, before we determined what the stocks finally were …*' *(consultant);* '*… the group had a different definition to every word that you could possibly come up with*' *(data manager)*. Because their backgrounds varied significantly, the members did not share common interpretations of the same issues. In addition, the discussion was mainly among NHS participants, while the third-sector service providers felt left out of this discussion: '*… A lot of it went straight over my head, I'll have to admit because the things they talked about… I understood the principles of what they were saying but a lot of it… that the NHS are talking to each other so they understand what they're talking about ... NHS talk' (manager 2)*.

The diversity of group members' positions was a challenge to unite these fragmentary snippets and abstract the "bigger picture": '*…., they were little in the big picture. They had their views and were based on personal experience rather than on proven science*' *(commissioner)*. The group members mistrusted more abstract and conceptual explanations and they insisted that the model structure should be made more detailed. Therefore, the consultant decided to adopt an inductive case-based approach to construct the model structure from details in their narratives.

During the reification process, similar issues became visible. The members needed to find the data to parameterise the model and start testing it. There was a problem with finding available data sources, and the data were not complete, so this stage was protracted: '*… it was difficult for me to sit there as a local manager trying to link in your facts and figures with our facts and figures, because I didn't have those figures in hand'* (*manager 4*). Some of the data was less robust: '*That research basically sat on one of the clipboards in one hospital ten years ago…' (commissioner)* — so there was

doubt about the accuracy of the model as a representation of what really happens. This doubt made the members sceptical, reinforced their disbelief in the abstract view of modelling. However, chasing data made them keep asking questions, and they continued in cycles, immersing into a re-examination of definitions as they tried to distil a coherent story *' ... I think that's why we learnt so much because we had to keep on looking for data ...' (PCT manager)*. By this time, the group started to see a shared story, even if they acknowledged that this story might not be completely accurate, but at least plausible and coherent, which encourage them revise their previous anecdotal beliefs.

The group members also encountered a series of 'surprises' that challenged their prior beliefs. It was both a confusing and illuminating turning point for the members when experiments shocked their assumptions. The model started to have a voice, it was as if it kept on saying to them to not only look for data, but also to re-think how the system works and which interventions mattered. Some people started to feel that they did not really know the answers anymore, '… *they thought oh, actually not, maybe our assumptions were wrong, and they later changed their approach,' (PCT manager),* and started asking themselves why '*...there was a little bit of why, because when we could see that, the effects of some of the things which we may have logically thought would have had an effect, didn't, or because something else did at some point' (data manager)*. This was a recursive process: as they used more case based reasoning they changed the model, which in turn gave them results they did not expect, triggering more model based analysis: '*... I saw their understanding change the resource for certain things. They obviously see how it changed the model ...' (manager 2)*. An interesting dialogue emerged in which people were speculating in terms of what the model would *say in response to their explanations: in this "dialogue" the model triggered a change in the way people made sense*.'

The first surprise was when they confronted the fact that the service is more complex than they originally thought. Reflecting on this experience, one participant considered '*Is it the fact that it highlighted that actually it's a lot more complicated than what we think? You know, we think... working in this field, we think it's quite simple' (manager 3)*. Initially some had questioned the value of modelling for what they perceived to be a simple system *'... I can see them working on this sort of work with, you know, like BP and British Airways. Do you know what I mean? I can see that, but for the town, this little process of getting people into treatment services isn't that complicated…' (manager 4).* This prior expectation was challenged when members saw that patients go in and out of remission and therefore go into cycles within the service, mixing across pathways that were initially presumed to be linear. The consultant used the analogy of the road network to describe the phenomenon: '… *It surprised me, so then it would surprise them even more, ... you tend to stay within a cycle. It's very, very difficult to get out of it completely. The only way out of that true cycle is to die…' (data manager)*. This insight of the complex nature of the issues generated a lot of discussion on '*... where they were, what they were doing and what needed to be done' (manager 5)*.

The second surprise was whether prevention or treatment services would have a greater impact on admissions. The model challenged the presumption that established treatment services were effective. The model showed that the key intervention that could work long-term to reduce admissions of chronic patients was a public health measure: *'...when people started to see the effect when changing some of the parameters, some of the things you might have expected to make a huge difference might have made a difference in the beginning but then levelled out. I was very sceptical about the whole thing, but when the model showed what the effect was, it was quite surprising to me to see that' (data manager)*; '*We looked at a whole range of interventions, both proactive and reactive, ... and nothing really made too much difference. That was another thing that was worth learning, actually ....'* and *'.... they got out of that the single most effective thing they could do is (identified specific policy intervention). ... There's quite a bit of resistance to that theory.*' *(consultant)*. Naturally, there was much resistance to this revelation. Interventions that were taken for granted to be effective suddenly did not seem to work.

The third surprise caused strong arguments because the model highlighted the importance of the chronic, long-term side of the condition as the main cause of admissions. Previously all the group members believed the cause to be short-term, acute issues. The model changed their stories – and led them to think in a model-based way instead of trusting only their prior cases.

The effects of the experiments were so dramatic that group members questioned again the accuracy of the model structure and the adequacy of the data. Whenever they saw surprising results,

they went back to almost the beginning to redefine parts of the model structure: '…*understandings were actually growing because each time they put forward ideas… they went back to almost the beginning on a lot of occasions, you know…' (manager).* In the final analysis, many of the participants understood the value of model-based abstraction; the truth in the model lies in the way of perceiving the service, not in the results: *'... It is not so much an accurate tool but a way to think about things –projecting…' (manager 1).* This iterative process, changed the way they made sense: '*... but the most useful part of the whole process wasn't what we had at the end, it was the conversations we had along the way and the extra research it made us go out to do to find out answers to questions and ask more questions along the way, helped everybody understand it a lot more. Every time you come up with an answer, you end up with another question, which is good for research' (PCT manager)*.

In the final presentation to the senior management in the client, the model surprised them as well. However, because the Board members did not participate in the process, so the model had no effect as an agent on their priors and reasoning: '*... Because the slides were demonstrating there was a weakness in an area and she (senior manager) was saying, no, we've invested quite a lot of money in here and that those figures can't be right. I don't think the presentation captivated their attention enough' (manager 3)*. The senior management questioned the validity of the model.

4.2 Group B: from model to case-based reasoning

Most of the policy, academic and NHS people in group B were acquainted with each other's work and had interacted previously (e.g., through conferences). Most of the group members had a broad overview of the relevant issues and deep knowledge of scientific evidence around different interventions: *'...you know, it's something we've spent our lives thinking about' (clinician)*. In addition, many had previous experience in modelling the health service using other types of tools. They all had an abstract, 'helicopter' view of the service in question. The modelling consultants felt that a more model-based approach was appropriate for this group because there was a broad consensus already about the system and underlying concepts and data was available to or even known by the participants.

Despite their common ground, there was nevertheless disagreement regarding the problem definition: '… *the debate was about the brief, not about the model*' *(consultant)*. Their disagreement was whether the model should use hospital admissions as the outcome measure. It was difficult for them to set priorities because the policy agenda was ambiguous: *'... the problem probably was the case that this is still an emerging piece of work; a lot of hard work has gone into developing the objective so far, but it is still developing ....' (consultant)*. An important factor was the complexity of NHS and the transition towards a system with more private provision: '*... So, I think there are problems because we don't really know what sort of healthcare system we've got at the moment. We seem to be, you know, in this transition from the NHS as we all knew it, and a new kind of private sector, or at the moment it's a sort of quasi private sector with the doctors management trying to article over the cracks, I think' (lead clinician)*.

The construction phase started on model based reasoning. First, consultants sketched an initial model structure. They narrowed down the structure to fit the service. However, *'... narrowing the scope just to the NHS service did not paint the full picture of what other services in the rest of the system can contribute' (policy analyst)*. There were objections to this process by academics and clinicians who thought that the process should be more exploratory and that this approach resulted in an explanation that was too linear and too simplistic. Others expected a fast process that would produce a forecasting tool for commissioners and also disagreed with the deductive process: *'....start with a simpler explanation rather than the more complex to the simple' (economic advisor)*. Some thought *'...that the group understanding of the modelling process, and the group's input at discussion was out of sync' (clinician)*. The objectives of the group became contentious in the construction phase prolonging it; such that it took three or four meetings to come up with a final model structure.

The model structure produced was highly abstract. Initial simulation experiments predicted costs which seemed excessive to the group members. For this reason, they did not trust the model and assumed it was overly simplistic. They then decided to alter the model structure and to disaggregate some of the services through testing.

During reification, several data issues emerged. While the group had access to databases from the DoH and academic research, some of the data was based on limited evidence or not available on a local level, or inconsistent throughout the country. This lack of robust data raised questions: '*does the data exist; is the data correct; is the data suitable; is the data utilised properly'* (*statistician DoH*).

Questioning the data led to much creativity during reification and, generally, this group was experimenting with the model very actively. During experimentation, the participants faced the fact that the model commanded mental discipline constraining certain ways of thinking. One participant described this as the model exerting a '*gravitational force because of the focus it commands*', although '*sometimes the focus on building the model overshadows the effort to learn*' *(academic).* Most group members found experimenting with the model stimulating: *'... people would physically get up off the table and sort of go and point at the screen and say you know, "why?" (consultant)*. Testing was *'... Absolutely fantastic! It was an eye opener to me anyway ... useful in understanding the consequences of potential policies' (clinician) – '... Yes, yes, yes, most definitely, most definitely. I think being allowed to play with ideas, to play with sort of 'Monopoly' money if you like, well you know... It is primarily a creative tool' (economic advisor) – '...Yeah absolutely, it's the ultimate ''flight simulator'' for the public sector; you have a risk-free environment in which to play out policy changes and see whether the indications are going to work' (associate)*. Some members arranged individual meetings with the consultants to obtain more guidance for testing the model by themselves.

This is the point where people used scenarios from their own experience (case-based). They had to *'... choose evidence, challenge it and make extrapolations from it '*(*academic*); they had to challenge and let go of what they thought they knew. It was enjoyable for most and created new surprising insights. A surprising insight was to discover which services were actually the most expensive. This discovery *'created a shift, and people directed the shift within the model' (commissioner 1)*. After the experiments they tweaked the model structure and participants gained more confidence: *'... I felt reasonably comfortable, primarily because... what was coming out the work made intuitive sense, and so from a purely sense-check perspective, you know, it held true ...' (DoH statistician)*. It took two months to digest what the model was saying about the treatment, but during this time, the model generated a lot of discussion and a lot of curiosity. Although it was difficult to structure initially, the model helped them to think in a more sophisticated way: *'... it's about helping people be more sophisticated about the treatments...' (policy analyst)*.

The members of this group seemed more comfortable when the model surprised them disputing their priors. They responded by adopting a more case-based approach in the reification phase. They also faced new questions regarding how to convey these messages to local NHS organisations. However, the members were satisfied that the model provided a focus towards the right changes in commissioning and that the model added to the body of knowledge instead of just implementing it. Most importantly, people understood that problem solving is not about using a forecasting tool that provides estimates but, rather, the way they perceived a complex system.

5 CONCLUSIONS: THE ROLE OF THE SD MODEL AS AN AGENT

Regarding our initial question: "Can an SD model be conceptualized as exerting agency in group-model building, by influencing the reasoning of the group members – and if so, how?" the answer is that we observed in these groups that the SD model did exert this influence. The artefact changed the type of reasoning that prevailed in each group (from case to model reasoning and vice versa) when they enacted with the model.

Regarding the way this happened, we observed that the agency of the SD model influenced the reasoning of the groups through a series of 'shocks' or surprises that were resisted initially by the groups. When the first round of testing did not provide expected results, both groups changed the initial model structure until during experimentation they finally felt comfortable using a different type of reasoning. Therefore, conflict of expectations with model experimentation is the point where the artefact changed the mode of reasoning of group members and became an agent. When their expectations got out of the way, their stories changed. During this process, the model changed the way in which they made sense of the problem and how they understood the healthcare service.

These shocks or surprises activated the SD model as an agent. We saw that the artefact had an active role by enabling the groups to rethink events, agents, goals, ideas and results together in a

relational and temporal sequence. We saw how the model questioned definitions about the problem and assumptions about the system and the reliability of people's stories through constraining their assumptions (Jarrahi and Nelson, 2018). The model imposed its own syntax and constrains on the process of learning and exploration, but it also broadened a more holistic view of the problem by forcing people to ask questions they would not before, think in alternative ways and make sense of elements that they had not considered (Leonardi and Barley 2008). The model was not just a medium but became an agent in creating these messages, responding to ideas and initiating these conversations. Therefore, we saw how the SD model brought the fragmented stories together.

Finally, we observed that people who did not participate in the model building process and did not have the experience of interacting with the model and trained to think differently, were not able to accept and comprehend its results. However, this does not in itself mean that the model was embodying the "truth" which non-participants could not understand. Indeed, it might be the case that the insights the participants gained from the model were misleading and that doubts about the accuracy of the model were appropriate. Our study as a study of the modelling process needs to be agnostic in regards to this question – we saw that the model changed the understanding the beliefs participants had about the reality and helped them to integrate different pieces of information and knowledge held by different stakeholders. We could observe that the participation in the modelling process led to the consideration of additional information and that the results were in the main judged to be convincing by those involved, but a scientific assessment of the accuracy of the model would require a different type study; a study of the process of modelling and the role of the model in this process cannot arrive at conclusions about the accuracy of the model. What we can conclude is that simulation models can be conceptualized as agents in group model building processes where the group is tasked with analysing complex systems. The model as artefact changes how group participants deal with ill-structured, fuzzy problems and a diverse array of confusing information and how they gain confidence in their (model derived) understanding.

REFERENCES

Alderman, N., Ivory, C., McLoughlin, I. and Vaughan, R. (2005). Sense-making as a process within complex service projects. *International Journal of Project Management* ***23***: pp 380-385.

Barrett, M, and Oborn, E. (2010). Boundary object use in cross-cultural software development teams. *Human Relations* ***63*(8)**: pp 1199-1221.

Bluhm, D. J., Harman, W., Lee, T. W., and Mitchell, T. R. (2011). Qualitative Research in Management: A Decade of Progress. *Journal of Management Studies* ***48***: pp 1866-1891.

Blumer, H. (1969). *Symbolic interactionism: perspective and method.* NJ: Prentice-Hall.

Bresnen, M., Goussevskaia, A., and Swan, J. (2004). Embedding New Management Knowledge in Project-Based Organizations. *Organization Studies* ***25***: pp 1535-1555.

Carlile, P. R. (2002). A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. *Organization Science* ***13*(4)**: pp 442-455.

Chongthammakun, R., and Jackson, S. J. (2012, January). Boundary objects, agents, and organizations: Lessons from e-document system development in Thailand. In 2012 *45th Hawaii International Conference on System Sciences*, (pp 2249-2258). IEEE.

Cochran, J. K., Mackulak, G. T., and Savory, P. A. (1995). Simulation Project Characteristics in Industrial Settings. *Interfaces* ***25***: pp 104-113.

Dodgson, M,. Gann, D., and Salter, A. (2007). The impact of modelling and simulation on engineering problem solving. *Technology Analysis Strategic Management* ***19*(4)**: pp 471-489.

Eriksson-Zetterquist, U., Lindberg, K., and Styhre, A. (2009). When the good times are over: Professionals encountering new technology. *Human Relations* ***62:*** pp 1145-1170.

Ewenstein, B., and Whyte, J. (2009). Knowledge practices in design: The role of visual representations as 'Epistemic objects. *Organization Studies* ***30*(1)**, pp 7-30.

Garrety, K., Robertson, P. L., Badham, R. (2004). Integrating communities of practice in technology development projects. *International Journal of Project Management* ***22***: pp 351–358.

Glaser, B. G., and Strauss, A. L. (1967). *The discovery of grounded theory: strategies of qualitative research*. New York: Aldine.

Graham, D., Smith, S. D. and Crapper, M. (2004). Improving concrete placement simulation with a case-based reasoning input. *Civil Engineering and Environmental Systems* ***21*****(2)**: pp 137-150.

Hernes, T., and Maitlis, S. (2010). *Process, sensemaking, and organizing*. Oxford: OUP.

Hoegl, M., Weinkauf, K., and Gemuenden, H. G. (2004). Inter-team Coordination, Project Commitment, and Teamwork in Multi-team RandD Projects: A Longitudinal Study. *Organization Science* ***15***: pp 38-55.

Hovmand, P. (2014). *Community Based System Dynamics*, Springer, 2014.

Jarrahi, M. H., and Nelson, S. B. (2018). Agency, sociomateriality, and configuration work. *The Information Society* ***34*****(4)**: pp 244-260.

Jensen, C., Johansson, S., and Lofstrom, M. (2006). Project relationships – A model for analyzing interactional uncertainty. *International Journal of Project Management* ***24***, pp 4–12.

Kallinikos, J (2005). The Order of Technology: Complexity and Control in a Connected World. *Information and Organization* ***15*****(3)**: pp 185–202.

Knuuttila, T., and Voutilainen, A. (2003). A Parser as an Epistemic Artefact: A Material View on Models. *Philosophy of Science* ***70*****(5):** pp 1484-1495.

Koskinen, K. U., and Makinen, S. (2009). Role of boundary objects in negotiations of project contracts. *International Journal of Project Management* ***27***: pp 31–38.

Leonardi, P. M., and Barley. S. R. (2008). Materiality and change: Challenges to building better theory about technology and organizing. *Information and Organization* ***18*** **(3)**: pp 159–76.

Maitlis, S. (2005). The Social Processes of Organizational Sensemaking. *The Academy of Management Journal* ***48***: pp 21-49.

Orlikowski, W. J. (2007). Sociomaterial practices: exploring technology at work. *Organization Studies* ***28***: pp 1435-1448.

Papadimitriou, K., and Pellegrin, C. (2007). Dynamics of a project through Intermediary Objects of Design (IODs): A sensemaking perspective. *International Journal of Project Management* ***25****:* pp 437-445.

Simon, H. A. (1978). Rationality as Process and as Product of Thought. *The American Economic Review* ***68***: pp 1-16.

Sapsed, J., and Salter, A. (2004). Postcards from the edge: local communities, global programs and boundary objects. *Organization Studies* ***25***: pp 1515-1534.

Swan, J., Bresnen, M., Newell, S., and Robertson, M. (2007). The object of knowledge: the role of objects in biomedical innovation. *Human Relations* ***60*****(12)**: pp 1809-1837.

Tako, A. A., and Robinson, S. (2010). Model development in discrete-event simulation and system dynamics: An empirical study of expert modellers. *European Journal of Operational Research* ***207*****(2)**: pp 784-794.

Tako, A. A. (2008). *Development and use of simulation models in Operational Research: a comparison of discrete-event simulation and system dynamics* (Doctor of Philosophy), University of Warwick Retrieved from University of Warwick institutional repository: http://go.warwick.ac.uk/wrap/2984

Urban, G. L. (1974). Building Models for Decision Makers. *Interfaces* **4**, 1-11.

Yakura, E. K. (2002). Charting time: Timelines as temporal boundary objects. *Academy of Management Journal* ***45*****(5)**: pp 956–970.

AUTHOR BIOGRAPHIES

MARIA KAPSALI is a senior lecturer in the Business School, Manchester Metropolitan University. Previously, she was a Lecturer in Management Systems (specializing in Project/Operations Management) at Hull University Business School, a researcher and lecturer at Umeå University and a research associate at Imperial College Business School. Her interest is in project management, in particular the coordination of groups of creatives in innovation projects from a socio-technical systems perspective.

STEFFEN BAYER is Lecturer in Business Analytics at the University of Southampton and Programme Leader of the MSc in Business Analytics and Finance. Previously, he worked as Assistant

Professor in the Program in Health Services & Systems Research at Duke-NUS Medical School in Singapore, as well as a research fellow at Imperial College London and at the University of Sussex. His main research interest is the planning of health services using simulation modelling. He uses a variety of approaches in his research including qualitative research, system dynamics and discrete event simulation.

SALLY BRAILSFORD is Professor of Management Science at the University of Southampton. Sally obtained a BSc in Mathematics and then worked for several years as a nurse in the NHS before obtaining an MSc and then a PhD in Operational Research from Southampton. Her research is in the area of healthcare simulation modelling: to evaluate treatments and screening programmes, or to redesign and improve service delivery.

TIM BOLT is an Associate Professor of Economics at Saitama University, having previously worked in the School of Management and the Faculty of Health Sciences at Southampton University. His research focuses on health economics, discrete choice experiments and stated preference valuation methods.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.036

AGENT-BASED SIMULATION OF HOUSEHOLD BANDWIDTH DEMAND

Mathias Kern, Russell Ainslie, Doug Williams, Arnaud Jacquet, Nicola Wong

Applied Research, BT
Adastral Park, Martlesham, Ipswich
mathias.kern@bt.com, russell.ainslie@bt.com, doug.williams@bt.com, arnaud.jacquet@bt.com, nicola.wong@bt.com

ABSTRACT

In this paper we present a snapshot of our ongoing research into an agent-based model for simulating bandwidth demand at the household level. Motivated by Edholm's law on the exponential growth of internet traffic, our approach models households, persons and devices and their internet traffic generating interactions, and builds a picture of overall total and peak demand. Assumptions on the types, likelihoods and bandwidth requirements of various activities are key inputs to the simulation, and will allow us to test different hypotheses on future growth in the next phase.

Keywords: Agent-Based Simulation, Internet Traffic, Bandwidth, Household Model.

1 INTRODUCTION

The internet has seen phenomenal growth in users, traffic and applications over the last decades, and has now reached almost every aspect of human life. Figure 1 illustrates the exponential increase in global IP traffic since 1990 (Wikipedia: https://en.wikipedia.org/wiki/Internet_traffic, accessed 18 September 2019). This observation has led to the formulation of Edholm's law (Cherry 2004) which states that the required traffic data rates double every 1.5 years. However, internet traffic cannot continue on this growth trajectory forever, and a slow-down is expected at some point. This, of course, poses a significant challenge to businesses such as internet service providers, network infrastructure companies and network equipment producers. These businesses need to forecast likely future traffic and plan ahead and invest accordingly. In this paper, we introduce an agent-based model for the simulation of bandwidth demands at the domestic household level to draw conclusions about overall network bandwidth demands. We describe the makeup of the model, state key modelling assumptions, and discuss the current state of our on-going work and next steps.

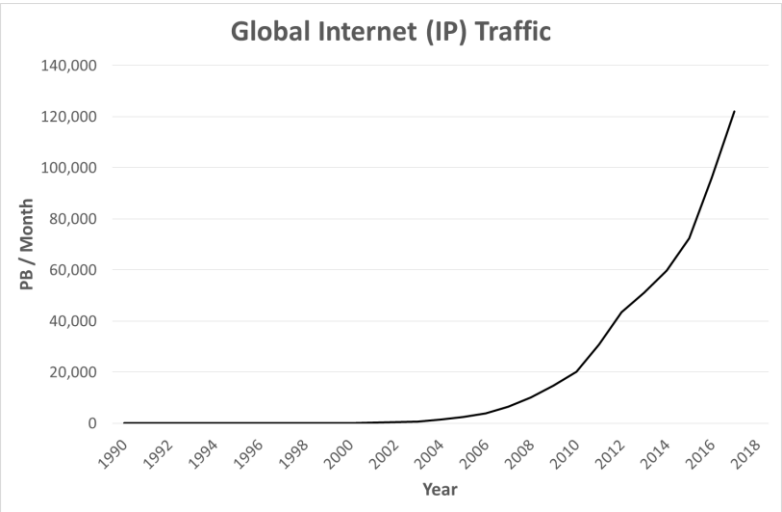


Figure 1 *Global internet (IP) traffic by year in Petabyte / month*

2 AGENT-BASED SIMULATION APPROACH

In order to approximate and predict overall and peak internet traffic demand for domestic consumers, an agent-based approach, a method applied successfully to many scientific domains (Niazi and Hussain, 2011), has been chosen to simulate traffic patterns generated by residential households. We model households and people and devices within these households as agents, and simulate key interactions between them, specifically activities that generate internet traffic.

Key modelling considerations for our household bandwidth simulation include:

i. The types of people modelled such as 'young adult (18-34 years)' or 'retiree (65 years+)'.
ii. The types of devices simulated, e.g. 'personal device' or 'generic household device'.
iii. The types of households, for example. 'couple (45-54 years) with no children' or 'single adult (35-44 years) with two dependent children', which describe the make-up of the household.
iv. The types of network connection available to a household, for instance 'fibre broadband' or 'copper broadband'
v. The categories of activities associated with people such as 'watch', 'game', 'listen', 'communicate' and 'read', and specific activity types such as 'watch paid video-on-demand' or 'online gaming'
vi. The generic activity types associated with household devices such as 'background activity' or 'download of update'.

These choices form the foundations of the household bandwidth model, and allow us to specify different scenarios. Each simulation scenario is composed of a number of partial or mini-scenarios, and these can be easily re-used and re-combined to form new overall scenarios. Below is a list of some of the central configurations for constructing a scenario:

i. The number of households to simulate.
ii. The relative frequency of different household types which facilitates sampling of households.
iii. The relative frequency of the different network connection types for each household type.
iv. The internet traffic and speed characteristics associated with each network connection type.
v. Bandwidth requirements for each activity type, including start-up requirements (e.g. when starting to stream a movie online) and on-going requirements (e.g. to maintain the on-going streaming of a movie).
vi. The probabilities of starting a new activity by a person or device, by hour and minute of the day and by day of the week, individually for each person and device type.

The simulation steps through the 7*24*60=10,080 minutes of a week minute by minute. In each step, we check for each person and device agent whether the agent has already an activity running, and/or whether a new activity should be started. For example, we allow person agents to have up to two activities live in parallel, e.g. watching an online video while also communicating at the same time, and decide stochastically whether and what activities to start. People and device agents send bandwidth requests to their respective household agents, which in turn might either fully satisfy or cap the actually allocated bandwidth, depending on whether the overall requested bandwidth is within or exceeds what the household's network connection can support. We collect activity and bandwidth data for each person, device and household agent for each minute cycle, and compute overall traffic, peak traffic and whether an agent's traffic is capped.

Three important aspects have to be combined to make the simulation truly functional. Firstly, the agent-based model itself, i.e. the 'mechanics' of the simulation, needs to be developed. Secondly, key scenario assumptions and inputs need to be compiled, e.g. information about the likelihoods and bandwidth requirements of different activities. And thirdly, actual bandwidth demand for a network of residential customers needs to be collected to facilitate the setup and tuning of the model and ensure that simulation results and real life observations match closely.

3 CURRENT STATE AND FUTURE WORK

The work on an agent-based model for the simulation of household bandwidth demand is an active area of research and development, and this paper presents a current snapshot of our thinking and progress. Although we initially considered and experimented with implementing the simulation in AnyLogic (https://www.anylogic.com, accessed 11 September 2019), the current solution is implemented in Java to allow for more flexibility and easier development and sharing across teams. Furthermore, we have compiled a list of scenario configurations and assumptions, and they combine information from a diverse range of data sources, including the Office for National Statistics (ONS) for trends on households (ONS: Families and households, 2018) and the Office of Communications (Ofcom) for trends in internet usage (OFCOM: Internet use and attitudes, 2019). We have also sourced current actual aggregate network bandwidth usage statistics, which mirror the patterns previously published in (Lord et al, 2016) shown in figure 2. This has allowed us to run simulations for recent/current bandwidth use scenarios, and simulation results and actual observations match well. In summary, we have developed an agent-based household bandwidth simulation model and can closely replicate current real life observations with it.

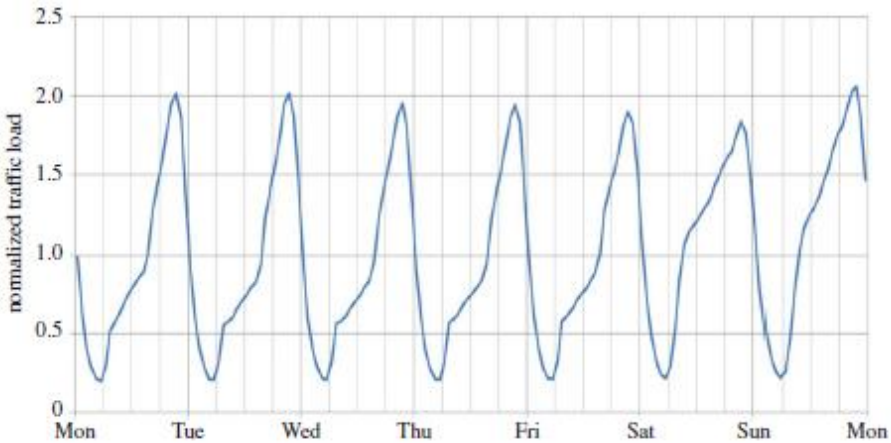


Figure 2 Typical distribution of traffic load across the week on a residential fibre network (1.0=average load) (Lord et al, 2016).

The next phase of our research and development work will address the following points:

- Support parallel running of simulations across a distributed computing infrastructure to facilitate large scale experiments.
- Build assumptions and data for future scenarios, including:
  - Do we anticipate completely new, as of now unknown, activities, for instance in the augmented or virtual reality space, or in the internet of things domain?
  - What are the possible bandwidth requirements for individual activities, e.g. for streaming online videos, over the next 10 to 20 years?
  - Will certain activities become more or less frequent going forward, e.g. with the move from terrestrial TV to online streaming?
  - Will we see a change in the online behaviour of certain age groups, i.e. person types, e.g. do we anticipate retirees to become more frequent internet users?
  - Do we expect a social shift in the make-up of households over the next decades, e.g. do we foresee a higher proportion of single person households?
- Run simulations for likely / best-case / worst-case future scenarios.
- Ultimately, provide guidance for answering the central question: Do the simulation results indicate further exponential growth of network traffic, or do they point to a slowing of the growth rate?
- Overlay the household model with different network topology assumptions such as full fibre.

REFERENCES

Cherry, Steven (2004). Edholm's law of bandwidth. *IEEE Spectrum*. 41(7): 58–60, doi:10.1109/MSPEC.2004.1309810.

Lord A., Soppera A., Jacquet A. (2016). The impact of capacity growth in national telecommunications networks. *Philos Trans A Math Phys Eng Sci.* 2016 Mar 6;374(2062).

Niazi, Muaz and Hussain, Amir (2011). Agent-based Computing from Multi-agent Systems to Agent-Based Models: A Visual Survey. *Scientometrics.* 89 (2): 479–499.

OFCOM: Internet use and attitudes (2019). https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/internet-use-and-attitudes, accessed 18 September 2019.

ONS: Families and households: 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2018, accessed 2 September 2019.

AUTHOR BIOGRAPHIES

MATHIAS KERN is a senior research manager in BT's Applied Research team, and his research interests focus on the application of AI, optimisation and simulation techniques to real-life resource management problems. He gained his PhD in Computer Science from the University of Essex researching population-based optimisation methods, and has been with BT since 2004.

RUSSEL AINSLIE is a research specialist in AI and optimisation at BT Research whilst completing a PhD in large-scale optimisation problems at Robert Gordon University. Starting from a background of a BSc in Maths and Physics from Edinburgh and a BSc and MSc in Computer Science from RGU, Russell has developed specific expertise in workforce management.

DOUG WILLIAMS is a research manager in BT's Applied Research team, specialising in Multimedia Analytics.

ARNAUD JACQUET is a research manager in BT's Applied Research team, with a particular research interest in the area of quality of experience.

NICOLA WONG received a BSc (Hons) Computer Science from the University of East Anglia in 2016. She is currently a researcher in the Operational Transformation research theme at BT Applied Research.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.037

A TWO-STAGE DUAL THRESHOLD AGENT-BASED MODEL OF INNOVATION ADOPTION

Dr. Arpan Y. Jani

University of Wisconsin – River Falls
Department of Computer Science and Information Systems
410 S. 3rd Street, River Falls WI 54022 USA
arpan.jani@uwrf.edu

ABSTRACT

Studies on diffusion of innovation have found that individuals and firms often do not recognize the benefits of an innovation in the early stages though they may eventually adopt it, which is reflected in the S-shaped diffusion curve. This could be attributed to information about an innovation spreading gradually. Even when ample information is available about an innovation, evaluation of information may be biased in favour of maintaining status quo. This paper presents an agent-based model which focuses on biased evaluation of information. It is assumed that agents have dual thresholds for evaluating competing information: one for maintaining the status quo choice and the other for switching to the alternative. Simulation experiments were conducted by varying the evaluation thresholds and the time frame over which the information was evaluated to investigate its influence on the adoption of an innovation. Results match the S-shaped diffusion curve under certain conditions.

Keywords: Diffusion of Innovation, Agent-based Modelling, Cognitive Biases

1 INTRODUCTION

Even though a new innovation may have clear benefits, it may take a long time for it to be adopted widely. Rogers' theory of diffusion of innovation (Rogers, 1962) identifies five categories of adopters: innovators, early adopters, early majority, late majority, and laggards. Innovators tend to be the first adopters whereas laggards tend to be among the last to adopt a new innovation. Geroski (2000) discusses alternate models of technology diffusion: the epidemic model, the probit model, population ecology based models, and the information cascade model. In the epidemic model it is assumed that the information spreads slowly over time and due to this the adoption rate differs across individuals (or firms) (Geroski, 2000). Probit models consider individual characteristics such as the switching costs or opportunity costs (Geroski, 2000) faced by individuals in adopting a new innovation. Information cascades perspective (Geroski, 2000) suggests that the early adopters of an innovation may go through a serious evaluation of the alternatives available before making a choice. Subsequent adopters may take a cognitive shortcut and be more likely to choose the same alternative instead of going through the extensive cognitive processing (e.g. see Bikhchandani et al., 1992).

While epidemiological models pose that diffusion of innovation is limited by the availability of information, studies in behavioral decision making have shown that even if all the information is available, how this information is evaluated may influence the choices made by individuals. Individuals may exhibit a confirmation bias (Nickerson, 1998) where they seek out information that supports their existing position and avoid information that could contradict their position. From this perspective, once individuals have made a choice, they may be more likely to seek information that justifies this choice even if other information may suggest that a better option is available. Individuals may also exhibit a status quo bias (Samuelson and Zickhauser, 1988) so that they may be more likely to stick to a chosen alternative even when a better choice is available.

Agent-based models incorporate simple decision rules while recognizing that agents make decisions based on the local context e.g. what information is available to them. This paper develops an agent-based model of diffusion of innovation where the focus is not just on availability of information itself but also how this information is evaluated. While gradually more information may be available about an innovation, biased processing of this information may lead to delays in adoption of the innovation. The model can also incorporate heterogeneity among agents in how carefully they evaluate the information and the varying time frame over which the information is gathered and evaluated.

2 A TWO-STAGE DUAL THRESHOLD MODEL OF DECISION MAKING

Threshold models have been used in social and behavioural sciences to explain social behaviour (Granovetter, 1978; Schelling, 1971). In psychology, threshold-based model of decision making have been proposed where individuals are assumed to make a choice once a threshold is reached (Curley, et al., 2018) which provides support for the choice. While a single threshold is simpler to consider in the context of decision making, it may be better to incorporate two thresholds in situations where the decision is to switch to an alternative: whether to maintain or reject the existing choice and to accept the alternate. It is as if individuals maintain dual standards for evaluation of the alternatives: one for the choice already made and the other for switching to a new alternative. This way one can consider the case where individuals may have a lower threshold for maintaining the current choice but may have a higher threshold for switching to an alternative.

Figure 1 presents a two-stage, dual threshold model of decision making. It assumes that an agent may have an initial belief i.e. benefits of product A, which would justify the initial choice made. This belief would have been formed based on the information that was available about product A. However, suppose a new innovation, product B, is also available in the market for which gradually more information becomes available. It is assumed that individuals have two thresholds, $\boldsymbol{\Theta_1}$ and $\boldsymbol{\Theta_2}$. $\boldsymbol{\Theta_1}$ ($0 <= \boldsymbol{\Theta_1} <= 1$) is the threshold of evidence needed for maintaining one's current belief (i.e. product A is the best option available and one should continue to adopt it) and $\boldsymbol{\Theta_2}$ ($0 <= \boldsymbol{\Theta_2} <= 1$) is the threshold of evidence required to change to the alternate belief (i.e. product B is superior and should be adopted). Individuals gather information about the products which provides varying evidence in support of both alternatives. Thus, at any given time, there is relative evidence $\boldsymbol{Ø_1}$ ($0 <= \boldsymbol{Ø_1} <= 1$) that supports the adoption of product A and relative evidence $\boldsymbol{Ø_2}$ ($0 <= \boldsymbol{Ø_2} <= 1$) that supports the adoption of product B.

Agents first compare the evidence $\boldsymbol{Ø1}$ with threshold $\boldsymbol{\Theta 1}$ and if the accumulated evidence clears this threshold, the agents do not need to engage in further information processing and would maintain their current belief (to continue to adopt Product A). This reflects the well-known confirmation bias. However, if there isn't sufficient evidence to support the initial belief i.e. the first threshold is not met, agents would look at the evidence for the alternative. If the evidence for the alternative is not sufficient, the agents will continue to hold the initial belief, thus reflecting the status-quo bias. If the relative evidence for the alternative is sufficient i.e. $\boldsymbol{Ø_2}$ is higher than the threshold $\boldsymbol{\Theta_2}$, the agent will be ready to change its belief. This concludes the first stage which may result in either the agents maintaining their current beliefs or becoming ready to change as doubts have appeared.

In the second stage, the agents who are ready to change their belief again look for evidence to support their initial beliefs and if there is sufficient relative evidence i.e. $\boldsymbol{Ø1}$ greater than the threshold $\boldsymbol{\Theta 1}$, the agent would go back to holding their initial belief. However, if the evidence is not sufficient for their initial belief, the agent will compare the relative evidence $\boldsymbol{Ø2}$ with threshold $\boldsymbol{\Theta 2}$. If the evidence clears the threshold, then the agent will change to the new belief (i.e. product B has superior benefits and should be adopted). If the evidence does not clear the threshold, then the agent would continue to be in the "ready to change" state. Hence, the two-stage dual threshold model represents the resistance to change that individuals have when considering adoption of a new alternative.

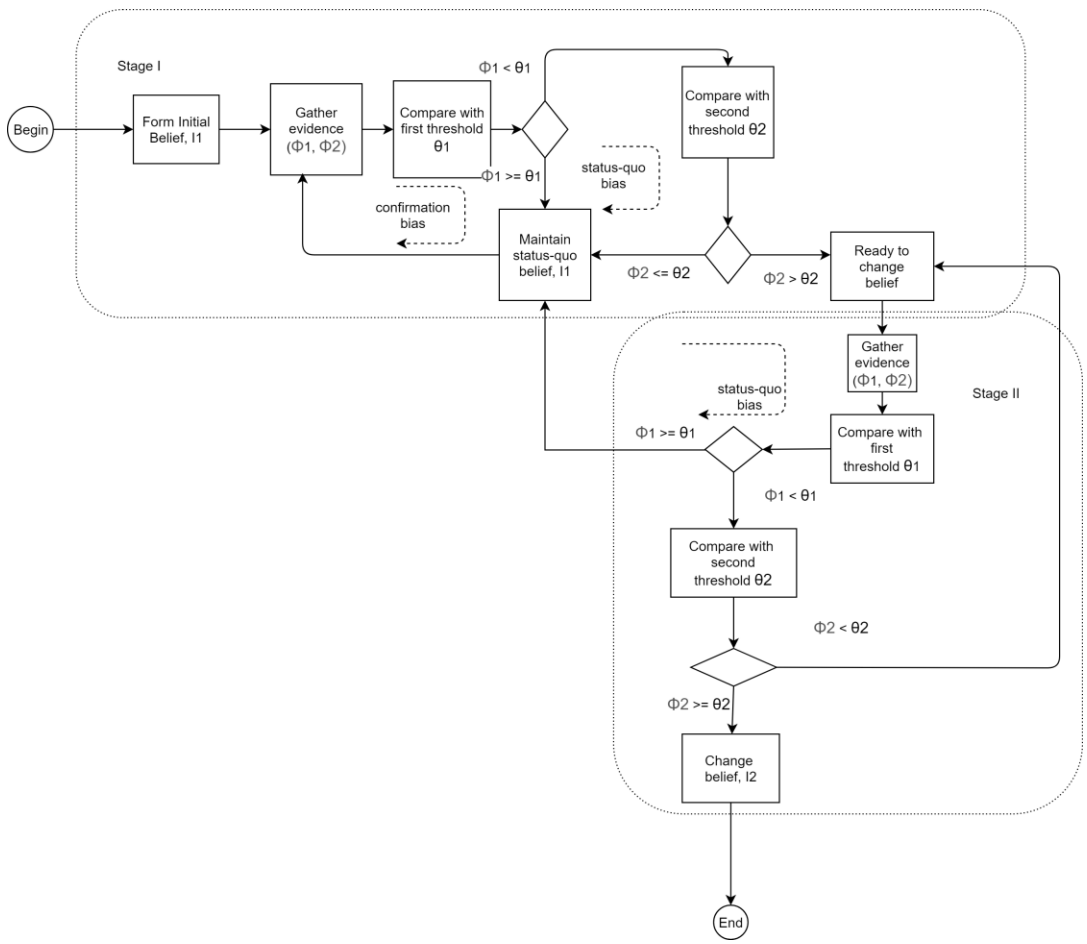


Figure 1 *Two-stage dual threshold model of belief change*

2.1 An Agent-based Model of Innovation Diffusion

An agent-based model of innovation diffusion that incorporates the two-stage dual threshold model of belief revision is presented here. The model was implemented in NetLogo (Wilensky, 1999). A torus grid with 100 x 100 size which wrapped both horizontally and vertically was used. This represented 10201 patches, with each patch representing a piece of information. The patch could be of one of two colours. The patch colour represented a piece of information that either supported belief A (e.g. Product A should be adopted) or belief B (e.g. product B should be adopted). A green patch represented information in support of product A and a yellow patch represented information in support of product B. Initially the entire grid was initialized to be green which indicated that all evidence was in support of product A (the current dominant technology). As product B, the new innovation, is introduced into the market, slowly more information becomes available. In each time step, one patch turned from green to yellow indicating that new information about product B was available. Each piece of information had credibility C that varied from 0 to 1, drawn from a uniform distribution. Hence, coming across information with higher credibility will lead to greater support for a particular belief.

C1 is the credibility of a piece of information about initial belief (i.e. product A should be adopted) and ***C2*** is the credibility of a piece of information about alternate belief (i.e. product B should be adopted).

$$0 \leq \boldsymbol{C1} \leq 1 \quad \text{where } \boldsymbol{C1} \sim \text{uniform } (0, 1)$$

$$0 \leq \boldsymbol{C2} \leq 1 \quad \text{where } \boldsymbol{C2} \sim \text{uniform } (0, 1)$$

The relative evidence $\boldsymbol{Ø_2}$ in support of the alternative grows over time since more information is available about the benefits of product B whereas relative evidence $\boldsymbol{Ø_1}$ decreases over time as new information crowds out the existing information.

2.1.1 Model Initialization

2000 agents were created and were assumed to have an initial belief supporting adoption of product A that they had currently adopted product A which was the dominant technology. These agents randomly moved around the grid and collected evidence in the form of information pieces with a certain credibility. The number of patches visited represented the information pieces that the agent came across, each with a certain degree of credibility. In the beginning of the simulation since most patches were green, the agents would only come across information supporting their current choice and hardly any that would support switching to the alternative. However, in each time step, one of the green patches changed to yellow indicating that over time the information supporting adoption of the alternative increased linearly whereas the information supporting the continued adoption of the default choice linearly decreased. Agents evaluated the collected evidence in a decision time frame that could range from 10 time steps to 100 time steps.

Let's say that during decision time frame *t*, an agent *i* has collected *m* pieces of evidence in support of the initial belief (i.e. continue to use product A) and *n* pieces of evidence in support of the alternate belief (adopt new product B). $\boldsymbol{Ø1_{it}}$ is the ratio of the sum total of the evidence collected in time frame ***t*** by agent ***i*** that is in favour of initial belief to the sum total of all evidence collected by agent ***i*** and $\boldsymbol{Ø2_{it}}$ is the ratio of sum total of the evidence collected in time frame *t* by agent ***i*** that is in favour of alternate belief to the sum total of evidence collected by agent ***i***.

Proportion of evidence that supports initial belief

$$Ø1_{it} = \left(\sum_{j=1}^{n} C1_j\right) / \left(\sum_{j=1}^{n} C1_j + \sum_{k=1}^{m} C2_k\right)$$

Proportion of evidence that supports alternate belief

$$Ø2_{it} = \left(\sum_{k=1}^{m} C2_k\right) / \left(\sum_{j=1}^{n} C1_j + \sum_{k=1}^{m} C2_k\right)$$

Each agent has evidence thresholds $\boldsymbol{\Theta_1}$ and $\boldsymbol{\Theta_2}$ for evaluating the evidence for the initial belief and alternate belief respectively. As discussed earlier, if there is weak relative evidence to support the initial belief but there is strong relative evidence to support the alternate belief, the agent will change to the state "ready to change belief." If there is a continued lack of support for the initial belief but strong support for the alternate belief, the agent will switch to the new belief.

2.1.2 Simulation Runs

The BehaviourSpace feature of NetLogo enables running of simulation experiments multiple times while changing the values of its parameters across these runs. The NetLogo simulation had the following three parameters: $\boldsymbol{\Theta_1}$, $\boldsymbol{\Theta_2}$, and Decision Time Frame. Table 1 shows the range of values that were varied across multiple simulation runs. The total runs were 250.

Table 1 *Parameter Values – Simulation Run*

Parameter	Range of Values	Number of values
Threshold 1 (***Θ1***)	0.1 – 0.9 increments of 0.2	5
Threshold 2 (***Θ2***)	0.1 – 0.9 increments of 0.2	5
Decision Time Frame (Ticks)	10 – 100 increments of 10	10
Total Simulation Runs		5*5*10 = 250

2.2 Results and Discussion

Data analysis was conducted using RStudio and R. For data visualization, ggplot2 package (Wickham, 2016) was used. Figure 2 shows the rate of adoption under varying evaluation threshold values and decision time frames. Each line chart in the panel of charts below shows the adoption rate for the combination of ***Θ1*** (Threshold 1) and ***Θ2*** (Threshold 2) values. The rows show ***Θ1*** value varying from 0.1 to 0.9 in increments of 0.2 and the columns show ***Θ2*** values varying from 0.1 to 0.9 in increments of 0.2. Each line chart also shows the adoption rate for various decision time frames from 10 time steps (ticks) to 100 time steps (ticks). Many of the line charts match the classic S-shaped diffusion curve for specific combinations of ***Θ1, Θ2*** and decision time frames.

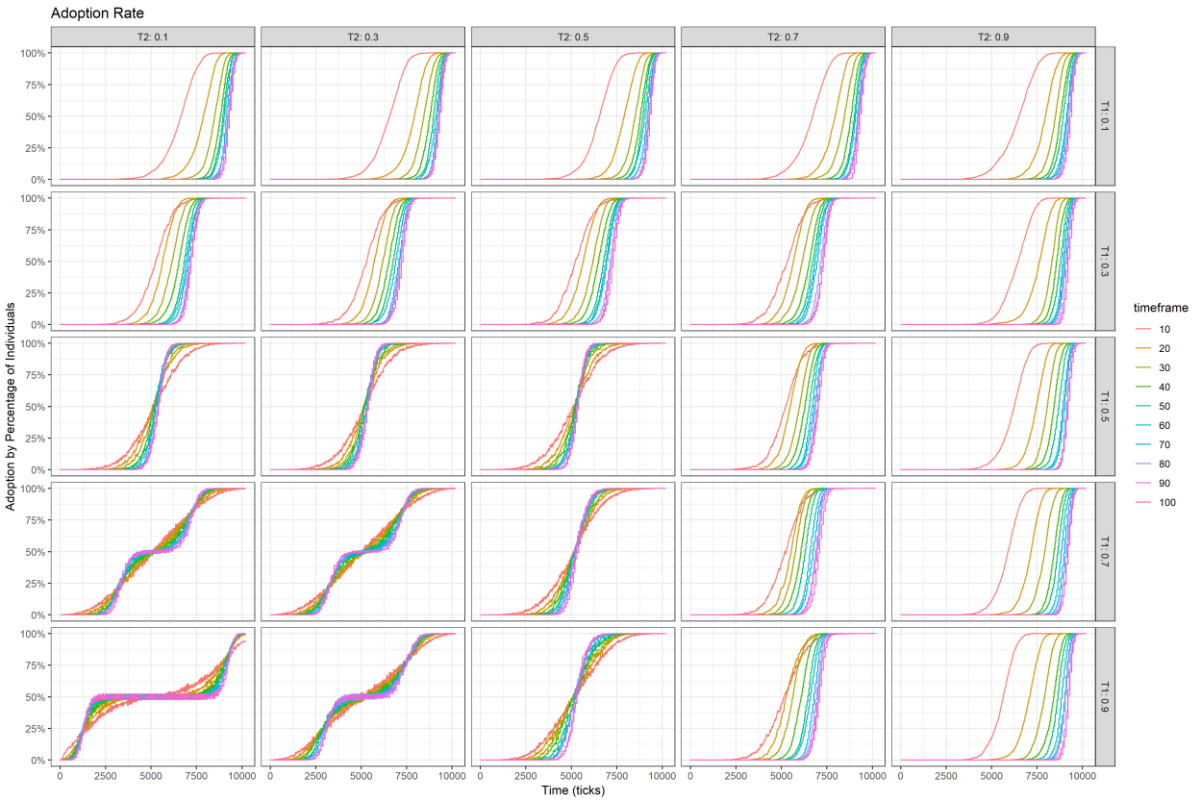


Figure 2 *Adoption Rates With Varying Evaluation Thresholds and Decision Time Frames*

Figure 3 shows the adoption at midway point of the simulation where rows show ***Θ1*** value varying from 0.1 to 0.9 and columns show ***Θ2*** values varying from 0.1 to 0.9. Each vertical column/bar shows the adoption rate for a given decision time frame. The decision time frame varies from 10 ticks to 100 ticks.

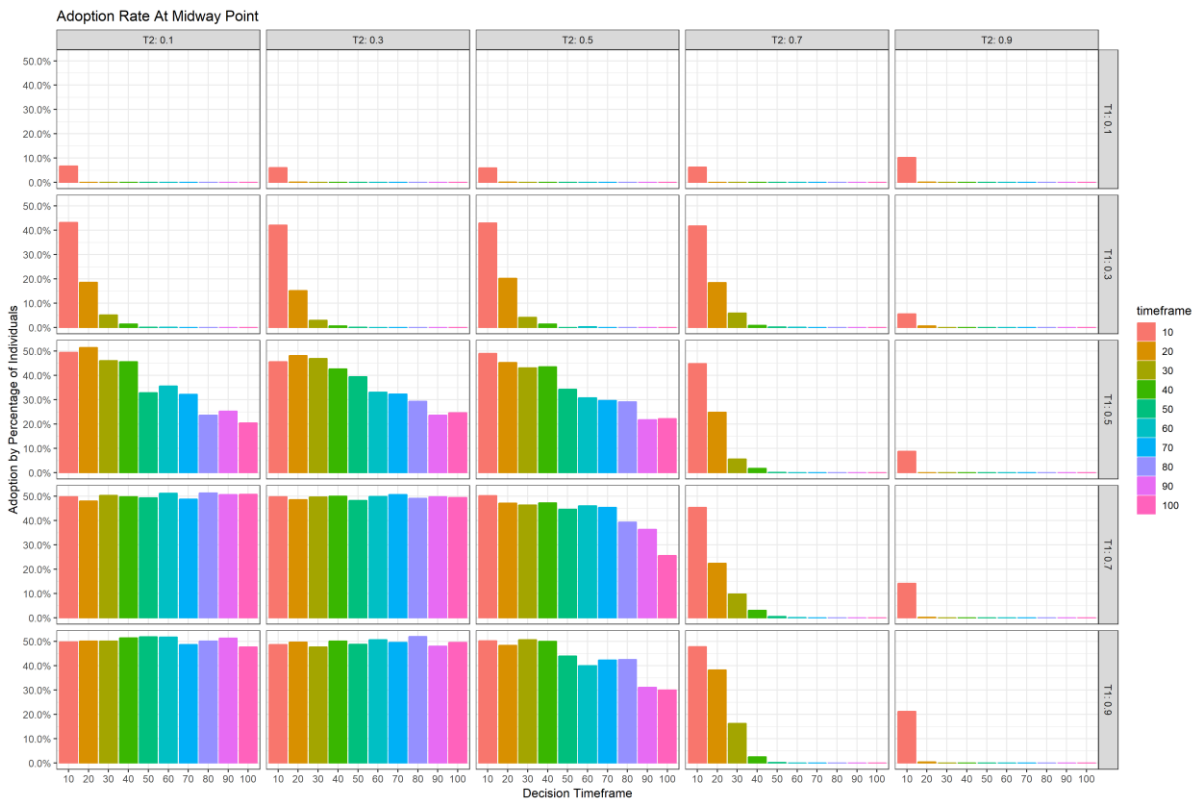


Figure 3 *Adoption percentage at mid-point of the simulation run*

In both figures 2 and 3 above, one can observe that as the decision time frame increases from 10 time steps (ticks) to 100 time steps (ticks), the adoption gets increasingly delayed and the transition to the alternative occurs more steeply.

As the first row of the panel in Figure 3 shows, when ***Θ1*** value is low (***Θ1*** = 0.1), the adoption is quite slow irrespective of the value of ***Θ2***. This is because the standard of evidence needed to support the default choice is quite low. When both ***Θ1*** and ***Θ2*** are in the middle range (***Θ1*** = 0.5, ***Θ2*** = 0.5), adoption rate is the highest with the lowest decision time frame but decreases as the decision time frame increases. As seen in the lower right hand panel of Figure 3, when ***Θ1*** is high (which means that the agents would need a very high evidence in order to continue with the default choice), and ***Θ2*** is low, the decision time frame is not too influential. The lower right hand corner represents high values for both ***Θ1*** and ***Θ2***, which means that one requires a very high evidence standard to either continue with the default choice or to switch. In this case, the adoption rate is also slow.

Figure 4 shows the overall behavior of the model. Even though the available information about an innovation increases linearly over time, the outcome i.e. adoption of this innovation unfolds in a non-linear manner. The shape of the adoption curve is influenced by the two threshold values and the decision time frame.

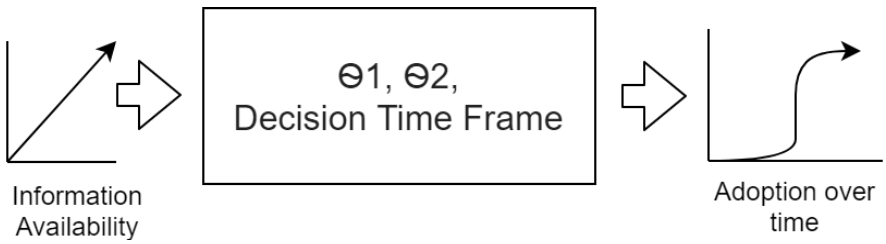


Figure 4 *Input, Output and Key Model Parameters*

3 CONCLUSION

An agent-based model of innovation diffusion is presented here that focuses on the biased evaluation of information available about an innovation. Compared to other models of diffusion of innovation, this is primarily a cognitive, information-processing model. It is assumed that greater evidence in support of the adoption of the innovation is available over time. Biased evaluation of this evidence influences the adoption pattern. The model shows that when there is low amount of information available about the innovation, there is not much adoption but once more information is available, under certain conditions, the adoption rate exceeds what the objective evidence may suggest. That is, everyone in the population adopts the new innovation in the later period as if complete evidence was in its favour. Thus, the adoption rate either lags or leads the availability of information.

REFERENCES

Bikhchandani S, Hirshleifer D and Welch I (1992). A theory of fads, fashion, custom, and cultural change in informational cascades. *Journal of Political Economy* **100(5)**:992–1026.

Curley L J, Maclean R, Murray J and Laybourn P (2018). Decision science: A new hope. *Psychological Reports* **122(6)**: 2417-2439.

Granovetter M (1978). Threshold models of collective behavior. *American Journal of Sociology* **83(6)**: 1420-1443.

Nickerson R S (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* **2(2)**: 175-220.

Rogers E M (1962). *Diffusion of Innovations*. Glencoe: Free Press.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Samuelson W and Zeckhauser R (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, **1**: 7-59.

Wickham, H (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wilensky, U (1999). NetLogo. http://ccl.northwestern.edu/netlogo/.

AUTHOR BIOGRAPHIES

Arpan Jani received his PhD in Business Administration from the University of Minnesota in 2005. He is currently an Associate Professor in the Department of Computer Science and Information Systems at the University of Wisconsin – River Falls. His research interests include agent-based modelling, information systems and decision support, behavioural ethics, and judgment & decision making under conditions of risk and uncertainty.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.038

DYNAMIC COALITIONS IN COMPLEX TASK ENVIRONMENTS: TO CHANGE OR *NOT* TO CHANGE A WINNING TEAM?

Darío Blanco-Fernández

University of Klagenfurt
Klagenfurt, Austria
dario.blanco@aau.at

Stephan Leitner

University of Klagenfurt
Klagenfurt, Austria
stephan.leitner@aau.at

Alexandra Rausch

University of Klagenfurt
Klagenfurt, Austria
alexandra.rausch@aau.at

ABSTRACT

Decision-makers are often confronted with complex tasks which cannot be solved by an individual alone but require collaboration in the form of a coalition. Previous literature argues that instability, in terms of the re-organization of a coalition concerning its members over time, is detrimental to performance. Other lines of research, such as the dynamic capabilities framework, challenge this view. Our objective is to understand the effects of instability on the performance of coalitions formed to solve complex tasks. To do so, we adapt the *NK*-model to the context of human decision-making in coalitions and introduce an auction-based mechanism for autonomous coalition formation and a learning mechanism for human agents. Preliminary results suggest that reorganizing innovative and well-performing teams is beneficial, but this is true only in certain situations.

Keywords:

self-organization, dynamic capabilities, complex and adaptive systems

1 INTRODUCTION

Decision-makers are often confronted with complex tasks, i.e., tasks composed of a certain number of (often highly) interrelated sub-tasks (Giannoccaro et al. 2018). Complex tasks are present in our everyday life, and can, for example, be found in the context of business operations, vaccine development, and construction projects, among many others. Such tasks' innate complexity makes it necessary for individuals to coordinate their efforts and to share their capabilities with other individuals to solve these tasks (Simon 1957). We refer to grouping up into teams as *coalition formation* and exclusively focus on coalitions of human decision-makers.

Previous research has identified stability (in terms of not allowing for the autonomous re-organization of coalitions over time) as a desirable characteristic of coalitions (Hsu et al. 2016). It is argued that stable coalitions come up with solutions to complex tasks associated with a higher *coalition performance* than the solutions of other, more unstable counterparts (Hsu et al. 2016). This is reflected in the motto *never change a winning team*. However, other lines of research argue that the stability in the composition of a coalition does *not* ensure that the coalition comes up with better-performing solutions (see, for example, Hennet and Mahjoub (2011) and Sless et al. (2018)). Given these mutually incompatible findings, it is, thus, unclear whether allowing for the autonomous re-organization of coalitions over time with different frequencies has a negative or a positive effect on performance. Can change make a winning team even better?

Our objective is to contribute to the literature of coalitions confronted with complex decision-making tasks by gaining insights into (i) the dynamics of coalition formation and (ii) the interrelations between the frequency of coalition re-organization over time and performance. The complex task environment is based on the *NK*-model for organizational decision-making (Levinthal 1997, Leitner and Wall 2015, Wall and Leitner 2020). Within this framework, we operationalize instability as the possibility to replace the members of a coalition over time and consider different frequencies at which such a replacement can take place. Coalitions that are allowed to replace their members more (less) frequently are regarded as being relatively more unstable (stable). We observe the solutions to the complex decision-making problem that the coalitions come up with and record the associated performances. Our approach also includes autonomous coalition formation, i.e., we allow human decision-makers to form coalitions by employing a mechanism based on a second-price auction, which is a standard method for self-organization in the field of Robotics (Rizk et al. 2019).

The remainder of this paper is organized as follows: Section 2 places our research endeavor in the context of the relevant literature. In Sec. 3, we introduce the agent-based model. The results of the simulation study are presented in Sec. 4. Finally, Sec. 5 discusses the results and provides some conclusions and an outlook on future research avenues.

2 RELATED LITERATURE

Coalition formation has been extensively studied in the *field of Economics*, see, for example, Banerjee et al. (2001) and İnal (2019). Research in Economics is particularly concerned with the allocation of a coalition's performance among its members, so that the members do not have incentives to exit a coalition and join another one (i.e., they put stability in the focus) (Banerjee et al. 2001, İnal 2019). A set of solution concepts, namely, the *core*, the *kernel*, the *nucleolus* and the *Shapley value* are studied to assure stability (Tremewan and Vanberg 2016).

Since our objective is not to find an allocation mechanism that ensures a stable coalition, our research more likely connects with previous work on task allocation in the *field of Robotics*. Research in this field has extensively studied the allocation of sub-tasks to autonomous robotic entities (such as drones which jointly carry out a specific task) as a mean for coalition formation. In this field, the objective is to find mechanisms that improve the efficiency of solving a task; frequently used mechanisms for coalition formation which have been shown to improve this efficiency are auction-based methods (Rizk et al. 2019).

Task allocation has also been studied in the *field of Managerial Science*, although not as extensively as in Robotics. The ultimate goal in this area is to understand how task decomposition into sub-tasks can improve performance. Researchers have frequently made use of the *NK*-framework for modelling task allocation and decision-making in agent-based systems (Rivkin and Siggelkow 2003, Wall 2018). Note, however, that instability is often not explicitly modelled in this line of research, as it is implicitly regarded to have negative effects on performance (Hsu et al. 2016).

The *dynamic capabilities* framework (Teece et al. 1997) provides a theoretical framework to investigate the potential effects of the (more or less frequent) re-organization of the coalition on performance. According to this framework, both the individual decision-makers (i.e., at the micro-level) and the coalition as a collective of decision-makers (i.e., at the macro-level) must constantly adapt to the *environment* to improve performance (Eisenhardt and Martin 2000). The environment to which individuals and the coalition adapt may be represented by the coalition's *task environment*, like in our case. The task environment is based on the *NK*-framework and is shaped by the complexity of the task that the coalition faces (Giannoccaro et al. 2018).[1] We follow the dynamic capabilities framework and conceptualize capabilities (which are subject to adaptation) at two levels:

- First, individuals have some capabilities to complete tasks, in terms of known actions. Each individual learns new or forgets existing actions over time to adapt their capabilities to the environment. We interpret this as the adaptation of capabilities process at the individual level.
- Second, coalitions are formed by individual agents. Over time, some members of a coalition may be replaced (in a self-organized manner), which we interpret as the adaptation at the coalition level.

[1]Details are provided in Sec. 3.

Following the dynamic capabilities framework, individuals and coalitions are regarded to be successful when they efficiently adapt to the environment in which they operate and, consequently, achieve higher performance levels. Having said this, the perspectives taken by the traditional economic view and the dynamic capabilities framework are mutually incompatible: The former regards stability in the composition of a coalition as beneficial (Hsu et al. 2016), whereas the latter regards the adaptation at all levels as the key to higher performance (Eisenhardt and Martin 2000).

3 THE MODEL

Overview Our research differs in several aspects from previous approaches that study coalition formation and task allocation in Economics, Robotics and Management Science. Regarding Economics, previous research has considered multiple coalitions at the same time and has focused on the mechanisms that prevent agents from switching between coalitions (see, for example, Banerjee et al. (2001) and İnal (2019)). Agents are usually allowed to freely leave or join the coalition (Banerjee et al. 2001, İnal 2019). In contrast, we focus on only one coalition and investigate the dynamics emerging within this coalition. We implement coalition formation as a self-organized process that follows a second-price auction and we endogenously define whether and how often the agents are allowed to join or leave a coalition. Furthermore, we model the agents in a way that they always have an incentive to participate in the coalition.

Research on coalition formation in Robotics is usually concerned with solving complex *physical tasks* by nonhuman entities (Rizk et al. 2019). However, we focus on coalitions formed by human decision-makers who have limited cognitive capabilities and face a complex *decision-making task*.

Finally, the interrelation of task allocation as a mean for coalition formation has not been extensively addressed in Management Science (Rivkin and Siggelkow 2003, Wall 2018). Moreover, the dynamic capabilities framework has not been properly related to coalition formation and task allocation literature. We contribute to this literature by explicitly addressing dynamic coalition formation and investigating the dynamics emerging from coalition formation processes in coalitions in which capabilities at both the individual and the collective level are subject to adaptation.

The task environment The complex task is modelled as a vector $\mathbf{d_t} = (d_{1t} \dots d_{Nt})$ consisting of N binary decisions, with K inter-dependencies among decisions. This follows the NK-framework for organizational decision-making, which was first introduced by Levinthal (1997). As decisions are binary, there exist 2^N solutions to the decision-making problem, each with an associated performance. The mapping of each solution to its associated performance is referred to as the performance landscape. At every time step $t = 1, \dots, T$ a coalition makes decisions $d_{it} \in \{0,1\}$ and each decision contributes c_{it} to coalition performance $C(\mathbf{d_t})$. Performance contributions follow a uniform distribution $c_{it} \sim U(0,1)$, whereby each contribution c_{it} is affected by decision d_{it} and K other decisions. The latter are denoted by $\overline{\mathbf{d}}_{it} = (d_{j_1 t} \dots d_{j_K t})$, where $i, j = 1, \dots, N$, $\{j_1, \dots, j_N\} \subseteq \{1, \dots, i-1, i+1, \dots, N\}$ and $0 \leq K \leq N-1$. Performance contributions are formalized by

$$c_{it} = f(d_{it}; \overline{\mathbf{d}}_{it}) \ . \tag{1}$$

Coalition performance is computed according to $C(\mathbf{d_t}) = \frac{1}{N} \sum_{i=1}^{i=n} c_{it}$. The parameter K shapes the complexity of the decision problem and, consequently, the ruggedness of the resulting performance landscape. Once a coalition is formed, it moves in the performance landscape following a hill-climbing-based search process for a solution that has a better associated performance (Levinthal 1997).

We set $N = 12$ and divide the entire decision problem into three sub-problems denoted by N_s, each consisting of four binary decisions; $s = (1 \dots S)$ indicates the *areas of expertise* that refer to the sub-problems. For each of the partial binary decision problems, there exist 4^2 solutions. These solutions represent capabilities within an area of expertise. The coalition strategy is the set of decisions made or the concatenation of solutions to each sub-problem (i.e., in each area of expertise), respectively, for a particular time step.

Concerning K, we consider three different levels of inter-dependencies, being $K = 3$, $K = 5$ and $K = 11$, and referred to as a low-, mid-, and high level of inter-dependencies, respectively. We also control for the structure of the inter-dependencies, distinguishing between three particular cases:

- In the *concentrated* scenario, each decision is inter-dependent with decisions of the same area of expertise or with decisions close to it (represented by the matrices under the headline *concentrated* in Fig. 1).
- In the *scattered* scenario, decisions are inter-dependent with decisions that are not in the same area of expertise (represented by the matrices under the headline *scattered* in Fig. 1).
- In the scenario with the maximum level of inter-dependencies, all decisions are inter-dependent with each other (represented by the matrix under the headline $K = 11$ *full* in Fig. 1).

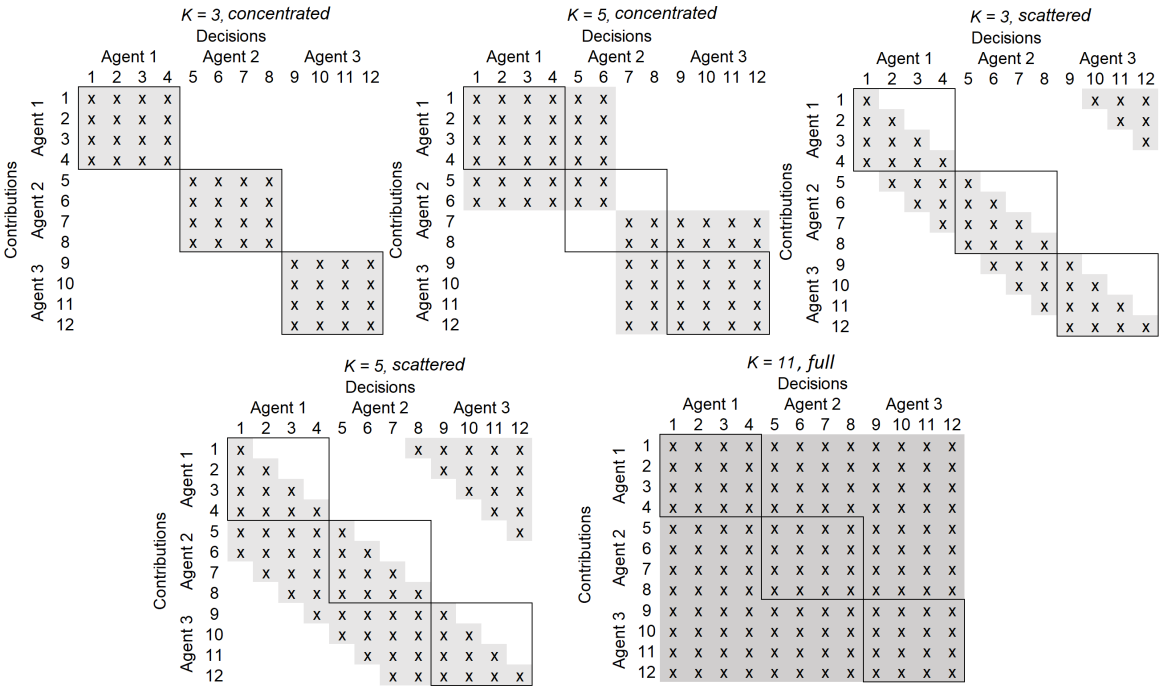


Figure 1: Matrices of inter-dependencies. 'x' stands for inter-dependencies between decisions and performance contributions. Areas of expertise are indicated by solid lines.

Agents and individual adaptation Agents are part of one of the three areas of expertise. They are utility maximizers. Their utility is a weighted sum of their individual contribution to coalition performance $\frac{\sum_{\forall d_{it} \in N_s} c_{it}}{||N_s||}$ and the contribution of other agents to coalition performance $\frac{\sum_{\forall d_{it} \in N_{\neg s}} c_{it}}{||N_{\neg s}||}$, where $N_{\neg s}$ represents the decisions located outside the agent's area of expertise (i.e., the *residual decisions*). In the utility function of an agent m, these two parts are weighted by parameters α and β. The utility function is formalized in Eq. 2. Agents are myopic, since they determine their choices with the objective of maximizing utility in the current period without taking future utility into consideration.

$$U_{mt} = \alpha \cdot \frac{\sum_{\forall d_{it} \in N_s} c_{it}}{||N_s||} + \beta \cdot \frac{\sum_{\forall d_{it} \in N_{\neg s}} c_{it}}{||N_{\neg s}||} \tag{2}$$

At the beginning of each simulation round, agents are assigned to an area of expertise (i.e., to a partial binary decision sub-problem), and they are endowed with some capabilities in their area of expertise. This means that they initially know one of the 4^2 possible solutions to the respective sub-problem. As agents do not have complete information about the solution space, they are boundedly rational (Simon 1957). Agents' capabilities might change over time, following a process of *individual adaptation of capabilities*. At every time step, agents can learn a new solution to their partial decision problem with probability p. This solution differs in one of the four values from the solution that they already know. In addition, with the same probability p, agents can forget a solution which they already know but which does not help in maximizing their utility in that particular time step. Regarding the adaptation of capabilities, we study three scenarios:

- In the *benchmark scenario* we set $p = 0$. Agents do not learn new or forget already known solutions to their partial decision problem.
- In the scenario with a *low probability of adaptation*, we set $p = 0.2$
- In the scenario with a *high probability of adaptation*, we set $p = 0.5$.

Coalitions and coalition formation As discussed above, self-organized coalition formation is modelled to follow a second-price auction. In second-price auctions, the top bidder wins and pays the second-highest bid (Vickrey 1961); agents are fully aware of the mechanism's functioning. Whenever an auction occurs, agents participate and bid the utility they expect from participating in a coalition (for an overview of similar approaches, see Rizk et al. (2019)). Since agents cannot observe the bids made by other agents or the solutions the other agents know, their expected utility is estimated by assuming that the residual decisions do not change compared to the previous period. Expected utilities $E(U_{mt})$ are computed using the expected contribution of each decision to coalition performance, which is formalized in Eq. (3). The expected contribution of a decision is a function of the decision itself at time t, d_{it}, the decisions within the area of expertise that are inter-dependent with the decision at time t, $\overline{\mathbf{d}}_{N_s t}$, and the residual decisions that are inter-dependent with the decision at time $t-1$, $\overline{\mathbf{d}}_{N_{\neg s} t-1}$.

$$E(c_{it}) = f(d_{it}; \overline{\mathbf{d}}_{N_s t}; \overline{\mathbf{d}}_{N_{\neg s} t-1}) \tag{3}$$

The expected utility function follows Eq. (2) and includes the agent's expectation about the contributions of individual decisions formalized in Eq. (3), so that

$$E(U_{mt}) = \alpha \cdot \frac{\sum_{\forall d_{it} \in N_s} E(c_{it})}{||N_s||} + \beta \cdot \frac{\sum_{\forall d_{it-1} \in N_{\neg s}} E(c_{it})}{||N_{\neg s}||} \ . \tag{4}$$

As the second-price auction assures that agents reveal their true preferences (Vickrey 1961), each agent bids the highest expected utility given the currently known subset of solutions. Once all agents have placed their bids, the top bidder for each area of expertise is determined. The top bidders of each area of expertise form the new coalition. Consequently, in our model, a coalition is always composed of three agents, i.e., one from each area of expertise.

Agents are given the possibility of collective adaptation every τ periods. Based on τ, we distinguish three levels of instability.

- In the case of a *stable coalition*, a coalition is formed only once (in the very first time step), and there is no further opportunity for re-organization (i.e., $\tau = 0$).
- In the case of a *mid-stable coalition*, there is the opportunity to reorganize the coalition every ten time steps (i.e., $\tau = 10$).
- In the case of an *unstable coalition*, there is the opportunity to reorganize the coalition at every time step (i.e., $\tau = 1$).

Once a coalition is self-organized, its task is to solve the complex decision problem introduced above. For simplicity, we assume that there is no further coordination or communication within a coalition nor outside of it.

Individual decision-making process and coalition strategy Agents are modelled to be utility maximizers. The agents who are part of the coalition are tasked with choosing a particular solution to their partial decision-making problem (i.e., in their area of expertise). This is done autonomously, with agents making their choices independently from each other. To make their choices, agents calculate the expected utility that each solution to their partial decision problem reports following Eq. (4). On this basis, each agent chooses the solution that promises the highest increase in expected utility (according to Eq. (4)). The overall strategy of the coalition is formed by concatenating the solutions chosen by the agents. Finally, the agents experience the resulting utility.

Process overview, scheduling, and main parameters The simulation model has been implemented in Python 3.7.4. Every simulation round starts with a preparation phase. The performance landscape is computed, and agents are randomly assigned to one area of expertise. They are given initial capabilities in the form of one initial solution to their partial decision-making problem. As there is no residual decision before the first time step, we compute a random N-dimensional bit-string as

starting point in the performance landscape. After preparation has ended, the agents form a coalition in the first time step of each simulation run following the previously outlined procedure.

After the coalition is formed, each member of the coalition is tasked with choosing a solution in their area of expertise at every time step. The coalition strategy is formed by concatenating the solutions proposed by each agent in their particular area of expertise. Agents experience the resulting utility considering the particular coalition strategy and its associated performance. Finally, at the end of every time step, individual adaptation occurs according to the mechanism of individual adaptation previously described. After τ periods the agents are given the possibility to reorganize the coalition following the auction-based mechanism.

This process is repeated over T time steps per simulation round. An overview of the process of the model is provided in Fig. 2.

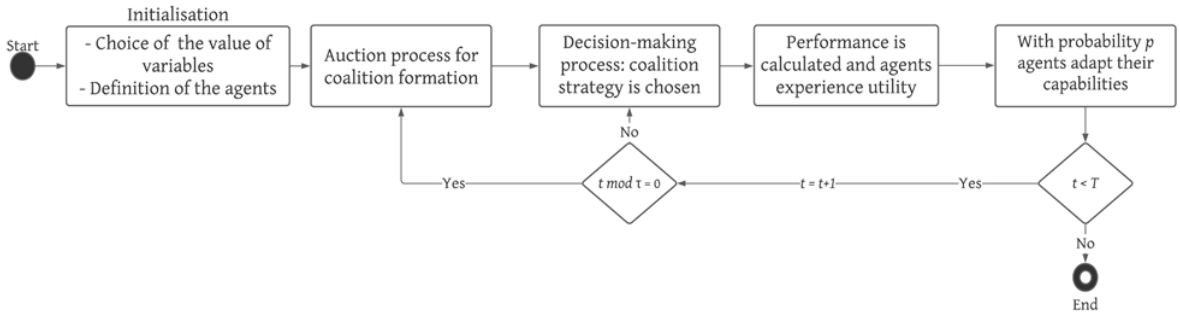


Figure 2: Process overview of the model. This figure represents the order in which every event occurs at a particular t

4 SELECTED RESULTS

4.1 Scenarios and performance measures

Table 1 summarizes the main variables and parameters used in our simulation study. Variables which are exogenously determined are (i) the level of inter-dependencies among sub-tasks (K), (ii) the structure of inter-dependencies, (iii) the frequency of re-organization at the level of the coalition (τ), and (iv) the probability of individual adaptation/learning (p). Given the ranges for these variables included in Table 1, we investigate 45 different scenarios. In each of these scenarios, we observe the results of the first 200 time steps because experiments have indicated that dynamics particularly emerge in this number of time steps. Based on the coefficient of variation, we fix the number of repetitions to 1,500.

Table 1: Main variables and parameters

Factor	Type	Description	Denoted by	Ranges
Variable	Independent	Level of inter-dependencies	K	$\{3;5;11\}$
	Independent	Structure of inter-dependencies	*Interaction matrix*	$\{Concentrated; scattered; full\}$
	Independent	Frequency of re-organization	τ	$\{0;1;10\}$
	Independent	Probability of individual adaptation	p	$\{0;0.2;0.5\}$
	Dependent	Coalition performance	$C(\mathbf{d_t})$	$\in [0;1]$
Parameter	Temporal	Time step	t	$\in [1;200]$
	Fixed	Time horizon	T	$\{200\}$
	Simulation	Simulation round	r	$\in [1;1,500]$
	Fixed	Weights	α, β	$\{0.5\}$

To assure that performances are comparable across scenarios, the observed coalition performance $C(\mathbf{d_t})$ achieved in a specific simulation round $r = 1\ldots,R$ at any time step $t = 1,\ldots,T$ is normalized by the maximum performance which is achievable on the landscape on which the simulation run is performed, $max(C)$. The normalized coalition performance at every time step t is averaged across the 1,500 simulation rounds of each scenario, resulting in the normalized average performance

$$\tilde{C}_t = \frac{1}{R}\sum_{t=1}^{T}\frac{C(\mathbf{d_t})}{max(C)} \ . \tag{5}$$

We report the sum of the Manhattan Distance for each scenario, i.e., the distance between the normalized average coalition performance at each time step and the best performance attainable in that particular scenario (which, after normalization, is equal to 1). This allows us to report a measure that reflects, on average, how each scenario performs. We compute this performance measure according to

$$D = \sum_{t=1}^{T} (1 - \tilde{C}_t) \ . \tag{6}$$

In Fig. 3, we plot the Manhattan Distance for all scenarios resulting from the parameters included in Table 1. Each sub-plot represents the results for a different level of individual adaptation ($p = 0$ in contour plot 1, $p = 0.2$ in contour plot 2, and $p = 0.5$ in contour plot 3). The horizontal axes provide information about the complexity of the task environment (in terms of the level of inter-dependencies K and their structure). Within each contour plot, complexity increases from the left to the right. The vertical axes provide information about the frequency of coalition formation (τ); this frequency increases from the bottom to the top. Note that the performance measure plotted in Fig. 3 is the distance between the average performance achieved by a coalition and the maximum performance. Lower values, thus, indicate better performing coalitions.

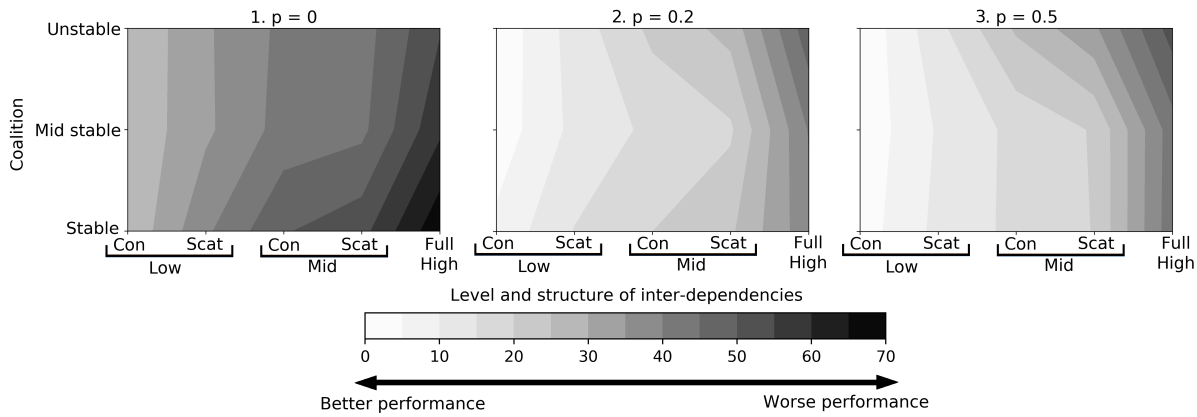


Figure 3: Contour plot. Contours are based on the Manhattan Distance (see Eq. (6)). Note: *Con* refers to a concentrated structure and *Scat* to a scattered structure.

4.2 Results

Results indicate that there exists a negative relationship between the *level of inter-dependencies* (K) and the performance achieved by a coalition: We can observe that increases in complexity lead to a decrease in coalition performance. If we move further to the right on the horizontal axis of the contour plots presented in Fig. 3, the Manhattan Distance increases, indicating that the distance between the maximum performance and the achieved performance increases. For $p = 0.2$, for instance, moving from low inter-dependencies ($K = 3$) and a concentrated structure to full inter-dependencies ($K = 11$) increases the Manhattan Distance from a value of less than 6 to a value of more than 40. This observation is robust across all frequencies of re-organization (i.e., all values of τ) and all considered probabilities of individual adaptation (i.e., all values of p).

In addition to the observations related to the level of inter-dependencies (K) on performance discussed above, the results suggest the following effect for the *structure of inter-dependencies*: Performances tend to decrease when we move from a concentrated to a scattered structure *within* one level of inter-dependencies. This observation is particularly pronounced for low levels of complexity and robust across all probabilities of individual adaptation (i.e., p) and all frequencies of re-organization at the level of the coalition (i.e., τ). For instance, moving from the concentrated to the scattered structure for mid-stable coalitions on the left-hand side of contour plot 2 in Fig. 3 (i.e., the case of low inter-dependencies ($K = 3$), a moderate probability of individual adaptation ($p = 0.2$), and a high frequency of re-organization ($\tau = 10$)), increases the Manhattan Distance from a value below 1.84 to a value beyond 11.33. That is an increase of 6 times in value. For a medium level of inter-dependencies ($K = 5$)

and the same values of $p = 0.2$ and $\tau = 10$, moving from the concentrated to the scattered structure leads to a relatively smaller increase in the Manhattan Distance from a value of 16.16 to a value of 19.20. This observation might be explained by the fact that higher levels of inter-dependencies do no longer allow for a fully self-contained structure, even in the concentrated case. This means that for high levels of K, the inter-dependencies cannot be completely located within an area of expertise (see Fig. 1). In other words, the marginal decrease in coalition performance is relatively low when the number of inter-dependencies across areas of expertise is already relatively high.

Results suggest that *individual adaptation* (p) – i.e., the capability to learn – is a key factor for performance. Coalition performance improves considerably when we move from $p = 0$ to $p = 0.2$ or $p = 0.5$ (see Fig. 3). For the case of no individual adaptation ($p = 0$), low levels of inter-dependencies ($K = 3$), a concentrated structure, and a mid-stable coalition ($\tau = 10$), endowing the agents with the capability to learn with a probability of 20% (i.e., moving from $p = 0$ to $p = 0.2$) decreases the distance between the performance achieved by this coalition and the maximum attainable performance from a value of 25.90 to a value of 2.82. Performance, thus, increases to a large extent when agents are endowed with learning capabilities. For higher levels of complexity (i.e., for higher values of K), the same effect can be observed, even though it is less pronounced. For $K = 11$ and $\tau = 1$, moving from $p = 0$ to $p = 0.2$ decreases the distance only from a value of 43.69 to a value of 40.34. For cases in which agents are already endowed with learning capabilities, increasing the probability of individual adaptation p – i.e., increasing the frequency of learning – only has a significant effect on performance when the complexity of the task environment is considerably high. This indicates that the positive effect of individual adaptation is only relatively high if non-adapting ($p = 0$) agents learn to adapt to the environment ($p > 0$). For agents who are already well-endowed with the capability of individual adaptation (in terms of a high p), the effect is non-significant. In other words, the marginal positive effect of endowing agents with the capability to adapt to the environment – in terms of learning – decreases with p.

Concerning the *frequency of adaptation at the level of the coalition* (τ), the results suggest a strong interaction between τ and the probability of individual adaptation. For high-levels of inter-dependencies ($K = 5$ and $K = 11$), the pattern in the contour plots presented in Fig. 3 appears to be substantially shaped by the probability of individual adaptation in the following way: For coalitions which are composed of agents who are not endowed with learning capabilities (i.e., $p = 0$, see contour plot 1 in Fig. 3), a frequent re-organization of the coalition appears to be beneficial for coalition performance. For coalitions which are composed of agents who learn with moderate probability (i.e., $p = 0.2$, see contour plot 2 in the middle of Fig. 3), re-organization appears not to play a central role when it comes to performance. If we move to coalitions which are composed of individuals who learn very frequently (i.e., $p = 0.5$, see contour plot 3 in Fig. 3), the pattern appears to shift into the opposite direction, so that stability in the composition of a coalition tends to have positive effects on performance. If coalitions face less complex tasks (i.e., if we move to the left on the horizontal axes in the contour plots presented in Fig. 3), this pattern becomes insignificant, so that the effect of the frequency of re-organization at the coalition level no longer has effects on performance. This observation might be explained by the fact that both individual adaptation and re-organization at the coalition level have similar consequences from the coalition's perspective. The former allows the agent to find new solutions to his or her partial decision problem which are, then, contributed to solving the problem which the entire coalition faces. The latter, by replacing the members of a coalition, can introduce new members who might bring in new solutions to solve the decision-making problem the coalition faces. From the perspective of the coalition, both individual and collective adaptation foster the innovativeness of the coalition: In case of agents who are not endowed with learning capabilities (i.e., $p = 0$), a coalition assures innovativeness by frequently replacing members. For coalitions already composed of agents who often learn and make progress (i.e., $p = 0.5$), there is no need to replace members, since the coalition is already quite innovative. Our results indicate that too much innovativeness in the above sense is detrimental to coalition performance. In other words, if a coalition which already comes up with innovative ideas faces a complex problem, one would be well-advised to assure that the coalition does not re-organize and replace members – this perfectly translates to "never change a winning team", at least do not do so when this team faces a very complex task.

5 CONCLUSION

5.1 Summary

From on the results presented in Sec. 4, four key findings can be derived. First, there is a negative relationship between the level of inter-dependencies among sub-tasks of a complex decision-problem (K), which is faced by a coalition, and the achieved performance. This finding is in line with previous literature, which finds that a higher degree of complexity leads to a larger number of local maxima (e.g., more rugged landscapes in the context of the NK-framework) and, consequently, a higher chance of getting stuck at a local maximum (Levinthal 1997, Leitner and Wall 2014).

Second, the structure of inter-dependencies among sub-tasks of a complex decision-problem appears to have substantial effects on performance at moderate levels of complexity. For complex decision problems, though, the impact of the structure of inter-dependencies is negligible. This finding is also in line with the results of previous studies on task allocation in complex environments, such as Wall (2018) and Hsu et al. (2016).

Third, endowing decision-makers with learning capabilities, so that the probability of adapting to the environment increases, is particularly fruitful for agents who just learn to adapt. For agents who are already well-trained in adapting to the environment, marginal positive effects decrease significantly. Results of the NK-model suggest that the positive effect of coming up with new solutions is relevant at the beginning of the simulation, but its relevance decreases over time (Levinthal 1997). This could explain why the marginal positive effects of adaptation diminish as p increases: Gains coming from individual adaptation would only be relevant for the first time steps.

Finally, whether adaptation at the coalition level is beneficial for performance is strongly affected by the characteristics of the members who form the coalition. While innovative coalitions that face complex decision problems should make every effort not to replace members, less innovative groups would be well-advised to replace members in a self-organized manner to increase performance. If the faced decision problem is of low (or in some cases moderate) complexity, the effect of (in)stability becomes insignificant. This is in line with the results of Wall (2018) and Hsu et al. (2016).

5.2 Conclusive remarks

This research is a first attempt to systematically integrate the dynamic capabilities framework into problems of coalition formation in complex task environments. Results are consistent with and extend the main insights from previous literature, such as Leitner and Wall (2014), Levinthal (1997) and Wall (2018). Our research is, however, not without its limitations. First, we consider human decision-makers who are rational in the sense of perfect foresight when evaluating the performance on their landscape (i.e., they make no errors when predicting expected utilities) (Hendry 2002), they do not suffer from decision making biases (Kahneman et al. 1982), and they can correctly handle complex decision problems (i.e., they have the cognitive capacity to handle decision problems irrespective of their complexity) (Pennington and Hastie 1986). Moreover, we do not consider coordination mechanisms within coalitions; in our model, every agent acts in a fully autonomous manner. Future research might want to include mechanisms to coordinate decisions within a coalition, such as suggested in Wall (2018). Our research can also be expanded by introducing heterogeneity in the adaptation of capabilities or shocks that change the shape of the solution space. All these features could lead to new insights within the contexts of complex decision-making and dynamic coalition formation.

ACKNOWLEDGMENTS

We would like to thank Joop van de Heijning and Ravshanbek Khodzhimatov (University of Klagenfurt, Austria) for their valuable assistance in the implementation of the model.

REFERENCES

Banerjee, S., H. Konishi, and T. Sönmez. 2001. "Core in a simple coalition formation game". *Social Choice and Welfare* 18 (1): 135–153.

Eisenhardt, K. M., and J. A. Martin. 2000. "Dynamic capabilities: What are they?". *Strategic Management Journal* 21:1105–1121.

Giannoccaro, I., G. Massari, and G. Carbone. 2018. "Team resilience in complex and turbulent environments: The effect of size and density of social interactions". *Complexity* 2018 (6): 1–11.

Hendry, J. 2002. "The principal's other problems: honest imcompetence and the specification of objectives". *Academy of Management Review* 27 (1): 98–113.

Hennet, J. C., and S. Mahjoub. 2011. "Coalitions of firms in manufacturing networks: Stability and optimality issues". In *Proceedings of the 18th World Congress The International Federation of Automatic Control, Milano*, 6419–6424. IFAC.

Hsu, S. C., K. W. Weng, Q. Cui, and W. Rand. 2016. "Understanding the complexity of project team member selection through agent-based modeling". *International Journal of Project Management* 34 (1): 82–93.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Leitner, S., and F. Wall. 2014. "Multiobjective decision making policies and coordination mechanisms in hierarchical organizations: Results of an agent-based simulation". *The Scientific World Journal* 7:1–12.

Leitner, S., and F. Wall. 2015. "Simulation-based research in management accounting and control: an illustrative overview". *Journal of Management Control* 26 (2-3): 105–129.

Levinthal, D. A. 1997. "Adaptation on rugged landscapes". *Management Science* 43 (7): 934–950.

İnal, H. 2019. "The existence of a unique core partition in coalition formation games". *Games and Economic Behavior* 114:215–231.

Pennington, N., and R. Hastie. 1986. "Evidence evaluation in complex decision making.". *Journal of personality and social psychology* 51 (2): 242–258.

Rivkin, J. W., and N. Siggelkow. 2003. "Balancing search and stability: Interdependencies among elements of organizational design". *Management Science* 49 (3): 290–311.

Rizk, Y., M. Awad, and E. W. Tunstel. 2019. "Cooperative heterogeneous multi-robot systems: A survey". *ACM Computing Surveys* 52 (2): 29:1–29:31.

Simon, H. A. 1957. *Models of man, social and rational*. John Wiley and Sons.

Sless, L., N. Hazon, S. Kraus, and M. Wooldridge. 2018. "Forming k coalitions and facilitating relationships in social networks". *Artificial Intelligence* 259:217–245.

Teece, D. J., G. Pisano, and A. Shuen. 1997. "Dynamic capabilities and strategic management: organizing for innovation and growth". *Strategic Management Journal* 18 (7): 509–533.

Tremewan, J., and C. Vanberg. 2016. "The dynamics of coalition formation – A multilateral bargaining experiment with free timing of moves". *Journal of Economic Behavior and Organization* 130:33–46.

Vickrey, W. 1961. "Counterspeculation, auctions, and competitive sealed tenders". *The Journal of Finance* 16 (1): 8–37.

Wall, F. 2018. "Emergence of task formation in organizations: Balancing units' competence and capacity". *Journal of Artificial Societies and Social Simulation* 21 (2): 1–26.

Wall, F., and S. Leitner. 2020. "Agent-based computational economics in management accounting research: Opportunities and difficulties". *Journal of Management Accounting Research* (11).

AUTHOR BIOGRAPHIES

DARIO BLANCO-FERNÁNDEZ is a PhD student in the fields of Social Sciences and Economics at the University of Klagenfurt. He is currently in his second year. His email address is dario.blanco@aau.at and his ORCID ID is 0000-0002-8275-701X

STEPHAN LEITNER is an Associate Professor at the Department of Management Control and Strategic Management at the University of Klagenfurt. His email address is stephan.leitner@aau.at and his ORCID ID is 0000-0001-6790-4651

ALEXANDRA RAUSCH is an Associate Professor at the Department of Management Control and Strategic Management at the University of Klagenfurt. Her email address is alexandra.rausch@aau.at and her ORCID ID is 0000-0002-9275-252X

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.039

ONTOLOGY DERIVED CONCEPTUAL MODELING FOR SIMULATION

Nurul Saleh

Brunel University London
London, United Kingdom
nurul.mdsaleh@brunel.ac.uk

David Bell

Brunel University London
London, United Kingdom
david.bell@brunel.ac.uk

ABSTRACT

An ontology is one of the way to represent domain knowledge into a human-understandable and machine-readable format. Meanwhile, an ontology in simulation has been used as a conceptual model to explicitly describes the modelers' perspective of the domain. This study proposes a rigorous method that systematically extracts domain concepts, synthesizes processes within the domain and build an ontology for simulation modelling - a Minimal-Viable Simulation Ontology (MVSimO). MVSimO can be viewed as a derived conceptual model that supports modelling and simulation through abstraction and simplification of the domain. The novel approach presented curates the modelers' perspective of the real-world by extracting concepts from existing knowledge and synthesizes the processes involved (demonstrated in A&E departments). The effectiveness of this method is reviewed by comparing MVSimO ontological model to the existing model. Evaluation results are encouraging, providing possibilities to improve an ontology for simulation when access to experts is limited.

Keywords:

Ontology, Conceptual Modeling, Simulation Modeling, Formal Concept Analysis

1 INTRODUCTION

The Semantic Web relies heavily on the underlying data structure for the purpose of comprehensive and transportable machine understanding. The term "Semantic Web" refers to the Web of linked data and the technologies enable creation of data stores on the Web, build vocabularies, and rules for handling data. Linked data of Semantic Web stack are empowered by technologies such as RDF, SPARQL, OWL and SKOS. Therefore, the success of Semantic Web depends strongly on the proliferation of ontologies, which facilitates knowledge acquisition by ontology engineers with the help of domain expert.

To build an ontology-based application, a conceptual model used in application development is typically supported by domain experts to gather domain requirements (Robinson 2013). For example, in developing a simulation, conceptual modeling represents a composition of concepts which help to view the abstraction of the real world system. Meanwhile, the modeler or researcher attain the optimal knowledge of the domain with the semantic representation of ontology, are able to improve the translation of real-world knowledge to model representation. This increase the chances of getting the correct simplification of the domain by making the domain assumption explicit (Noy, McGuinness, et al. 2001). Although ontologies have been introduced as conceptual model for a semantically-defined application, the development of the new ontology can be tedious and costly (Lonsdale, Embley, Ding, Xu, and Hepp 2010, Simperl 2009). The research presented in this paper is a development of a Minimal Viable Simulation Ontology (MVSimO) as an ontology derived conceptual model for simulation modeling.

For a complex domain, the observed phenomena generate the understanding and help the modeler in the abstraction and simplification process of the real-world. These processes underpin the work in this

paper by extracting the important elements and concepts of the A&E departments and develop MVSimO. This paper focuses on the first phase of MVSimO development which it reuses existing ontologies from the healthcare domain that are directed toward simulation for A&E departments. The phase where processes are synthesized using Formal Concept Analysis will not be discussed in detail. The definition of MVSimO is adopted from the definition of Minimal Viable Product by (Ries 2009): "*the version of a new product which allows a team to collect the maximum amount of validated learning about customers with the least effort*". Domain knowledge from real data and generic pathways is used to perceive the domain understanding of A&E departments. The study is conducted by reusing the domain knowledge, thus saving the time and effort spent in requirement gathering and ontology design process. During the evaluation of MVSimO, an existing simulation model developed by experts is used for validation to gain an insight into the proposed work.

This study follows a design research approach which (1) identifies the problem area and its relevance from a real-world environment and previous research, (2) develops the model as a design artefact, and (3) evaluates the model through a relevant scenario. The paper is structured as follows. Section 2 discusses the background of the study, simulation in A&E departments, ontology representation and Formal Concept Analysis (FCA) with respect to a conceptual model. Section 3 looks at the domain conceptualization for simulation modeling from the perspective of A&E departments, and explain the detailed map of A&E pathways. Section 4 briefly discusses the mathematical approach of domain attributes exploration using FCA. Finally, Section 5 explains the development of MVSimO and Section 6 discusses the evaluation conducted for this study and the conclusion respectively.

2 Study Background

2.1 Simulation in A&E Departments

The emergency department, widely known as accident and emergency or A&E department in healthcare services plays a major role to save people's lives, and more importantly to reduce death and disease rate in public (Aringhieri, Bruni, Khodaparasti, and van Essen 2017). The departments refer to the sub-system that provides medical treatment to urgent need patients, and is the most critical unit since they are one of the first unit responsible in treating life-and-death situation (Gul and Guneri 2015). Figure 1 shows the process journey in A&E departments. The figure presented in the project "*A Better A&E*" by PearsonLloyd shows the different stages from patient's check-in, assessment by staff, receiving treatment from medical staff, and the outcomes of the process.

Due to the complexity of A&E, simulation has been used by decision makers to understand the processes and behaviours(Aringhieri, Bruni, Khodaparasti, and van Essen 2017, Traore and Zeigler 2018). Simulation models conduct a preliminary test or trial changes for a safe and efficient care deliverable implementation (Günal and Pidd 2009, Aringhieri, Bruni, Khodaparasti, and van Essen 2017) and at the same time helps in finding the optimal solution to overcome problems appear within the department (Gul and Guneri 2015). Attributes to simulate real world situations help decision makers to predict the output of a proposed solution in real-world phenomena using Discrete-event simulation (Lebcir, Demir, Ahmad, Vasilakis, and Southern 2017), System Dynamic (Pidd 2014) and Agent-based simulation (Chahal, Eldabi, and Young 2013). For the purpose of this study, DES is used due to the widespread agreement of its generality among simulation models (Guizzardi and Wagner 2010).

2.2 Ontology and FCA

To ensure the usefulness of simulation model in a complex and heterogeneous system like A&E departments, the model is required to accommodate the behaviours and interaction within the domain (Isern and Moreno 2016, Baboolal, Griffiths, Knight, Nelson, Voake, and Williams 2012). Consequently, an ontology is generally assumed to play a role in setting up a common ground to provide knowledge sharing among subject domain and also describe, standardize and represent an object or instance in the domain (Grolinger,

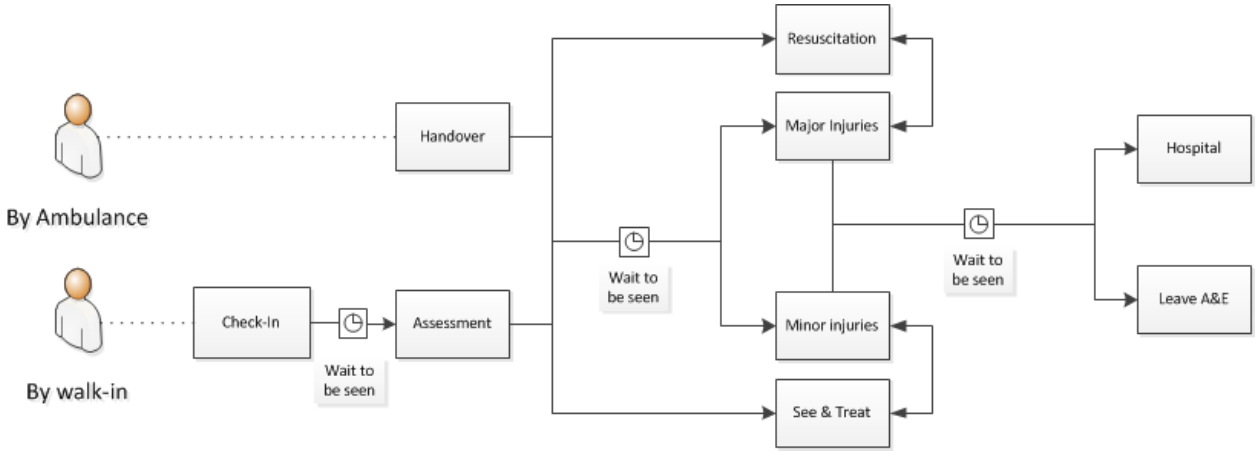


Figure 1: A&E Process Flow

Capretz, Marti, Srivastava, Grolinger, and Capretz 2012, Huang 2016). This is because the basic foundation of ontology itself is a formal specification of conceptualization (Gruber 1993). Ontologies can accurately define a domain using classes, properties, relationships and instances hence support the determine the content of the simulation model to build the model that contains explicit detail of the domain. The novelty of the proposed approach in this paper, is that ontology is not simply build by reusing exsiting ontologies, but instead synthesize the processes performed in the departments using Formal Concept Analysis (FCA) approach.

FCA was developed in the early 1980s as a mathematical perception for concept formalization and conceptual thinking (Wille 1982). According to its philosophical definition (Wille 1982), a concept is composed of a set of objects as its extensions and a set of attributes as its intentions. Examining the process and entities in A&E departments using FCA, gives the ability to draw new relationships that can be used to represent MVSimO class and properties. Taken together, this suggests that the aggregation of ontology and FCA during conceptual modeling may help to better understand the domain and its emerging issues, thus making the proposed model more usable and beneficial even when access to experts' opinions are limited.

3 A&E Case Study

3.1 Domain Conceptualization

In the first phase of MVSimO development, the existing ontologies are used as to represent the domain. The requirement is defined to determine which ontology to be selected and reused to best represent the domain and the application to be modeled. Here, the domain are conceptualized to discover the candidate ontologies. Domain conceptualization phase is conducted based on general model theory by (Stachowiak 1973) - mapping, reduction and pragmatic. **Mapping:** To map the process in the A&E departments, the process flow 1 is divided into modules. In the setting for simulation of A&E departments, the modules are "healthcare" and "hospital". These modules are determined based on the intention on what to model. **Reduction:** The A&E department is a large system involving several resources and heterogeneous patient type within a complex and well-organized process (Ghanes, Wargon, Jemai, Jouini, Hellmann, Thomas, and Koole 2014). Hence, ones need to identify the assumptions and ontological commitment that each module should comply to. The description of the conceptual model is the commitments outline in the specifications requirements of MVSimO. **Pragmatic:** In order to obtain a suitable ontology to reuse for MVSimO, in general, the ontologies should comply to the processes conducted in A&E departments. From these three general rules, five modules are deduced: 'healthcare', 'hospital', 'emergency department', 'process' and 'patient data'. These modules are used as keywords to search for existing ontology available on the internet.

Generally the modules obtained from this process represent subclasses of MVSimO. Considering ontology repositories that are under active development, Bioportal [1] and Ontobee [2] are used to search for existing ontologies using modules-represented keyword obtained.

The search returned 31 results for BioPortal and 23 results for Ontobee. From the total of 54 ontologies minus duplicate ontologies from the same keywords, only 7 ontologies fit the criteria as outline by (Malone, Stevens, Jupp, Hancocks, Parkinson, and Brooksbank 2016). The 7 selected ontologies are:

- (HEIO) Regional Healthcare System Interoperability and Information Exchange Measurement Ontology
- (OMRSE) Ontology of Medically Related Social Entities
- (GENEPIO) The Genomic Epidemiology Ontology
- (OOSTT) Ontology of Organizational Structures of Trauma centres and Trauma systems
- (TRIAGE) Nurse Triage
- (TRANS) Nurse Transitional
- (RNPRIO) Research Network and Patient Registry Inventory Ontology

The discovery process follows the guideline developed by (Malone, Stevens, Jupp, Hancocks, Parkinson, and Brooksbank 2016), to ensure that the ontology is about a specific domain of knowledge, or in this case an appropriate amount of knowledge to cover the modules. Assumption is made that BioPortal and Swoogle are contributed by domain experts from the Semantic Web community and are under active development. From the selected ontologies, ontology merging and integration are conducted to select suitable objects, instances and properties to be reused in MVSimO. This discovery process starts from domain conceptualization that leads to modules identification and selecting candidate ontologies able to replace the role of experts in determining the abstracted and simplified simulation to be modeled.

3.2 Space-Time-Process Map

To map processes involve in A&E departments to discrete-event simulation paradigm, a new detailed pathways called Space-Time-Process (STP) map is created from a generic pathways of A&E process flow. The method of extracting the process element based on space, time and event for the ontology development has adopted the dataset analysis as proposed by (Sider et al. 2001) and (De Cesare, Juric, and Lycett 2014). To obtain the STP map, the process element based on space, time and event has adopted the dataset analysis as proposed by Sider et al. (2001) and De Cesare et al. (2014), for the ontology development method.STP map (Figure 2) is produced as a means to improve the representation of the domain by providing a new dimension to the existing process flow. The three-dimensional map illustrates the processes in the department and depicts the events in space and time dimension blocks(De Cesare, Juric, and Lycett 2014). The mapping process extracts possible process elements for the development of MVSimO model based on DES paradigm and provides the first insight from a modeler's perspective. From the map, the processes are separated into blocks of activity and entity. The blocks are lined into sequence order to resemble the flow in A&E that have been categorised into 4 main processes: *Check-In*, *Assessment*, *Treatment*, and *Outcome*. Each process block is then labelled as process, activity and entity.

The process begins with a patient check-in, either by walking to the hospital or by using an ambulance service, patient then waits to be assessed by a nurse before proceeds to the triage assessment, receives treatment by a nurse or a medical staff, and finally waits for the outcome of whether to be discharged from the hospital or admitted to the ward. The processes take place in their designated space or location in the department within specific time. For example: A process of *Check-in* within t_1 and t_2, has an event of *Patient Check-in* with *Receiptionist,Nurse* as the entities. The process-event block represents each event that will be modeled for the discrete-event simulation. During the transition from one process to another,

[1] http://bioportal.bioontology.org/
[2] http://www.ontobee.org/

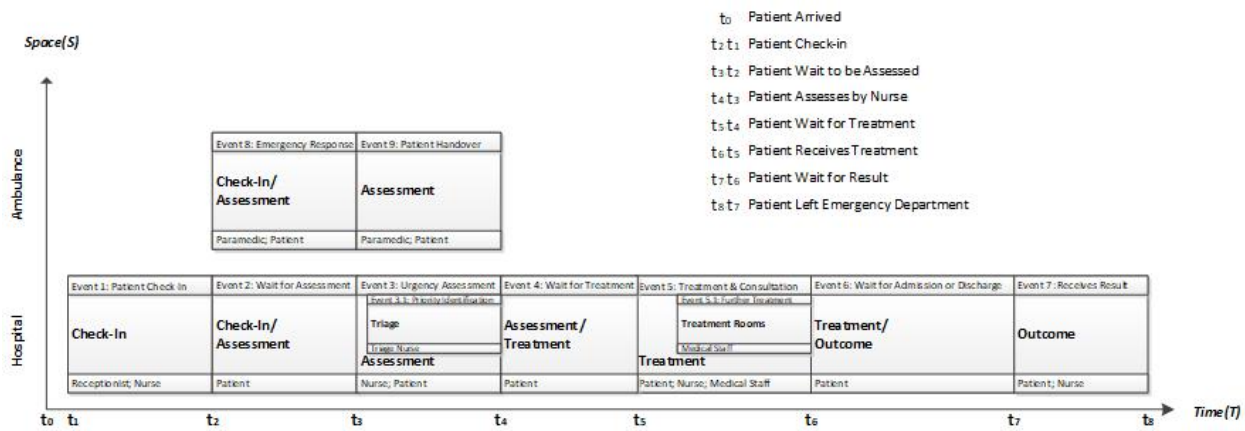


Figure 2: A Space-Time-Process of 'Patient Flow in A&E Departments'

the patient has to wait in a queue. This suggests a delay in the process, for example between t_2 and t_3 of the process *Check-In/Assessment*, the event of *Wait for Assessment* with the entity of *Patient* occurs. In another example of *Assessment* event, there is also a sub-event which occurs in a sub-location in the department. The *Triage* process is conducted by a Triage Nurse to run *Priority Identification*.

STP map, in a three-dimensional perspective (space, time and event), help in deciding how the A&E process elements can be extracted and modelled. Starting from here, all steps taken are based on STP map as it outlines the processes of A&E; determines object's roles and boundary, as well as level of details; and models events for the discrete-event simulation. The application of STP map in MVSimO, makes various kinds of objects, properties and relations between classes and their instances comprehensible. Beside, STP map act as a bridge for the modeler and the user, and translating the real-world process into an ontological definition. The rationale behind STP in this study is to identify the process elements which can then be applied in the FCA-conceptual exploration to extract suitable concepts, and their taxonomic and non-taxonomic relation.

4 Process Elements Extraction

For further development of MVSImO in this study, A&E data from a London Hospital is used to represent objects and attributes of the domain and by using conceptual exploration approach in FCA, the dependencies between the attributes are described and the concepts are determined. Firstly, FCA formalizes the notion of concept relative to a formal context - objects and attributes mapped into a cross-table. The formalization of concept is based on modeler's perspective on what to model, for the purpose of this: the A&E pathways. The motivation of concept exploration is to find new relationship for classes in MVSimO, and to have that the focus is given to two objectives; **first**, to produce a relevant concept and **second**, to construct the minimal set of implications from the concepts. The followings are the formal definitions of FCA to achieve these two objectives:

- **Definition 4.1** Formal context
  A formal context is a triplet *(X, Y, I)* where X is a set of objects and Y is a set of attributes and I is a binary relation between X and Y, i.e., $I \subseteq X \times Y$. $(x,y) \in I$ indicates that the object x has attribute y.
- **Definition 4.2** Intent and Extent
  Let (X, Y, I) be a context, X' $\subseteq X$ and Y' $\subseteq Y$, the function Intent maps a set of objects to the set of attributes, whereas the function Extent maps a set of attributes to the set of objects:
  Intent (X') = y $\in Y'$ — $\forall y \in Y'$, (x,y) $\in R$
  Extent (Y') = x $\in X'$ — $\forall x \in X', (x,y) \in R$

For X' $\subseteq X$, Intent (X') is the set of attributes owned by all objects of X', and Extent(Y') is the set of all objects that own the attributes Y'. The two functions form a Galois connection and formal concepts.

- **Definition 4.3** Formal Concept
  A Formal Concept C in a context is a pair (X', Y') that satisfies Y' = Intent (X') and X' = Extent(Y') i.e., C is a Formal Concept $\Leftrightarrow$ for X' $\in CandY' \in C$, Extent(Intent(X')) = X',
  and symmetrically, Intent(Extent(Y')) = Y'.
- **Definition 4.4** Implications
  An implication A $\Rightarrow B$ holds in (X,Y,I) if and only if B $\subseteq A$", which is equivalent to A' $\subseteq B$'. It then automatically holds in the set of all concept intents

The attributes, including the transformed single-value data are selected based on processes from STP - *Check-In*; *Assessment*; *Treatment*; and *Outcome*. These attributes are merely selected according to the processes through A&E pathways to generate cross-table. The process elements in STP defined as process name (e.g.Outcome), process entity (e.g.Patient), process date and time (e.g.Mon: am), and process location (e.g.Dept: A and E). This step is to obtain concepts with process-related attributes. This process reduced the dataset by only includes the columns with process element hence makes implicit knowledge discovery easier by focusing only on the context of the model, and also makes the representations of FCA concept more process-oriented. The reduced dataset focused on process element by including the attributes of:

Age (Age)
EMAttendanceDate (Attendance Date/Time)
EMModeofArrivalDescription(Mode of Arrival)
AttendanceDisposal (Outcome - Admitted or Discharged)
DepartmentDescription (Department)

ConExp software (Yevtushenko 2000) is used to plot the cross table. ConExp generates implication basis and Concept Lattice diagram to determine any new relationship for MVSimO classes.

4.1 Implication to Class Relation Translation

The result from FCA implicitly described the process relation in A&E department. Using the logical reasoning method by (Xiao Hang Wang, Da Qing Zhang, Tao Gu, and Hung Keng Pung 2004), the implications result from FCA implicitly described the process relation in A&E department.Considering attribute *Age* and *Attendance Date/Time* belonging to *Patient* class (Subject) in MVSimO, only attributes *Mode of Arrival, Admission or Discharge* and *Department* are taken as class properties to describe the process in A&E. Based on logical reasoning of the first-order predicate - a subject, a verb and an object and referring to the classes in MVSimO, a subject is a class e.g *Patient*, a verb is a properties-derived implication, and an object is a class e.g *Department*. As a result from the implications, five concepts with two or more objects are selected as concepts. Table 1 shows the list of formal concepts in natural language statement.

5 MVSimO Development

For the development of MVSimO, existing ontology of DeMO is adopted as guideline. An ontology for discrete-event modeling and simulation (DeMO) by (Silver, Miller, Hybinette, Baramidze, and York 2011), provides taxonomies for a discrete-event simulation model that captures the essential features of the real world system. The refinement of MVSimO classes has taken suitable classes from DeMO includes relation properties from conceptual exploration process from previous section. Using method of event scoping and event harmonization from framework by (Bell, De Cesare, Iacovelli, Lycett, and Merico 2007), the first version of MVSimO ontological model is created by deriving the semantic content of the domain. The framework incorporates STP map to replicate the A&E domain; and the domain knowledge from existing ontology (DeMO). The development is supported by the process relations from Formal Concept Analysis.

Table 1: Selected Formal Concepts

Formal concept, expressed in natural language	**No.of Object**
Fri: pm, MOA:Brought in by Ambulance, Dept: A and E, Age: 65 and above, AD:Admitted to hospital bed/LODGED Patient	3
Thu: pm, MOA:Other, Dept: Mount Vernon MIU, Age: 65 and above, AD: Discharged - did not require followup	2
Sun: pm, MOA:Other, Dept: UCC, Age: 0-4, AD: Discharged - did not require followup	2
Mon: am, MOA: Brought in by Ambulance, Dept: A and E, Age: 19-64, AD: Admitted to hospital/LODGED Patient	2
Wed: pm, MOA: Other, Dept: mount Vernon MIU, Age: 5-18, AD: Discharged - did not require followup	2

Incorporating process relations and classes from DeMO enables the detailing of the relationship of the real world knowledge. The artefact can be regenerated to incorporate new ontologies to extract new classes and relationships. This step is necessary to allow for the flexibility of the framework, and to enable the ontology engineer to go back to this step to add new classes and properties and create new ontology. The process of existing ontology adoption in this activity has resulted in the creation of new *classes*, *properties* and *individuals*. Apart from class properties from the existing ontologies, as mentioned earlier, new properties of MVSimO are also derived from the real-world knowledge through the formal concept analysis. Individuals or instances are from the A&E data. Table 2 summarizes the decisions made and actions taken to elements into MVSimO. Elements are combined to create the initial ontological model.

Table 2: Elements and Class Properties

Elements	**MVSimO Class**
Process hasOccuranceOf Event	Process, new class (Event)
Patient codedBy Patient ID	Patient, new class (Patient ID)
Patient waitsIn Queue	Patient, new class (Queue)
Queue codedBy Date/Time	new class (Queue), new class (Date/Time)
Event takesPlaceAt Location	new class (Event), renamed (Facility) to (Location)
Patient servedBy Event	Patient, new class (Staff) with subclass (Nurse)
Event supportedBy Staff	new class (Event), new class (Staff) with subclass (Nurse)

In model harmonization activity, the first-cut domain model is combined with DeMO to show the scoped event to be model. To be more specific, event of *getPatientCheckIn* and *getPatientAssessment* are presented. This activity uses a process-oriented ontology subclass in DeMO named the *ProcessOrientedModel* or PIModel. Ontologically, this enables an explicit mapping between real-world DES elements and the domain they serve. The harmonized model is mapped into DeMo PIModel accordingly, and may be translated later to an XML and then into a simulation model. This allows researchers, domain experts and modellers to share a common understanding of the concepts and the relationship of the domain. Harmonized

model enables an explicit mapping between a process (with its parts), and the domain it serves supported by A&E data.

Figure 3 presents the harmonization model derived from the event of *getPatientCheckIn* and *getPatientAssessment*. The processes are defined within DeMO and their parameter are typed in relation to the respective classes. From the diagram, *Process* and *Patient* classes are derived from MinDO; , *Location*, *Event*, *Staff*, *Nurse*, *Queue*, *Patient ID* and *Date/Time* are new classes or renamed classes after the decisions made on which elements to be taken into MVSimO. Class *Check-In*, *Assessment*, *Treatment*, *Outcome*, *Paramedic* and *Medical Staff* are from Space-Time-Process map, and finally *Activity*, *Entity*, *Queue*, *Location*, *Resources*, and *Process* are mapped-out from DeMO.

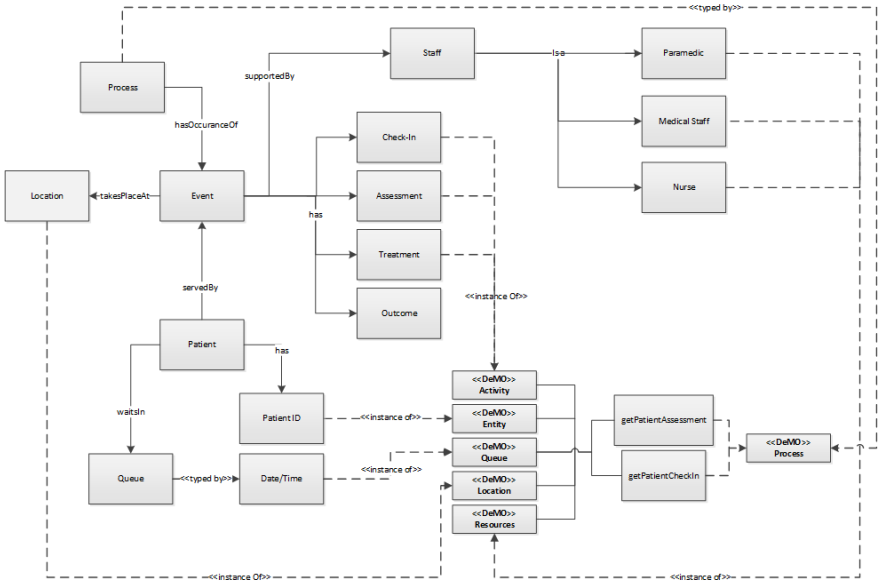


Figure 3: The Harmonized Model

Table 3: MVSimO-Cumberland Objects Assessment

Category	**Cumberland Model**	**MVSimO**
Data Group	Patient Age	Age
	Arrival Method	EMModeofArrivalDescription
	Arrival Time	EMAttendanceDate
		DepartmentDescription
		AttendanceDisposal
Objects	Departments	Process
	Demographic Data	Activity
	Resources	Location
		Patient
		Resources
		Queue

A comparison of the model presented in this study with existing simulation model, the Cumberland model (Bell, Cordeaux, Stephenson, Dawe, Lacey, and O'Leary 2017) was conducted. The comparison evaluates the functional adequacy and quality of the model by conducting similarity analysis for categories of: *Data Group* and *Objects*. Table 3 shows the similarity assessment of the elements in MVSimO and Cumberland model. From element-by-element comparison with 10 elements extracted from MVSimO and 6 elements extracted from Cumberland, 5 of the elements are similar. It can be concluded that 83% of

elements are overlapped and 4 new elements are produced from MVSimO. With a promising comparison result, the process-based ontological model approach presented in this study, a tedious work in collecting domain knowledge and gathering experts' opinions to validate the model can be reduced. The result can also be improved with more elements provided a rigour conceptualization in FCA stage.

6 Conclusion

The concept for later version of MVSimO are easily characterized by the modeler or given to adaptation to produce a viable simulation model.Despite these achievements, several issues still need to be addressed in the future by full ontology development, in particular related to more semantic content, considering relations and combining knowledge sources. For future research, the harmonized model of MVSimO can be followed for different concepts of processes to be covered for the development of simulation model in other domain. The development of MVSimO can be revised at every stage to ensure the knowledge obtained by the end of the research is close enough to have a metamodel of A&E departments. The result of our method heavily depends on A&E domain where discrete-event simulation is being used to show the sequence of events in the departments. In addition, we are going to explore the possibility to acquire ontology data from different domain and design the ontology to suit other type of simulation like agent-based simulation and system dinamic.

REFERENCES

Aringhieri, R., M. E. Bruni, S. Khodaparasti, and J. T. van Essen. 2017. "Emergency medical services and beyond: Addressing new challenges through a wide literature review". *Computers and Operations Research* 78 (August 2016): 349–368.

Baboolal, K., J. D. Griffiths, V. A. Knight, A. V. Nelson, C. Voake, and J. E. Williams. 2012. "How efficient can an emergency unit be a perfect world model". *Emergency Medicine Journal* 29 (12): 972–977.

Bell, D., C. Cordeaux, T. Stephenson, H. Dawe, P. Lacey, and L. O'Leary. 2017. "Designing effective hybridization for whole system modeling and simulation in healthcare". *Proceedings - Winter Simulation Conference*:1511–1522.

Bell, D., S. De Cesare, N. Iacovelli, M. Lycett, and A. Merico. 2007. "A framework for deriving semantic web services". *Information Systems Frontiers* 9 (1): 69–84.

Chahal, K., T. Eldabi, and T. Young. 2013. "A conceptual framework for hybrid system dynamics and discrete event simulation for healthcare". *Journal of Enterprise Information Management* 26 (1/2): 50–74.

De Cesare, S., D. Juric, and M. Lycett. 2014. "Toward the automation of business process ontology generation". *Proceedings - 16th IEEE Conference on Business Informatics, CBI 2014* 1:70–77.

Ghanes, K., M. Wargon, Z. Jemai, O. Jouini, R. Hellmann, V. Thomas, and G. Koole. 2014. "a Comprehensive Simulation Modeling of an Emergency Department: a Case Study for Simulation Optimization of Staffing Levels". *Proceedings of the 2014 Winter Simulation Conference*:1421–1432.

Grolinger, K., M. A. M. Capretz, J. R. Marti, K. D. Srivastava, K. Grolinger, and M. A. M. Capretz. 2012. "Ontology – based Representation of Simulation Models Ontology – based Representation of Simulation Models". *The Twenty-Fourth International Conference on Software Engineering and Knowledge Engineering (SEKE)*:432–437.

Gruber, T. R. 1993. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing. KSL 93-04". *Knowledge Systems Laboratory, Stanford University*.

Guizzardi, G., and G. Wagner. 2010. "Towards an ontological foundation of discrete event simulation". *Proceedings - Winter Simulation Conference*:652–664.

Gul, M., and A. F. Guneri. 2015. "A comprehensive review of emergency department simulation applications for normal and disaster conditions". *Computers and Industrial Engineering* 83:327–344.

Günal, M. M., and M. Pidd. 2009. "Understanding target-driven action in emergency department performance using simulation.". *Emergency medicine journal : EMJ* 26:724–727.

Huang, Y.-L. 2016. "Appointment standardization evaluation in a primary care facility". *International Journal of Health Care Quality Assurance* 29 (6): 675–686.

Isern, D., and A. Moreno. 2016. "A Systematic Literature Review of Agents Applied in Healthcare". *Journal of Medical Systems* 40 (2): 1–14.

Lebcir, R., E. Demir, R. Ahmad, C. Vasilakis, and D. Southern. 2017. "A discrete event simulation model to evaluate the use of community services in the treatment of patients with Parkinson's disease in the United Kingdom". *BMC Health Serv Res* 17 (1): 50.

Lonsdale, D., D. W. Embley, Y. Ding, L. Xu, and M. Hepp. 2010. "Reusing ontologies and language components for ontology generation". *Data and Knowledge Engineering*.

Malone, J., R. Stevens, S. Jupp, T. Hancocks, H. Parkinson, and C. Brooksbank. 2016. "Ten Simple Rules for Selecting a Bio-ontology". *PLoS Computational Biology* 12 (2): 1–6.

Noy, Natalya F and McGuinness, Deborah L and others 2001. "Ontology development 101: A guide to creating your first ontology".

Pidd, M. 2014. "The ways forward : A personal view of system dynamics and discrete-event simulation". *Discrete-Event Simulation and System Dynamics for Management Decision Making*:318–336.

Ries, E. 2009. "Minimum Viable Product: a guide". *Startup Lessons Learned*.

Robinson, S. 2013. "Conceptual modeling for simulation". *2013 Winter Simulations Conference (WSC)*:377–388.

Sider, T. et al. 2001. *Four-dimensionalism: An ontology of persistence and time*. Oxford University Press on Demand.

Silver, G. A., J. A. Miller, M. Hybinette, G. Baramidze, and W. S. York. 2011. "DeMO: An Ontology for Discrete-event Modeling and Simulation". *Simulation* 87 (9): 747–773.

Simperl, E. 2009. "Reusing ontologies on the Semantic Web: A feasibility study". *Data & Knowledge Engineering* 68 (10): 905–925.

Stachowiak, H. 1973. *Allgemeine modelltheorie*. Springer.

Traore, M. K., G. Z. R. D., and B. Zeigler. 2018. "Modeling and simulation framework for value-based healthcare systems".

Wille, R. 1982. "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts". In *Ordered Sets: Proceedings of the NATO Advanced Study Institute held at Banff, Canada, August 28 to September 12, 1981*, edited by I. Rival, 445–470. Dordrecht: Springer Netherlands.

Xiao Hang Wang, Da Qing Zhang, Tao Gu, and Hung Keng Pung. 2004. "Ontology based context modeling and reasoning using OWL". *IEEE Annual Conference on Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second*:18–22.

Yevtushenko, S. 2000. "Documentation - The Concept Explorer". *Proceedings of the 7th national conference on Artificial Intelligence*:127–134.

AUTHOR BIOGRAPHIES

NURUL SALEH received a BSc (Hons) Information System from Multimedia University, Malaysia and obtained her Master's Degree in Software Engineering from Universiti Putra Malaysia. She worked in industrial sectors before completed her PhD in Computer Science at Brunel University in 2019. She is currently a researcher in the Department of Computer Science at Brunel University.

DAVID BELL is a reader in the Department of Computer Science at Brunel University where he teaches and carries out research into Software Services, Cybersecurity and Modelling & Simulation. Before returning to university, he worked for several small and large software development companies and then in Investment Banking becoming technology director leading Emerging Markets and Trade Capture initiatives globally.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.040

MAINTAINING DOMAIN SPECIFIC SIMULATION MODELLING FRAMEWORKS - A CASE STUDY ON MODELLING HYPER ACUTE STROKE PATHWAYS

Dr. Durk-Jouke van der Zee
Friso Postema

University of Groningen
P.O. Box 800, 9700 AV, Groningen
THE NETHERLANDS
d.j.van.der.zee@rug.nl,
fh.postema@gmail.com

Dr. Thomas Monks

University of Exeter Medical School
South Cloisters, St. Luke's Campus, Heavitree Road, Exeter EX1 1TE
t.w.m.monks@exeter.ac.uk

Willemijn J. Maas
Dr. Maarten M.H. Lahr
Dr. Maarten Uyttenboogaart,
Prof. dr. Erik Buskens

University Medical Center Groningen
P.O. Box 30001, 9700 RB, Groningen
THE NETHERLANDS
w.j.maas@umcg.nl
m.m.h.lahr@umcg.nl,
m.uyttenboogaart@umcg.nl,
e.buskens@umcg.nl

ABSTRACT

Conceptual simulation models capture system essentials in terms of modelling objectives, model inputs, outputs and content. Their impact on simulation study success is undenied. Guidance towards high quality conceptual models is relevant, as modelling is not easy. Modelling frameworks offer guidance by specifying what to model, by identifying modelling activities, and offering good practices and methods in doing so. In this article we explore the needs for maintenance of modelling frameworks and good policies for doing so, starting from a case study. We show how an existing modelling framework addressing hyper acute stroke pathways is extended to meet system requirements set by a new stroke treatment. We clarify how extensions imply significant efforts that may legitimate the definition of maintenance policies, clarifying what, when and how to maintain a framework. In turn, maintenance effectiveness and efficiency may rely on modelling framework setup, being transparent from a maintenance point of view.

Keywords: Conceptual modelling, Modelling Frameworks, Hyper Acute Stroke Pathway

1 INTRODUCTION

Conceptual modelling (CM) for simulation, boils down to a process of abstraction in which essential elements of a real or would be system are captured in terms of modelling objectives, model inputs, outputs and content (Robinson, 2008b). Essentially, the conceptual model serves as a linking pin between the initial problem situation and the setup of a coded model and its intended use. Clearly, quality of the conceptual model highly impacts on success of a simulation study.

CM is certainly not easy (Law, 1991). It requires bringing together domain specific knowledge and insights, and disciplinary knowledge, especially operations research, statistics, engineering, and computer science. Furthermore, modelling activities are subject to the specifics of the business

context (budgetary constraints, resource availability, time frame etc.) and (possibly conflicting) stakeholder interests. Clearly, this puts high demands on the analyst's skills. At the same time, it clarifies the need for guidance in doing so.

Robinson (2008a) distinguishes three basic approaches on simulation model development: principles of modelling, methods of simplification, and modelling frameworks. Principles of modelling refer to the general case of simulation modelling. Important examples concern, the need for model simplicity, the advocated policy of incremental modelling, and the good use of metaphors, analogies, and similarities in model creation. Methods of simplification focus on the possibility of reducing model scope and/or its level of detail in order to enhance its feasibility and/or utility, while safeguarding its validity (Van der Zee, 2019). Modelling Frameworks go beyond aforementioned approaches by specifying what to model by providing a procedural approach for detailing a model in terms of its elements, their attributes, and their relationships.

In recent years, several modelling frameworks have been developed. The main differences among modelling frameworks concern their intended field of application, scope and process support. Modelling frameworks tend to address rather broad classes of systems, like operations systems (Robinson 2008b), supply chains (Van der Zee and Van der Vorst, 2005), health systems (Kotiadis et al., 2014), the military (Pace, 1999; Pace, 2000) or the general case, i.e., discrete event dynamic systems (Arbez and Birta, 2010). Furthermore, differences among frameworks are found concerning their scope. Whereas some frameworks focus on capturing just model content (Arbez and Birta, 2010), others consider a wider angle by including an exploration of the problem context, project and modelling objectives, and/or the experimental frame, i.e., model inputs and outputs (Kotiadis, 2007; Robinson 2008b). For overviews of modelling frameworks, see Robinson (2008a, 2019), Karagoz and Demirors (2011), Van der Zee et al. (2011) and Furian et al. (2015).

In recent work we considered the development of domain specific modelling frameworks. They offer refined support for the analyst as a net effect of bringing in domain specific knowledge that is helpful in specifying model inputs, outputs and content (Monks et al., 2017). We found how the development of domain specific modelling frameworks could be legitimated vs. more general modelling frameworks by considering (i) problem complexity – being such that it requires joint effort from both experts in the domain of interest and modelling in order to tackle it, (ii) importance of the problem – being of high value to either industry or society or both, and (iii) future demand and uptake of the framework is sufficient to warrant the effort to develop it. As we felt that aforementioned requirements were met for modelling hyper acute stroke systems (HASPs), we developed a domain specific modelling framework targeting respective systems. However, a newly available stroke treatment and its implications for stroke system setup, suggest a need for maintenance, i.e., extensions, of the modelling framework to capture new system elements and their workings. Motivated by this case example, in this paper, we explore the needs for modelling framework maintenance, and good policies in doing so. To do so we rely on first experiences in extending the initial domain specific modelling framework for hyper acute stroke systems by including domain specific knowledge on the new treatment, and its implication for stroke care organization (Postema, 2019). Insights obtained are assumed to be relevant for the setup, development, uptake and use of modelling frameworks.

This paper is structured as follows. In Section 2 we characterize the hyper acute stroke pathway, and the way the new treatment impacts its organization. Next, in Section 3 the updated modelling framework is discussed. In Section 4 we characterize extensions of the new framework relative to the existing framework, and consider their implications in terms of the need for maintenance and good maintenance policies. Finally, in Section 5, we summarize main conclusions.

2 ORGANIZING THE HYPER ACUTE STROKE PATHWAY - NEW TREATMENTS

Stroke can be categorized in two subtypes: ischemic and hemorrhagic stroke, respectively 85% and 15% of the patient population. Ischemic strokes occur when a cerebral artery is occluded due to a clot and disrupts blood circulation to the brain, whereas hemorrhagic strokes are usually caused by a rupture of a vessel. This paper focuses on ischemic stroke.

Two main reperfusion treatments are available for ischemic strokes, i.e., intravenous thrombolysis (IVT) and intra-arterial thrombectomy (IAT). Efficacy of both treatments is highly time dependent, as

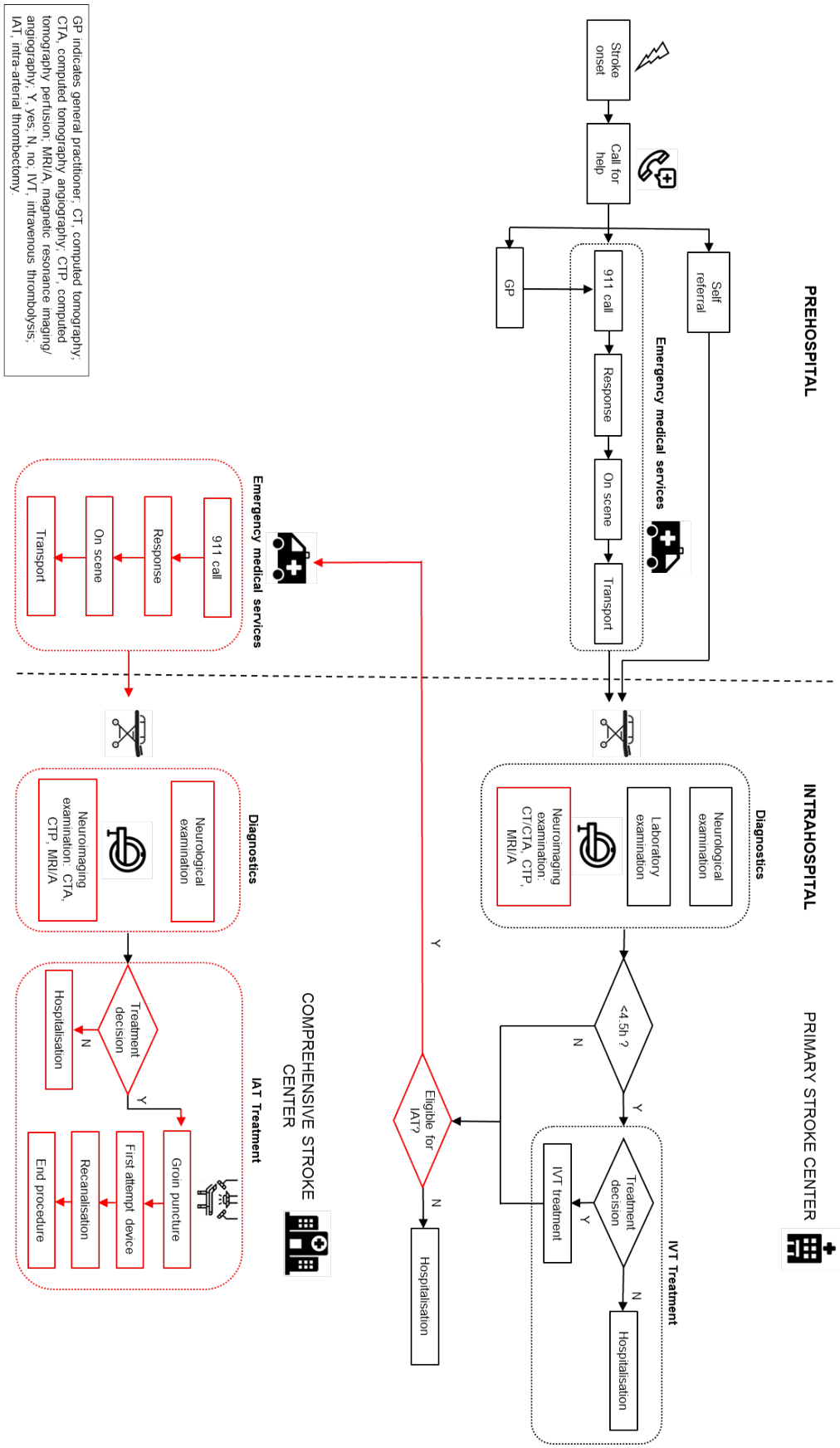


Figure 1 *Hyper Acute Stroke Pathway – Drip & Ship model, i.e., IVT treatment at Primary Stroke Center, follow-up IAT treatment at Comprehensive Stroke Center*

clarified by their respective windows of opportunity, i.e., 4.5 and 6 hours after onset (Emberson et al., 2014; Saver et al., 2016). Essentially, IVT leads to recanalization by dissolving the clot, whereas IAT attempts to remove the clot with mechanical devices. IAT is only applicable for stroke patients due to a large vessel occlusion (LVO), concerning about 7% of the stroke population (Chia et al., 2016). Whereas IVT is provided in many community hospitals acting as Primary Stroke Centers (PSCs), IAT is offered by a limited number of hospitals acting as Comprehensive Stroke Centers (CSCs). Decisions to restrict availability of IAT to designated centers are motivated by the relatively low number of stroke patients facing LVO, and high demands set on staff expertise and availability of specific resources. Current guidelines advocate that LVO patients are treated with IVT (in case of no contraindication for IVT) before receiving IAT treatment.

Whereas IVT is a well-standardised treatment that made an entry in 1995 (The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, 1995), IAT only emerged in recent years (Berkhemer et al., 2014). The availability of IAT and its proven efficacy make adjustments of the organization of the HASP a relevant issue. Two dominant organisation models emerged for serving LVO patients: the "Drip & Ship" model (DS) and the Mothership model (MS). Figure 1 characterizes the DS model. It clarifies how LVO patients may first be served by a PSC for IVT, and next be transported to the CSC for IAT treatment. Alternatively, those patients in proximity of a CSC may receive both IVT and IAT treatment at the same hospital, i.e., CSC, according to the so-called MS model. Both models imply an extension of the current IVT-based model. Changes implied relative to the IVT-based organisation model, concerning patient routing, new activities, and associated staff and resources are marked red in Figure 1.

Recently, many suggestions have been made to improve organisation models for IAT service, seeking to reduce onset to treatment time. Usually, time of treatment is related to groin puncture, see Figure 1. Suggested improvements of dominant models include, among others, expediting intra-hospital workflow, increasing EMS transportation speed by using helicopters, enabling on-call IAT services at PSCs by transporting doctors from a CSC to a PSC or pre-hospital diagnosis seeking to identify LVO patients in an early phase, allowing them to be routed to the CSC directly. Initial findings indicate how the success of proposed interventions is strongly dependent on regional characteristics like current stroke system set-up and resources, medical guidelines, and geography (Ciccone et al., 2019). Unfortunately, a one-fits-all solution does not exist. Hence, optimizing IAT services requires dedicated decision support on a regional scale. Past research has shown how simulation may be a well-qualified means for doing so (Monks et al. 2017).

3 TOWARDS AN UPDATED SIMULATION MODELLING FRAMEWORK FOR COMBINED IVT-IAT TREATMENT

This section discusses extensions of the modelling framework proposed by Monks et al. (2017) that is meant to support simulation conceptual modelling of stroke systems for IVT treatment only. Main focus will be on the approach taken for maintaining the framework, the nature of extensions - rather than their detail, and their impact on framework set-up - as measured by (the significance of) proposed changes. More detail is provided in Postema (2019) and in ongoing work.

3.1 Maintaining the framework - approach

The new modelling framework builds on the framework proposed by Monks et al. (2017) that refines the modelling framework proposed by Robinson (2008b), addressing the general class of operations systems, see Table 1 (Columns "Activity", "Detail"). In line with Robinson (2008b) Monks et al. (2017), distinguish between a number of key activities for specifying the conceptual model, see Table 1 (Column "Activity"). Each key activity is further decomposed in detailed activities. In addition, support in executing activities is offered by hinting at (i) good practices in executing them, (ii) (libraries or lists of) common choices made with respect to modelling objectives, model inputs, outputs and content (model components, attributes and their relationships) that appeal to the domain of interest, i.e. HASPs.

Table 1 *Framework extensions relative to the framework of Monks et al. (2017)*

Activity	Detail	Extensions
1. Understanding the problem situation	• Determine study population: Decide subcategories of the stroke population the study might focus on. • Assess current performance of the HASP: Interpret findings with respect to their relevance for setting the modelling objectives (2). • Map the current process: Create a starting point for determining model content (3) by building a process map that captures the status quo. • Explore decision variables for use in experimentation: Process mapping and the (initial) assessment of HASP's current performance serve as a vehicle to elicit hypotheses about delays and barriers to treatment. These hypotheses are candidate decision variables in model experimentation. Choice of decision variables may be linked to four areas(examples): ○ Pre-hospital logistics ○ Processes for identification of stroke patients in the ED ○ Communication between hospital departments ○ Work force scheduling	• Distinguish between LVO and non-LVO patients. • Establish key figures concerning speed of IAT treatment • Process maps may span both PSCs and CSCs • New areas to consider are: ○ Network topology: distribution of IAT services. ○ Communication between PSCs and CSCs. ○ Transports between PSCs and CSCs (patient or doctor transfers). ○ Pre-hospital routing decisions.
2. Setting the modelling objectives	• Select response measures and target performance levels: ○ Health outcomes (primary level): The logical and time-based results from a DES model can be used as input parameters to either clinical models of population benefit or a health economic model. ○ Logistic performance (secondary level): Treatment volume might be presented as an average rate or as a histogram of the likely range of treatment rates. ○ Target levels: Target performance levels may be set on logistic performance such as treatment rates and OTT, possibly using available benchmarks. • Determine model outputs, see response measures and target performance levels; include activity durations to assess cause and effect. • Determine model inputs that underlie experiments concerning alternative configurations of the HASP: Construct experiments that link one or more inputs to decision variables, see I. • Establish restrictions in solution finding: Consider the way budgets of care providers, physical space, and regulations etc. may restrict choice of configurations of the stroke pathway	○ Onset to groin time • Add activity durations related to patient transfer and IAT treatment. • See I decision variables (various possible inputs for areas – linking to choice of organizational model)
3. Determining model content	• Establish model scope: Identify model boundaries, by either including or excluding a representation of elements of the HASP under study as model components. Choice of model components is facilitated by a library of most common components of a HASP, see Monks et al. (2017). • Determine model detail, specifying attributes of model components. See Monks et al. (2017) for entity attributes that may be required across a broad range of objectives. • Specify assumptions underlying model content: Facilitate the interpretation of the model and its workings by making assumptions on the HASP under study explicit. See Monks et al. (2017) for common assumptions on HASP simulation models. • Making appropriate model simplifications: For common simplifications of HASP simulation models see Monks et al. (2017).	• Add model components representing intra hospital activities associated with IAT treatment, inter hospital patient and doctor transfer, and on-scene patient screening for LVO to library of components of a HASP (see Figure 1). • Include LVO/Non-LVO in patient classification. • Inter hospital patient transfers may be captured by delay functions.

Framework development relies on five sources of information: (i) literature reporting on simulation studies concerning IVT – compare Monks et al. (2017), (ii) literature reviews on alternative organization models for IAT (Dètraz et al., 2018, Ciccone et al., 2019), (iii) references as in (ii), (iv) interviews with domain experts and (v) authors' involvement in stroke research, including doing simulation studies. Initial validation of the framework is done by domain experts. Note that, so far, simulation studies concerning organization models for IAT are envisioned, but not reported in literature.

3.2 Framework extensions

Implementation of IAT treatment sets new requirements to HASP setup. In turn, these requirements are to be reflected in simulation study setup, in terms of modelling objectives, model inputs, outputs and content (Table 1 (Column "Extensions")).

Essentially, decision making on HASP setup may be considered at three levels, i.e., strategic, tactical and operational, thereby acknowledging the time horizon at which changes may be implemented, their impact on patient outcomes, and their associated costs and efforts. Choices with respect to the network topology, i.e., the distribution of stroke service over the region, are considered strategic decisions. IAT services are typically distributed over a restricted number of hospitals, due to the low number of patients involved, requirements set on expertise and (staff) resources, and natural and organic growth of stroke services already located and distributed in certain regions. The tactical level considers patient routing along stroke services, compare the DS and MS models. Finally, operational level decisions involve expediting care and transport services along the pathway. Decision variables related to each level, candidating as model inputs, are shown in Table 1 (Column "Extensions"), see activities 1 and 2.

Current clinical practice is dominated by the DS and MS organizational models. Taken together they organize regional stroke care. Both models involve system elements and activities that are new relative to the "classic" IVT-only based organization models. Newly proposed interventions (see Section 2) bring further elements and detail. Elements identified are to be reflected in new types of model components and their attributes for specifying model content, see Table 1 (Column "Extensions"), activity 3. In turn, new methods or rules for model simplification, or common assumptions, may guide component use and choice of their detail, in order to enhance model utility and feasibility.

Choice of model outputs is influenced by the new spectrum of system elements, and associated decision variables – linking to those elements open for change. To assess effects of new model inputs on system performance, further detailed outputs are required, especially for capturing delays associated with new care and transport services.

4 DISCUSSION – MODELLING FRAMEWORKS AS ASSETS

4.1 The need for maintenance

We observed a need for modifications of a domain specific modelling framework for HASP simulation, given the availability of new treatments, and innovations foreseen in service delivery, having significant impact on organizing HASPs. Such system changes are likely to occur, also in other domains. Hence, domain specific modelling frameworks, being tailored towards system specifics for a domain, cannot do without maintenance, in order to keep up their service levels. So far, the observed need for maintenance has not been acknowledged in literature.

4.2 What has been done?

Starting from several information sources, including literature, interviews with domain experts and authors' involvement in stroke research, including simulation studies, modelling framework extensions are developed. Extensions concerned (libraries or lists of) common choices made with respect to modelling objectives, model inputs, outputs and content (model components, attributes and

their relationships) that appeal to the domain of interest, i.e. HASPs. Essentially, changes, despite being significant, do not address the framework core, i.e., key modelling activities.

4.3 When to update?

The need to do maintenance was fostered by the authors' wishes of doing HASP simulation with IAT in the near future. While the framework serves their purposes, one may wonder about the timing of the initiative. At the decision moment no evidence of simulation studies with IAT was available. Furthermore, significant learning effects are observed in clinical practice in mastering the new DS and MS organization models, while new organization models are introduced at high pace.

4.4 Investment made

Efforts made, i.e., several months of full-time work, do classify modelling framework extensions as major maintenance. Much time is invested by one of the authors in familiarizing himself with the original framework – not being one of its developers, and not being involved in the domain before. Increased complexity and uncertainty associated with organization models for IAT further add to this, also see above (When to update?). Findings suggest how framework maintenance may boil down to a significant investment, requiring a clear project definition and tailored team – being familiar with the domain.

4.5 Design for maintenance

The observed need for modelling framework maintenance is clear. However, the way domain-based changes are to be related to the framework is somewhat less clear. Which elements have to be considered for possible modification? This raises issues like standards for framework set-up, and the modularity of their setup.

4.6 Maintenance policies – what and how to respond to?

Our case study considers an incident – an observed need for major maintenance of domain specific elements of the modelling framework due to significant changes of referent systems. What are possible other needs: improved modelling methodology, new methods, new best practices, ….? How to respond to this? Should we link this to timewise or status dependent triggers? The issue is relevant for simulation users in industry and education.

4.7 Limitations

Our findings are based on a single case study. Clearly, further studies are required to explore issues raised in greater depth. However, the study clarifies how the issue is very much there – in practice where modelling frameworks whether explicit or implicit are the practitioner's or student's assets that are in need of regular maintenance.

5 CONCLUDING REMARKS

In this article we explore the needs for maintenance of simulation modelling frameworks, requirements maintenance imposes on framework set-up, structure of maintenance policies and maintenance process. Starting from a case study on a new modelling framework for HASP simulation we establish a general need for maintenance of modelling frameworks – to benefit educators, students and simulation practitioners.

Case findings suggest how maintenance may take considerable efforts of developers in familiarizing with the original framework, the domain and its new facets. Therefore, policies are required clarifying what, when and how to maintain. Tailoring framework design towards its maintenance is considered instrumental for effective and efficient maintenance.

Future research is directed towards developing modelling frameworks for HASPs, including non-ischemic stroke patients, framework maintenance and framework modularity and standards. A starting point may be found in related research on software maintenance and quality (Malhotra and Chug, 2016; Stevenson and Wood, 2018).

REFERENCES

Arbez G and Birta L (2011). The ABCmod Conceptual Modelling Framework. In: Robinson S, Brooks R J, Kotiadis K. and Van der Zee D J (eds). *Conceptual Modelling for Discrete-Event Simulation*. CRC/Taylor & Francis: Boca Raton, pp 133-178.

Berkhemer O A, Fransen P S, Beumer, D et al. (2015). A Randomized Trial of Intraarterial Treatment for Acute Ischemic Stroke. *New England Journal of Medicine* **372**: 11-20.

Chia NH , Leyden J M, Newbury J, et al. (2016). Determining the Number of Ischemic Strokes Potentially Eligible for Endovascular Thrombectomy: A Population-Based Study. *Stroke* **47**:1377-1380.

Ciccone, A, Berge, E and Fischer, U (2019). Systematic review of organizational models for intra-arterial treatment of acute ischemic stroke. *International Journal of Stroke* **14(1)**: 12–22.

Détraz, L, Ernst, M and Bourcier, R (2018). Stroke Transfer and its Organizational Paradigm Review of Organizational Paradigms and the Impact on Outcome. *Clinical Neuroradiology* **28**: 473–480.

Emberson J, Lees K R, Lyden P, et al. (2014). Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. *Lancet* **384(9958)**: 1929-1935.

Furian, N, O'Sullivan, M, Walker, C, Vössner, S and Neubacher, D (2015). A conceptual modeling framework for discrete event simulation using hierarchical control structures. *Simulation Modelling Practice & Theory* **56**: 82–96.

Karagoz N A and Demirors O (2010). Conceptual Modelling Notations and Techniques. In: Robinson S, Brooks R J, Kotiadis K. and Van der Zee D J (eds). *Conceptual Modelling for Discrete-Event Simulation*. CRC/Taylor & Francis: Boca Raton, pp 179-210.

Kotiadis K, Tako A and Vasilakis C (2014). A participative and facilitative conceptual modelling framework for discrete event simulation studies in healthcare. *Journal of the Operational Research Society* **65(2)**: 197-213.

Law, A M (1991). Simulation Model's Level of Detail Determines Effectiveness. *Industrial Engineering*, **23 (10)**: 16-18.

Malhotra, R and Chug, A (2016) Software Maintainability: Systematic Literature Review and Current Trends. *International Journal of software engineering and knowledge engineering* **26**(8): 1221-1253.

Monks, T, van der Zee, D J, Lahr, M, Allen, M, Pearn, K, James, M A, Buskens, E and Luijckx, G J (2017). A framework to accelerate simulation studies of hyperacute stroke systems. *Operations Research for Health Care* **15**: 57-67.

Pace D K (1999). Development and documentation of a simulation conceptual model. In *Proceedings of the 1999 Fall Simulation Interoperability Workshop*. Available via www.sisostds.org.

Pace D K (2000). Simulation conceptual model development. In Proceedings of the 2000 Spring Simulation Interoperability Workshop. Available via www.sisostds.org.

Postema, F H (2019). *A modeling framework for hyperacute stroke care simulation with endovascular treatment*. MSc thesis, University of Groningen.

Robinson S (2008a). Conceptual modelling for simulation Part I: definition and requirements. *Journal of the Operational Research Society* **59(3)**: 278-290.

Robinson S (2008b). Conceptual modelling for simulation Part II: a framework for conceptual modelling. *Journal of the Operational Research Society* **59(3)**: 291-304.

Robinson S (2019). Conceptual modelling for simulation: Progress and grand challenges. Journal of Simulation (in press).

Saver J L, Goyal M, van der Lugt A, et al. (2016). Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-analysis. *JAMA* **316(12)**: 1279.

Stevenson, J and Wood, M (2018) Recognising object-oriented software design quality: a practitioner-based questionnaire survey. *Software Quality Journal* **26**(2): 321-365.

The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (1995). Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine* **333**:1581-1587.

van der Zee, D J (2019). Model Simplification in Manufacturing Simulation - Review and Framework. *Computers & Industrial Engineering*: **127**: 1056-1067.

Van der Zee D J and Van der Vorst J G A J. (2005). A modeling framework for supply chain simulation – Opportunities for improved decision-making. *Decision Sciences* **36(1)**: 65-95.
Van der Zee D J, Brooks R J, Robinson S and Kotiadis K (2011). Conceptual Modelling: Past, Present and Future. In: Robinson S, Brooks R J, Kotiadis K. and Van der Zee D J (eds). *Conceptual Modelling for Discrete-Event Simulation*. CRC/Taylor & Francis: Boca Raton, pp 473-490.

AUTHOR BIOGRAPHIES

DURK-JOUKE VAN DER ZEE is associate professor of Operations at the Faculty of Economics and Business, University of Groningen, The Netherlands. He holds a MSc and Phd in Industrial Engineering. His research interests include health care logistics engineering, simulation methodology and applications, simulation & serious gaming, and manufacturing planning & control. He is a member of the INFORMS-SIM. His webpage is http://www.rug.nl/staff/d.j.van.der.zee.

FRISO POSTEMA is a junior researcher at the Faculty of Economics and Business, University of Groningen, The Netherlands. Currently he is working as a junior supply chain analyst in the private sector. His research interests include simulation methodology and applications in health care logistics.

THOMAS MONKS is an Associate Professor of Health Data Science at the University of Exeter Medical School and The Institute of Data Science and AI, University of Exeter. He is also Turing Fellow at The Alan Turing Institute and Honorary researcher at the Clinical Operational Research Unit, UCL. He holds a BSc in Computer Science and Mathematics, MSc in Operational Research and PhD in Simulation Modelling. Between 2014 and 2019 he led an applied healthcare Data Science team funded by the NIHR. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His speciality is simulation of healthcare systems of unscheduled and emergency care. He was co-chair of the UKs Simulation Workshop 2016 and 2018. His webpage is: https://www.turing.ac.uk/people/researchers/thomas-monks

WILLEMIJN J. MAAS is a PhD Candidate at the Department of Neurology and the HTA unit of the Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands. Her research interests include the organization of acute stroke care and discrete-event simulation.

MAARTEN M. H. LAHR is a postdoctoral researcher at the HTA unit of the Department of Epidemiology, and member of the Healthy Ageing Team, University of Groningen, University Medical Center Groningen, The Netherlands. His research interests include organizational models for acute stroke care and discrete-event simulation. His website is http://www.rug.nl/staff/m.m.h.lahr/.

MAARTEN UYTTENBOOGAART is an interventional neurologist at the department of Neurology and Radiology, University Medical Center Groningen. He is involved in the Cerebrovascular Unit, and Principal Investigator at the Research School of Behavioural and Cognitive Neurosciences Groningen (BCN) and KOLFF institute. His research topics are clinical research on prognostic factors in and treatments for acute ischemic stroke, clinical research on neuroprotection in acute stroke and research on organization and implementation of acute stroke care.

ERIK BUSKENS is Professor of Health Technology Assessment University Medical Center Groningen, and faculty of Economics and Business University of Groningen, The Netherlands. His research interests among others encompass the association and mutual dependencies between technological advances and the factual organization of care to attain the desired health and possibly economic benefits.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.041

WHY MACHINE LEARNING INTEGRATED PATIENT FLOW SIMULATION?

Dr. Tesfamariam M. Abuhay

University of Gondar
Gondar, Ethiopia
tesfamariam.mabuhay@uog.edu.et

Dr. Adane L. Mamuye

University of Gondar
Gondar, Ethiopia
adane.letta@uog.edu.et

Prof. Stewart L. Robinson

Loughborough University
Loughborough, UK
S.L.Robinson@lboro.ac.uk

Dr. Sergey V. Kovalchuk

ITMO University
Saint Petersburg, Russia
kovalchuk@itmo.ru

ABSTRACT

Patient flow analysis can be studied from clinical and/or operational perspective using simulation. Traditional statistical methods such as stochastic distribution methods have been used to construct patient flow simulation sub-models such as patient inflow, Length of Stay (LoS), Cost of Treatment (CoT) and Clinical Pathway (CP) models. However, patient inflow demonstrates seasonality, trend and variation over time. LoS, CoT and CP are significantly determined by patients' attributes and clinical and laboratory test results. For this reason, patient flow simulation models constructed using traditional statistical methods are criticized for ignoring heterogeneity and their contribution to personalized and value-based healthcare. On the other hand, machine learning methods have proven to be efficient to study and predict admission rate, LoS, CoT, and CP. This paper, hence, describes why coupling machine learning with patient flow simulation is important and proposes a conceptual architecture that shows how to integrate machine learning with patient flow simulation.

Keywords: Patient Flow Simulation, Machine Learning, Health Services Research, Conceptual Modelling.

1 INTRODUCTION

Patient flow analysis to and in a hospital has been one of the hot research areas in health services research and health economics (Kreindler, 2017; Gualandi et al, 2019). It can be conducted from clinical and/or operational perspective (Côté, 2000) in a single unit/department (e.g., ambulatory care unit (Santibáñez, 2009), intensive care unit (Benjamin and Christensen, 2012), emergency department (Konrad et al, 2013; Cocke et al, 2016; Hurwitz et al, 2014), surgery department (Antonelli et al, 2014; Azari-Rad, 2014)) or in multiple units/departments (e.g., Abuhay et al (2016 and 2020), Kovalchuk et al (2018), Suhaimi et al (2018)) of a hospital using simulation.

Simulation allows to represent complex systems (Anatoli, 2013) and produces a range of data to support decision making (Monks et al, 2016). Patient flow analysis using simulation can be used to reduce the chances of failure, to meet specifications, to eliminate unforeseen bottlenecks, to prevent under or over-utilization of resources, to reduce crowding, to improve clinical pathways and system performance (Maria and Anu, 1997). According to Gunal (2012), simulation methods that are employed to study patient flow are generally classified into three categories: Discrete-Event Simulation (DES), Agent-Based Simulation (ABS) and System Dynamics (SD).

To construct components/sub-models (e.g., patient inflow, Length of Stay (LoS), Cost of Treatment (CoT) and clinical pathway (CP) models) of a patient flow simulation, traditional statistical methods such as stochastic distribution (discrete and continues) methods have been used. However, patient inflow or admission rate data demonstrate seasonality and trend and it also varies from hour to hour, day to day, week to week, month to month and year to year (Nas and Koyuncu, 2019). This makes modelling patient admission rate using stochastic (discrete and/or continues) distribution difficult. LoS, CoT and CP are also significantly determined by a patient's attributes such as age, gender, comorbidity, genomic makeup and clinical and laboratory test results (Bramkamp, 2007; Noohi et al, 2020; Siddiqui et al, 2018; Zhang et al, 2010). For this reason, patient flow simulation models are criticized for ignoring heterogeneity (Zaric, 2003) and their contribution to personalized medicine (Schleidgen et al, 2013) and value-based healthcare (Brown, 2005; Traoré, 2019) is now in question.

On top of that, decision makers have a doubt on validity and credibility of patient flow simulation due to significant uncertainty in the patient flow simulation models (Kovalchuk et al, 2018). This, in turn, affects acceptance level and applicability of patient flow simulation models. Patient flow simulation thus needs an accurate estimation model of patient arrival, LoS, CoT and CP (Nas and Koyuncu, 2019).

On the other hand, Electronic Health Record (EHR) (Ambinder, 2005) presents an opportunity by generating big data that can be employed to construct data-driven clinical and/or operational decision support tools that facilitate modelling, analysing, forecasting and managing healthcare.

Machine Learning (ML) (Ngiam and Khor, 2019), using EHR data as an input, has been widely used to study, discover patterns and predict patients' admission rate or demand for healthcare (Asheim, 2019; Luo, 2017; Hong, 2018), LoS (Daghistani, 2019; Taleb, 2017), CoT (Bremer et al, 2018; Jödicke et al, 2019), and CP (Kovalchuk et al 2018; Allen et al, 2019; Prokofyeva and Zaytsev, 2020; Funkner, 2017), to mention a few.

This paper, hence, aims to describe why coupling machine learning with patient flow simulation is important and proposes a conceptual framework that shows how to integrate machine learning with patient flow simulation.

The proposed architecture may improve credibility and acceptance of patient flow simulation as it expands the knowledge stock of general and domain-specific conceptual modelling (Robinson, 2020) of patient flow simulation. It may also foster personalized medicine and value-based healthcare because both concepts promote individual-patient-based healthcare with high-quality, low cost and wide access instead of "one-model-fits-all" approach (Schleidgen et al, 2013; Brown, 2005; Traoré, 2019).

The rest of the paper is organized as follows: Section 2 discusses related works, Section 3 presents why machine learning integrated patient flow simulation is important, Section 4 illustrates conceptual architecture of the proposed model and Section 5 presents conclusion.

2 RELATED WORKS

Several studies employed computer simulation methods such as SD, DES, and ABS for modelling patient flow in a single or multiple departments. However, there is still a high variation of uncertainty in patient flow (Kovalchuk et al, 2018) because the performance of patient flow models highly depends on input variables/data.

To minimize the uncertainties and improve accuracy of patient flow simulation, data-driven methods were proposed to supplement the existing simulation modelling techniques. For example, Kovalchuk et al (2018) studied simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. ML was applied to identify classes of clinical pathways, capturing rare events and variation in patient.

Nas and Koyuncu (2019) proposed an Emergency Department (ED) capacity planning using a Recurrent Neural Network (RNN) and simulation approach. The main objective of this study was to determine patients' arrival times and optimum number of beds in an ED by minimizing the patients' LoS. The outcomes of ML model, hourly patient arrival rates prediction model, was used as input variables.

However, these studies inadequately described why machine learning integrated patient flow simulation was important. Previous studies also did not propose an architecture that shows how to couple machine learning with patient flow simulation.

3 WHY MACHINE LEARNING INTEGRATED PATIENT FLOW SIMULATION?

The patient flow simulation model mainly contains two sub-models: patient inflow simulation model and in-hospital patient flow simulation model. The first sub-model simulates patients' arrival to a specific department of a hospital in hourly, daily, weekly or monthly period. Discrete stochastic distribution methods, mainly Poisson distribution (Banks, 2005), were used to develop patient admission rate simulation model. However, patient inflow data exhibits trend, seasonality and/or variation as shown in Figure 1 and Figure 2 (All Figures in this paper were generated using the Acute Coronary Syndrome (ACS) patients' data collected from the Almazov National Medical Research Centre[1], Saint Petersburg, Russia).

Zhang et al (2020) investigated emergency patient flow forecasting in the radiology department and their result implied that ward patient visits had significant nonlinear trend. i.e., the patient arrival problems are generally related hourly, daily, weekly, or monthly. This significantly affect planning and allocation of resources such as bed, health professionals, and diagnosis and treatment equipment.

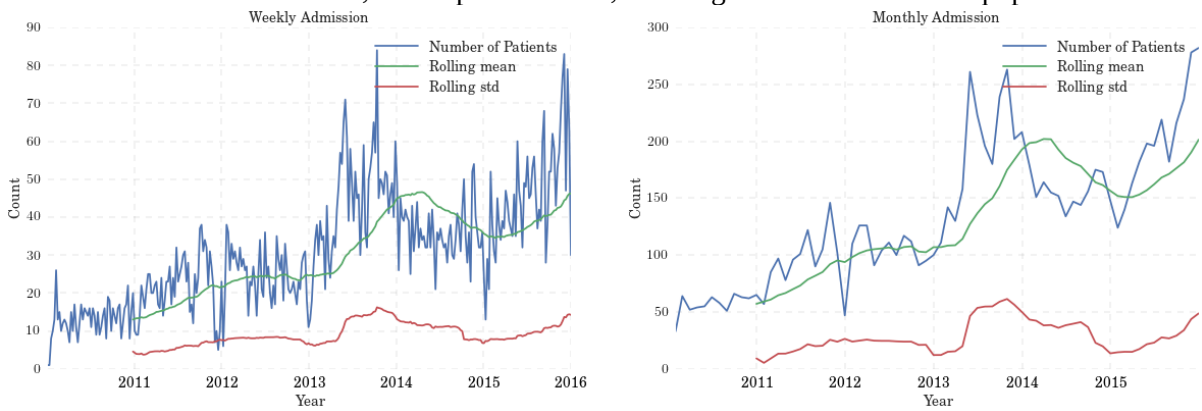


Figure 1 *Trend and Seasonality of Weekly and Monthly Patient Inflow*

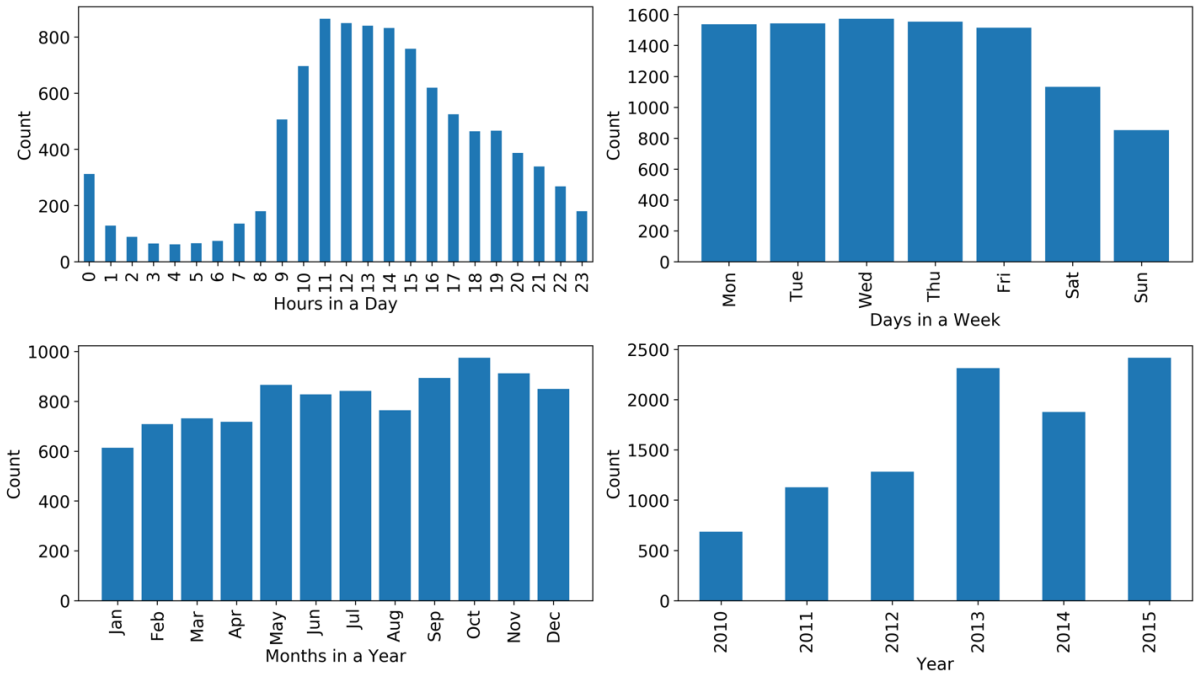


Figure 2 *Variation of Patient Inflow over Hours, Week Days, Months and Years*

So as to provide timely, personalized and value-based healthcare, understanding the trend of patient flow to a hospital and develop a prediction model that can identify and understand trend, seasonality

[1] http://www.almazovcentre.ru/?lang=en, accessed February 2021.

and variation of patient flow is vital. Hence, instead of modelling patient inflow with stochastic distribution (counting functions) methods, ML algorithms for time series analysis and prediction (Brockwell, 2010) are ideal because they are able to pinpoint seasonality and trend and predict the future patient inflow, in different time period, based on historical data. Woo et al (2018) also showed that comprehensive elements of patient history (e.g., previous healthcare usage statistics, past medical history, historical labs and vitals, prior imaging counts, and outpatient medications, and demographic details such as insurance and employment status) improved patient inflow prediction performance significantly.

Machine learning, therefore, would have twofold objectives: simulating patient inflow and/or predicting future patient inflow. This allow effective and efficient planning and resources allocation, while reducing over- and/or under-utilization.

The second sub-model, which is in-hospital patient flow simulation model, mimics movement of patients through multiple clinical and/or operational processes in a single or multiple departments. In the case of multiple departments, the in-hospital model simulates movement of patients from one department to another based on transaction matrix. This model may have sub-models such as LoS estimation model, CoT estimation model and CP estimation model.

The LoS is a significant indicator of the effectiveness and efficiency of a hospital management. LoS has been used as a surrogate to evaluate the utilization of resources, quality and efficiency of care, costs of treatment, patient experience, and planning capacity in a hospital (Papi, 2016; Verburg et al, 2014). Both LoS and CoT sub-models are usually developed using univariate density estimation methods such as Lognormal, Weibull, and Gamma. However, (Ickowicz et al, 2016; Lee et al, 2011; Houthooft et al, 2015) mentioned that none of them seemed to fit satisfactorily in a wide variety of samples.

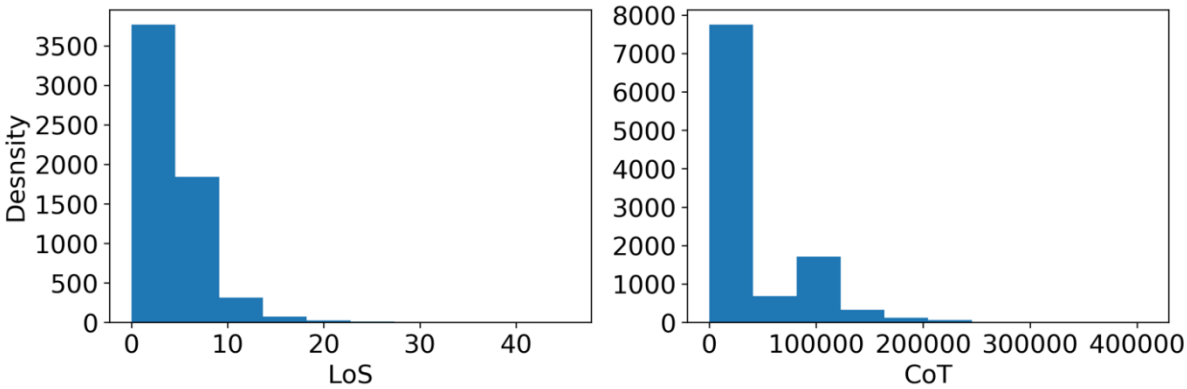


Figure 3 *Distribution of Length of Stay (LoS) in Days and Cost of Treatment (CoT) in RUB*[2]

Instead, the assumption of heterogeneous sub-populations would be more appropriate because the probability distribution of both LoS and CoT is positively skewed, multi-modal (See Figure 3) and significantly vary between diagnosis-related groups (DRGs) and correlated with characteristics of patients such as age, gender, number of comorbidity (see Figure 4) and the like (Daghistani et al, 2019; Bramkamp et al, 2007; Noohi et al, 2020; Siddiqui et al, 2018; Ickowicz et al, 2016). This limits the use of inference techniques based on normality assumptions (Ickowicz et al, 2016).

According to Ngiam and Khor (2019), ML provides flexibility and scalability compared with traditional statistical methods because it allows to analyse diverse data types and incorporate them into predictions for disease risk, diagnosis, prognosis, appropriate treatments, LoS and CoT. Thus, modelling LoS and CoT using ML while constructing patient in-hospital flow simulation allow considering heterogeneous sub-populations based on their characteristics. This may improve accuracy and credibility of patient flow simulation model.

[2] Russian Federation Currency

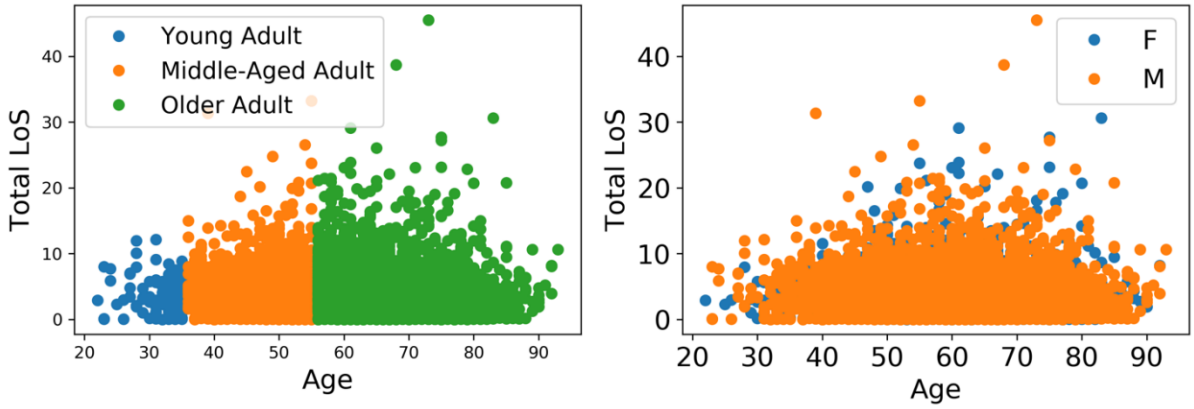


Figure 4 *Distribution of LoS based on Age Category and Gender*

Another sub-model of patient flow simulation is CP prediction model. CP specify the categories of care, activities, and procedures that need to be conducted for a group of patients until they are discharged from the hospital (Aspland et al, 2019). Modelling the processes in a healthcare system plays a large role in understanding its activities and serves as the basis for increasing the efficiency of medical institutions (Prokofyeva and Zaytsev, 2020).

Clinical pathways have been generally modelled based on probability law using transaction matrix. However, patient in-hospital flow is very complex process due to dissimilar and multiphase pathways and the innate uncertainty and variability of care processes due to patients' attributes and their previous history. On top of this, CP identification and prediction involves analysis of comprehensive patient information (Kovalchuk, 2018; Aspland et al, 2019) that cannot be achieved or modelled with transaction matrix. Hence, pathway analysis and prediction model using ML while constructing in-hospital patient flow simulation may foster personalized medicine and improve efficiency of healthcare provision.

4 HOW TO INTEGRATE MACHINE LEARNING WITH PATIENT FLOW SIMULATION?

Figure 5 shows how machine learning can be coupled with patient flow simulation. Electronic medical record (EHR), a digital version of a patient's paper chart, generates huge amount of data that can be used to construct data-driven decision support tools that facilitate modeling, analyzing, forecasting and managing operational and/or clinical processes of healthcare.

Different kind of data about admission rate, LoS, CoT, characteristics of patients (e.g., age, gender, genetic makeup and etc.), clinical pathways in the form of event log, clinical and laboratory test results can be extracted from the EHR. This would allow to develop data-driven patient flow simulation (Ambinder, 2005).

Instead of modelling patient inflow using stochastic distribution, it can be formulated as time series problem and modelled using ML algorithms for times series problem. So that factors affecting patient flow to a hospital can be considered and seasonality and trending nature of patient inflow can be captured.

Patient inflow can be modelled as a univariate or multivariate data. In the case of modelling patient inflow as univariate data, hourly, daily, weekly, monthly or yearly number of patients only is extracted and used as an input. Whereas, patient inter-arrival rate (hourly, daily, weekly, monthly or yearly) can be integrated with third-party data such as weather, demographic structure of a population, pandemic and natural and/or human made disasters and used as an input to model patient inflow as multivariate data. Data integration and preprocessing tasks can be applied so as to handle missing values, normalize the data and make the data suitable for ML algorithms.

The patient inflow prediction model can be developed for hourly, daily, weekly, monthly or yearly period. Zhang et al (2020) predicted emergency patient flow in the radiology department by constructing six linear (autoregressive integrated moving average and least absolute shrinkage and selection operator) and nonlinear models (linear-and-radial support vector regression models, random forests and

adaptive boosting) and considering the lag effects and corresponding time factors. The data was collected from the radiology department and the performance of the models was measured using mean absolute percentage error.

Khaldi et al (2019) attempted to forecast weekly patient visits to ED by combining Artificial Neural Networks (ANNs) with a signal decomposition technique named Ensemble Empirical Mode decomposition (EEMD). Seven years of univariate time series data of weekly demand was collected from ED and the proposed models were evaluated using root mean square error (RMSE), mean absolute error (MAE) and correlation coefficient (R).

The ML-based patient inflow prediction model is the starting point of the patient flow simulation process. It can be also attached to each department so that demand and supply can be analyzed, forecasted and managed at department or operational level, if the simulation model encompasses multiple departments/units.

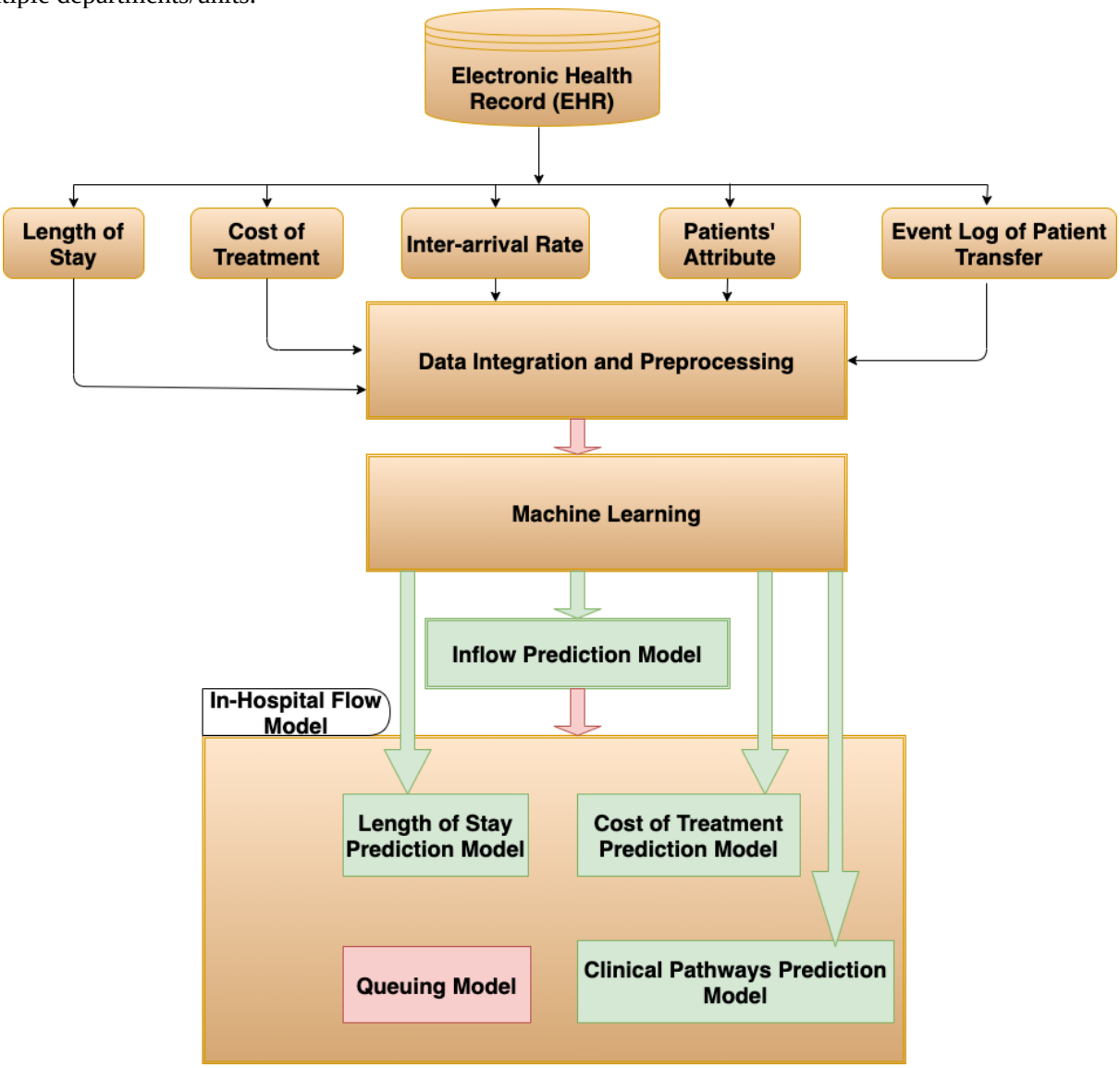


Figure 5 *Conceptual architecture of the proposed model*

Different factors, including characteristics of patients, affect LoS, CoT and CP. Hence, formulating LoS and CoT as regression problem allow consideration of different determinate factors such as characteristics of patients, clinical and laboratory tests results and co-morbidities. So that sub-populations of patients could be managed accordingly and unnecessary LoS can be reduced, which results in decreased risk of healthcare acquired infection, improvement of quality of treatment, reduction of CoT, and increased availability of free beds for needy patients (Baek et al, 2018).

Linear and nonlinear models can be used to model and predict LoS. Baek et al (2018), for example, analyzed LoS using electronic health records and machine learning techniques of multiple regression analysis. Muhlestein et al (2019) trained 29 ML algorithms, including tree-based models, linear classifiers, support vector machines, neural networks, and naïve Bayes classifiers, on 26 preoperative variables, collected from publicly available NIS database, to predict LoS following craniotomy for brain tumor. The root mean square logarithmic error (RMSLE) was applied to evaluate the performance of the ML and top performing algorithms were combined to form an ensemble. Mekhaldi et al (2020) compared two ML methods, the Random Forest (RF) and the Gradient Boosting model (GB), to predict the LoS based on an open source dataset and the Mean Square Error (MAE), the R-squared ($R2$) and the Adjusted R-squared (Adjusted $R2$) metrics were used to evaluate the performance of the models. The same fashion can be also applied to analyze and model CoT as both LoS and CoT demonstrate similar behavior.

The ML-based LoS and CoT prediction models will be attached to each department or healthcare service in a department so that demand and supply can be analyzed, forecasted and managed at operational level. In order to attach LoS and CoT prediction model to each department/healthcare services in the in-hospital patient flow simulation, generator of characteristics of patients should be constructed using continues and/or discrete probability functions according to the data type of the patients' characteristics.

Clinical pathways or patient trajectories analysis and prediction can be formulated as a clustering problem using clustering methods. Event log data about movement of patients, in combination with patient characteristics, previous clinical history, current laboratory and clinical test results, can be used to model and predict clinical pathways in a single department or in the entire hospital. Prokofyeva and Zaytsev (2020) analyzed clinical pathways in medical institutions using hard and fuzzy clustering methods and public data. Allen et al (2019) studied significance of machine learning to analyze clinical pathway and enhance clinical audits. Secondary data collected during routine care were used, following a comparison of methods, a random forests method was chosen and the model was validated using stratified tenfold validation. Allen et al (2019) proposed data-driven modeling of clinical pathways using EHR to cluster patients into groups based on their movements/clinical pathways during their stay in hospital. Kovalchuk et al (2018) studied simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. ML was applied to identify classes of clinical pathways, capturing rare events and variation in patient using clustering.

5 CONCLUSION

The aim of this paper is to describe why machine learning integrated patient flow simulation is important and to propose a conceptual architecture that shows how to integrate machine learning with patient flow simulation.

Traditional statistical methods such as stochastic distribution (discrete and continues) methods have been used to construct sub-models (e.g., patient inflow, Length of Stay (LoS), Cost of Treatment (CoT) and clinical pathways models) of patient flow simulation model. However, patients' admission data demonstrate seasonality, trend and variation over time. LoS, CoT and clinical pathways are also significantly determined by a patient's attributes such as age, gender, comorbidity, genomic makeup and clinical and laboratory test results. For this reason, patient flow simulation models were criticized for ignoring heterogeneity and their contribution to personalized medicine and value-based healthcare is now in question.

On the other hand, machine learning methods have proven to be efficient to study and predict patients' admission rate, bed capacity, LoS, CoT, and clinical pathways. This paper, hence, describes why coupling machine learning with patient flow simulation is important and proposes a conceptual architecture that shows how to integrate machine learning with patient flow simulation models.

ACKNOWLEDGMENTS

The authors thank Almazov National Medical Research Centre (Saint Petersburg, Russia) for providing anonymized data for this study.

REFERENCES

Kreindler S. A. (2017), "The three paradoxes of patient flow: An explanatory case study," BMC Health Serv. Res., vol. 17, no. 1.

Gualandi R. et al (2019), "Improving hospital patient flow: a systematic review," Business Process Management Journal. Emerald Group Publishing Ltd.

Côté M. J. (2000), "Understanding Patient Flow," Decis. Line.

Santibáñez P. et al (2009), "Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency's ambulatory care unit through simulation," Health Care Manag. Sci., vol. 12, no. 4, pp. 392–407.

Benjamin B. A. and Christensen A. (2012), "Improving ICU patient flow through discrete-event simulation," Massachusetts Institute of Technology.

Konrad R. et al (2013), "Modeling the impact of changing patient flow processes in an emergency department: Insights from a computer simulation study," Oper. Res. Heal. Care, vol. 2, no. 4, pp. 66–74.

Cocke S. et al (2016), "UVA emergency department patient flow simulation and analysis," 2016 IEEE Syst. Inf. Eng. Des. Symp., pp. 118–123.

Hurwitz J. E. (2014), "A flexible simulation platform to quantify and manage emergency department crowding," BMC Med. Inform. Decis. Mak., vol. 14, no. 1, p. 50.

Antonelli D. et al (2014), "Simulation-Based Analysis of Patient Flow in Elective Surgery," in Proceedings of the International Conference on Health Care Systems Engineering, vol. 61, pp. 87–97.

Azari-Rad S. et al (2014), "A simulation model for perioperative process improvement," Oper. Res. Heal. Care, vol. 3, pp. 22–30.

Abuhay T. M. et al (2016), "Simulation of Patient Flow and Load of Departments in a Specialized Medical Center," in Procedia Computer Science, vol. 101, pp. 143–151.

Abuhay T. M. et al (2020), "Constructing Holistic Patient Flow Simulation Using System Approach," in Computational Science – ICCS 2020, vol. 12140, Nature Publishing Group, pp. 418–429.

Kovalchuk S. V. et al (2018), "Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification," J. Biomed. Inform., vol. 82, pp. 128–142.

Suhaimi N. et al (2018), "BUILDING A FLEXIBLE SIMULATION MODEL FOR MODELING MULTIPLE OUTPATIENT ORTHOPEDIC CLINICS," in 2018 Winter Simulation Conference (WSC), pp. 2612–2623.

Anatoli D. and German R. (2013), "PROSPECTIVE HEALTHCARE DECISION-MAKING BY COMBINED SYSTEM DYNAMICS, DISCRETE-EVENT AND AGENT-BASED SIMULATION," in Proceedings of the 2013 Winter Simulation Conference, pp. 270–281.

Monks T. et al (2016), "A modelling tool for capacity planning in acute and community stroke services," BMC Health Serv. Res., vol. 16, pp. 1–8.

Maria A. (1997), "Introduction to modeling and simulation" in Proceedings of the 29th conference on Winter simulation - WSC '97, pp. 7–13.

Gunal M. M. (2012), "A guide for building hospital simulation models," Heal. Syst., vol. 1, no. 1, pp. 17–25.

Nas S. and Koyuncu M. (2019), "Emergency Department Capacity Planning: A Recurrent Neural Network and Simulation Approach," Comput. Math. Methods Med., vol. 2019.

Bramkamp M. et al (2007), "Determinants of costs and the length of stay in acute coronary syndromes: A real life analysis of more than 10 000 patients," Cardiovasc. Drugs Ther., vol. 21, no. 5, pp. 389–398.

Noohi S. et al (2020), S. Kalantari, S. Hasanvandi, and M. Elikaei, "Determinants of Length of Stay in a Psychiatric Ward: a Retrospective Chart Review," Psychiatr. Q., vol. 91, no. 2, pp. 273–287.

Siddiqui N. et al (2018), "Hospital length of stay variation and comorbidity of mental illness: A retrospective study of five common chronic medical conditions," BMC Health Serv. Res., vol. 18, no. 1, p. 498.

Zhang W. et al (2010), "A two-stage machine learning approach for pathway analysis," in Proceedings - 2010 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2010, pp. 274–279.

Zaric G. S. (2003), "The Impact of Ignoring Population Heterogeneity when Markov Models are Used in Cost-Effectiveness Analysis," Med. Decis. Mak., vol. 23, no. 5, pp. 379–386.

Schleidgen S. et al (2013), C. Klingler, T. Bertram, W. H. Rogowski, and G. Marckmann, "What is personalized medicine: Sharpening a vague term based on a systematic literature review," BMC Med. Ethics, vol. 14, no. 1, p. 55.

Brown G. (2005), "Value-based medicine: the new paradigm," Curr. Opin. Ophthalmol., vol. 16, no. 3, pp. 139–140.

Traoré M. K. et al (2019), "Modeling and simulation framework for value-based healthcare systems," Simulation, vol. 95, no. 6, pp. 481–497, Jun. 2019.

Ambinder E. P. (2005), "Electronic Health Records," J. Oncol. Pract., vol. 1, no. 2, pp. 57–63.

Ngiam K. Y. and Khor I. W. (2019), "Big data and machine learning algorithms for health-care delivery," Lancet Oncol., vol. 20, no. 5, pp. e262–e273.

Asheim A. et al (2019), "Real-time forecasting of emergency department arrivals using prehospital data," BMC Emerg. Med., vol. 19, no. 1, p. 42.

Luo L. et al (2017), "Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models," BMC Health Serv. Res., vol. 17, no. 1, p. 469.

Asheim A. et al (2019), "Real-time forecasting of emergency department arrivals using prehospital data," BMC Emerg. Med., vol. 19, no. 1, p. 42.

Hong W. S. et al (2018), "Predicting hospital admission at emergency department triage using machine learning," PLoS One, vol. 13, no. 7, pp. 1–13.

Daghistani T. A. et al (2019), "Predictors of in-hospital length of stay among cardiac patients: A machine learning approach," Int. J. Cardiol., vol. 288, pp. 140–147.

Al Taleb A. R. et al (2017), "Application of data mining techniques to predict length of stay of stroke patients," in 2017 International Conference on Informatics, Health and Technology, ICIHT 2017.

Bremer V. et al (2018), "Predicting therapy success and costs for personalized treatment recommendations using baseline characteristics: Data-driven analysis," J. Med. Internet Res., vol. 20, no. 8.

Jödicke A. M. et al (2019), "Prediction of health care expenditure increase: How does pharmacotherapy contribute?," BMC Health Serv. Res., vol. 19, no. 1, p. 953.

Allen M. et al (2019), "Can clinical audits be enhanced by pathway simulation and machine learning? An example from the acute stroke pathway," BMJ Open, vol. 9, no. 9.

Prokofyeva E. and Zaytsev R. (2020), "Clinical pathways analysis of patients in medical institutions based on hard and fuzzy clustering methods," Bus. Informatics, vol. 14, no. 1, pp. 19–31.

Funkner A. A. et al (2017), "Data-driven modeling of clinical pathways using electronic health records," in Procedia Computer Science, vol. 121, pp. 835–842.

Robinson S. (2020), "Conceptual modelling for simulation: Progress and grand challenges," J. Simul., vol. 14, no. 1, pp. 1–20.

Banks J. (2005), Discrete-event system simulation, Fourth. PRENTICE-HALL INTERNATIONAL SERIES IN INDUSTRIAL AND SYSTEMS ENGINEERING.

Zhang Y. et al (2020), "Emergency patient flow forecasting in the radiology department," Health Informatics J., p. 146045822090188.

Brockwell P. J. (2010), "Time series analysis," in International Encyclopedia of Education, Elsevier Ltd, pp. 474–481.

Papi M. et al (2016), "A new model for the length of stay of hospital patients," Health Care Manag. Sci., vol. 19, no. 1, pp. 58–65.

Verburg I. W. M. et al (2014), "Comparison of Regression Methods for Modeling Intensive Care Length of Stay," PLoS One, vol. 9, no. 10, p. e109684.

Ickowicz A. et al (2016), "Modelling hospital length of stay using convolutive mixtures distributions," Stat. Med., vol. 36, no. 1, pp. 122–135.

Lee A. H. et al (2001), "Determinants of maternity length of stay: a Gamma mixture risk-adjusted model.," Health Care Manag. Sci., vol. 4, no. 4, pp. 249–55.

Houthooft R. et al (2015), "Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores," Artif. Intell. Med., vol. 63, pp. 191–207.
Aspland E. et al (2019), "Clinical pathway modelling: a literature review," Heal. Syst., pp. 1–23.
Khaldi R. et al (2019), "Forecasting of weekly patient visits to emergency department: Real case study," in Procedia Computer Science, vol. 148, pp. 532–541.
Baek H. et al (2018), M. Cho, S. Kim, H. Hwang, M. Song, and S. Yoo, "Analysis of length of hospital stay using electronic health records: A statistical and data mining approach," PLoS One, vol. 13, no. 4.
Muhlestein W. E. (2019), "Predicting Inpatient Length of Stay after Brain Tumor Surgery: Developing Machine Learning Ensembles to Improve Predictive Performance," Clin. Neurosurg., vol. 85, no. 3, pp. 384–393.
Mekhaldi R. N. (2020), "Using Machine Learning Models to Predict the Length of Stay in a Hospital Setting," in Advances in Intelligent Systems and Computing, vol. 1159 AISC, pp. 202–211.

AUTHOR BIOGRAPHIES

TESFAMARIAM ABUHAY received a BSc in Management Information Systems (MIS) from Haramaya University of Ethiopia in 2010. He completed his MSc in IT at the University of Gondar, Ethiopia in 2015 and his PhD in Computer Science at ITMO University, Russia in 2019. He is currently an Assistant Professor and the Vice Dean of College of Informatics at the University of Gondar, Ethiopia. Dr Abuhay has an extensive background in data-driven patient flow analysis and simulation for the past 5 years. http://www.uog.edu.et/foi/personnal/dr-tesfamariam-mulugeta/

ADANE MAMUYE is an Assistant Professor at the College of Informatics at the University of Gondar, Ethiopia. He is and project coordinator of Capacity Building Mentorship Program (CBMP) and he is also Co-Director of eHealth Lab Ethiopia. He received his Ph.D from University of Camerino, Italy, in 2017. https://ehealthlab.org/our-team/adane-l-mamuye/

STEWART ROBINSON is a Professor of Management Science and Dean of School of Business and Economics at Loughborough University, UK. He completed his BSc and PhD in Management Science (Operational Research) from Lancaster University and was Fellow of the Operational Research Society. http://www.stewartrobinson.co.uk/

SERGEY KOVALCHUK is an Associate Professor and a senior researcher at the eScience Research Institute and a lecturer at the High-Performance Computing Department at ITMO University. His education includes a degree in software engineering from Orenburg State University (Orenburg, Russia) in 2006 and a PhD degree (Candidate of Technical Sciences) in the field of mathematical modelling, numerical calculations, software systems from ITMO University (Saint Petersburg, Russia) in 2008 (thesis subject "High-Performance Software System for Metocean Extreme Events Simulation"). https://en.itmo.ru/en/viewperson/1248/Sergey_Kovalchuk.htm

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.042

A HYBRID SIMULATION MODELLING FRAMEWORK FOR COMBINING SYSTEM DYNAMICS AND AGENT-BASED MODELS

Ms. Le Khanh Ngan Nguyen

Department of Management Science
University of Strathclyde
Nguyen-le-khanh-ngan@strath.ac.uk

Prof Susan Howick

Department of Management Science
University of Strathclyde
Susan.howick@strath.ac.uk

Dr. Itamar Megiddo

Department of Management Science
University of Strathclyde
Itamar.megiddo@strath.ac.uk

ABSTRACT

System dynamics (SD), discrete event simulation (DES) and agent-based model (ABM) are three different simulation modelling methods widely applied to support decision-making in complex systems across various disciplines. However, single simulation modelling approaches can face significant challenges representing the multi-dimensional nature of complex systems composed of interactive and interconnected constituents with dynamic behaviours. Combining different simulation methods offer an opportunity to overcome these challenges and to capture important characteristics and behaviours of such systems. Despite the growing interest and popularity in this approach, guidance for designing and utilizing hybrid models, especially for those combining SD and ABM, is scanty. This paper aims to review and consolidate the existing theoretical guidance/frameworks on combining these two simulation methods. Based on this literature review, we propose an initial framework for combining SD and ABM.

Keywords: System dynamics, agent-based model, hybrid simulation, theoretical frameworks.

1 INTRODUCTION

Although system dynamics (SD) and agent-based model (ABM) are different in terms of their philosophical approaches, both methods possess strong explanatory capabilities and can be combined and/or integrated (Bobashev et al, 2007; Phelan, 1999). The top-down approach of aggregated feedback of SD and the bottom-up approach of ABM may complement one another in a hybrid simulation modelling design to provide useful insights and realistic aspects of problems of complex systems. Combining SD and ABM enables problem owners to deal with different factors of system complexity, including micro, meso, and macro perspectives; strategic, tactical and operational levels; and detail and dynamic complexity (Begun et al, 2003; Morel & Ramanujam, 1999). The scope of this paper is to consolidate the existing guidance on how to combine the two simulation modelling methods and propose a framework for developing hybrid SD-ABM models.

2 OVERVIEW OF SYSTEM DYNAMICS AND AGENT-BASED MODEL

2.1 System Dynamics

From an SD perspective, the interaction among the elements within a system and their interaction with the environment generate the characteristic behaviour of that system (Pidd, 1998). SD is a simulation modelling method which represents the structure of complex systems as accumulations (stocks), rates (flows), feedback and time delays, and examines their behaviour over time (Sterman, 2000). Stocks (or "levels") are defined as aggregation or accumulations of inflows and outflows over a period of time. Feedback exists when a change in a variable in the system impacts other variables in the system and these variables then, in turn, influence the initial variable. Delays represent the time it takes to measure and report information, make decisions or update stock that causes outputs to lag behind inputs.

2.2 Agent-Based Model

ABM is a simulation method for modelling autonomous, dynamic and adaptive systems and is formed on the basis of three key concepts which are agency, dynamics, and structure (Borshchev & Filippov, 2004; Gunal, 2012). Agency means that agents are autonomous entities with specific properties, actions and possibly goals. Dynamics is the development, change, and evolvement of both agents and their environment over time. Structure is emergent as a result of agent interaction. Agents live in the environment, sense it and decide what action to employ at a certain time on the basis of the current state of the environment and their own state and defined decision rules. Agents can have explicit targets to minimize or maximize, and they can also learn and adapt based on their experiences. Such interactions result in the update of agents' internal state or decision on their next actions. The lower-level autonomy and interaction lead to the concept of dynamics at the system level. The system changes and patterns emerge as agents and their environment evolve or co-evolve over time. The core idea of ABM is that a model composed of agents that interact with one another and their environment can effectively demonstrate many (if not most) phenomena and real-world systems (Wilensky & Rand, 2015).

3 THE EXISTING THEORETICAL GUIDANCE ON COMBINING SD AND ABM

SD and ABM have already been used separately to study the same problem in some areas which has led to interesting outcomes. For example, Scholl (2001) compared SD and ABM literature on the bullwhip phenomenon, which arises in supply-chain management, and Rahmandad and Sterman (2008) studied literature using these methodologies to model "networking" problems such as innovation diffusion and AIDS dissemination through needle sharing (Rahmandad & Sterman, 2008; Scholl, 2001). These reviews indicate differences and similarities between results and explanations of the studied phenomena in the two simulation modelling methods. In addition to supply chain management and diffusion, SD and ABM methods were also compared in areas such as ecology (Norling, 2007) and biology (Wakeland et al, 2004). Applying the two methods separately to study the same problem provide fruitful insights, cross-validation, and triangulation of results (Phelan, 2004). While the early works focus on the use of one simulation modelling method to validate outputs generated by the other and triangulate outputs, a growing number of studies using hybrid SD-ABM approaches have shown the diversity in the designs of hybridization of the two methods. We conducted a review of literature on the existing theoretical guidance to design a hybrid simulation model that combines SD and ABM and summarized the results of different designs for a hybrid SD-ABM model in Table 1. Although some of the studies included in this table provide guidance on mixing SD and discrete event simulation (DES), mixing analytic and simulations modelling, or mixing methods in general, the hybrid designs they proposed can be used for mixing SD and ABM. It should be noted that we did not include Lättilä et al (2010), Onggo (2014), and

Djanatliev and German (2015) in Table 1. Although these studies provide guidance on mixing simulation methods, they did not provide a description of specific hybrid designs.

We identified and classified the existing combinations of SD and ABM into six designs. As the literature uses different sets of terminology to describe similar designs, we will not explain all terminology but only the general ideas for each design. Detailed explanations can be found in the referenced papers. When using a parallel design, SD and ABM are used to develop independent models either to address different aspects of the same problem which are better suited with one particular simulation method or to represent the same problem for direct comparison. Results of these models are ultimately combined to solve the same problem or compared to enhance confidence in output produced by each model. A sequential design includes two or more separate sub-models embedded in different simulation modelling methods in which one model is used to inform the other. One simulation is initially run, and it produces output before terminating; the second simulation starts to run, using as input the output of the first simulation. The information and/or data are passed only once from the first to the second simulation. The output of the second simulation represents the final output of the hybrid model. An interaction design comprises different sub-models developed using different simulation modelling approaches which are considered equally important and interact cyclically during run time. Interactions between sub-models occur several times in each direction. A sequential design can be considered as a special case of the interaction design when the interaction occurs once and in one direction only. Integration is an approach that combines different simulation modelling methods to create one seamless hybrid model in which it is impossible to explicitly distinguish between the SD and ABM parts and to identify where one simulation approach ends and the other begins. This design offers a coherent view of the problem which enhances continuous flows of information and feedback and captures interactive effects within a system. Although several studies concur on the definition of an integration design, only Swinerd and McNaught (2012 & 2014) describe in detail different ways to develop an integrated hybrid model. They proposed three designs which belong to the integrated class, including agents with rich internal structure, stocked agents, and parameters with emergent behaviour. This design combines different simulation modelling methods to form one unified hybrid model in which one method dominates and is enhanced by elements of another. As enrichment and integration designs share many similarities, there seems to be a continuum from enrichment to full integration in hybrid simulation modelling designs depending on the relative dominance between the adopted simulation approaches. An enrichment design uses an element of one simulation method to enhance the main method without the need to build an additional model, while integration brings together two full methods to create something new. A dynamic design allows the dynamic switching between SD and ABM in the structure of a model. Its reported application is to efficiently depict the process of an ongoing epidemic. The extent to which sub-models in a hybrid simulation model are coupled depends on its design and are increasingly coupled in the following order: parallel (genuinely independent), sequential (loosely coupled), interaction, dynamic, enrichment, and integration (inseparably coupled).

Table 1 *The existing theoretical guidance/frameworks for combining SD and ABM*

References	**Designs for a hybrid SD-ABM model**					
	Parallel	**Sequential**	**Interaction**	**Enrichment**	**Integration**	**Dynamic**
(Shanthikumar & Sargent, 1983)	Class I	Class III, IV			Class II	
(Bennett, 1985)	Comparis-on			Enrichment	Integration	
(Kim & Juhn, 1997)				Multi-Agent Dynamics where a hybrid model is constructed with the principles of SD and using array variables to represent the individual agents		

(Parunak et al, 1998)				Agents modeled using the equations of SD. An agent can be part of a bigger SD.		
(Akkermans, 2001)				Using SD to model the logic of individual agents.		
(Schieritz & GroBler, 2003)				Using SD to model the internal decision logic or cognitive structure of the agents in an ABM.		
(Borshchev & Filippov, 2004)				SD sub-models inside discretely communicating agents. Agents live in an environment whose dynamics is modeled using SD.		
(Lorenz & Jost, 2006)				Using SD structures to create entities for an ABM. An "active" environment		
(Bobashev et al, 2007)						Hybrid threshold model
(Martinez-Moyano et al, 2007)		Scenario exploration and Crisis response			Intertwined models	
(Chahal & Eldabi, 2008)			Hierarchical format	Process - Environment format	Integration format	
(Brailsford et al, 2010)					The "Holy Grail"	
(Vincenot et al, 2011)				Case 1, 2, and 3		Case 4
(Swinerd & McNaught, 2012, 2014)	Interfaced class	Sequential class			Integrated class including: Agents with rich internal structure, stocked agents, parameters with emergent behaviour	
(Chahal et al, 2013)		Cyclic interaction	Parallel interactions			
(Wallentin & Neuwirth, 2017)					"Super-agents"	Dynamically switching hybrid model
(Morgan et al, 2017)	Parallel	Sequential	Interaction	Enrichment	Integration	

4 LIMITATIONS OF EXISTING GUIDANCE FOR HYBRIDIZING SD AND ABM

There are three major limitations of the studies shown in table 1 when providing guidance on combining SD and ABM Firstly, they do not specify the processes that modellers need to take and which aspects they need to consider to reach a decision on the design of a hybrid model. Secondly, we note that such

guidance is established at a high level and it is, therefore, still quite abstract and not straightforward for problem owners to apply in solving a specific problem. Lastly, most of the existing hybrid simulation modelling studies focus on dealing with issues of particular domains such as inter-organizational network development in Akkermans (2001) or supply chain management in Schieritz and Größler (2003), rather than offering a broader but more detailed guidance specifying when, why, and how to combine SD and ABM approaches. In this paper we focus on describing "how" the two methods can interact and exchange data and information. The paper seeks to provide a detailed and practical stepwise instruction that specifies what steps modellers need to take and what they need to do in each step to build a hybrid simulation model. This is achieved through building on the existing guidance on hybrid SD-ABM modelling and reflecting on existing hybrid SD-ABM studies and the process of building a hybrid model. A guideline presented in a consistent and structured format will assist the selection of appropriate model designs in future studies, which will further facilitate and enhance the efficiency of the process of developing a hybrid model and the usefulness of the created models.

5 A HYBRID SD-ABM SIMULATION MODELLING FRAMEWORK

We propose a hybrid SD-ABM simulation modelling framework that aids modellers in the conceptual design phase of the development of a hybrid model. We expand and elaborate the existing work in this research area and focus on the processes that are essential to structure the problems into hybrid simulation models of which discussing different designs of hybridization is one part. In the proposed framework, the iterative process of developing a hybrid simulation model has been divided into 9 steps. The first three steps aim to specify the characteristics of the problem of interest on which modellers determine whether an individual SD or ABM or a hybrid SD-ABM is most suited to modelling the problem. In the fourth step, modellers determine different modules within a hybrid model, the hierarchy, and levels of abstraction for each of them. A simulation modelling method is chosen for each module in the fifth step. After defining the flows of information among modules in the sixth step, in the next step, the modeller decides on a design to combine these modules (i.e., combing different simulation modelling methods into a hybrid model). Finally, interfaces between modules and updating rules are defined.

5.1 Conceptualizing the Problem

Before developing a simulation model, it is vital to be clear about the nature of a problem under investigation and the objectives of the model. This helps to identify the scope of the model and the level of detail required and, therefore, the choice of appropriate simulation modelling methods (Roberts et al, 2012; Robinson, 2008). In addition to reviewing literature describing the problem and existing models addressing related problems, modellers should widely consult with relevant stakeholders and experts to refine the problem definition and develop clear, agreed modelling objectives. Defining the objectives of the model is an iterative process as deepening understanding of the problem gained from a modelling process may alter objectives. A more detailed understanding of the problem also guides modelling decisions. For example, building a simulation model to guide a healthcare practice or a public health policy should carefully and explicitly define the problem characteristics such as the target population, the healthcare setting, the cost of different interventions and how it can be modelled, the health outcomes of importance for that population, and the time horizon adequate to capture differences in outcome across interventions.

5.2 A Problem-Oriented Approach to Choosing Between a Single and Hybrid Simulation Modelling Method

Once the problem is conceptualized, it is important to identify whether the problem of interest can be modeled and solved using a single simulation modelling method or requires a hybrid simulation approach. Each simulation modelling method has strengths and limitations, making it better suited for specific

problems and less so for other ones (Scholl, 2001). The selection of different simulation modelling methods should, therefore, be driven by the problem characteristics. Lättilä, et al (2010) has listed some different "problematic situations" where one of the simulation modelling methods is preferred to use. This helps answer the questions on why and when it is appropriate to use hybrid simulation models. Modellers will choose a hybrid simulation modelling approach that combines the strengths of SD and ABM if one simulation paradigm has difficulty to capture the complexity of the problem on its own.

5.3 Determining Modules, Hierarchy, and Levels of Abstraction

A model can consist of several components called "modules". A module should principally be self-contained and bounded with predefined interfaces (input and output) to the external world, including other modules. In a hybrid simulation model, we find it useful to consider a module as one logical component of a hybrid model developed using one of the simulation modelling methods (Onggo, 2014). In an integrated hybrid model, the boundary between modules is not explicit because the interfaces between modules are intertwined. In this case, we still can dissect the model into smaller components, where each component will be considered as a module that can receive a set of inputs and transform them into a set of outputs. Djanatliev and German's 2015 work aligns with this idea; they raised the necessity to define independent problem areas within a specific domain scope and to model each area using one of the simulation methods (Djanatliev & German, 2015). There are several options to perform this task. For example, in dealing with the problems in healthcare we can explore the problems in different healthcare settings such as hospitals and long-term care facilities. We can also use a hierarchical breakdown to study the problems in healthcare at a national/regional level (macro), an institutional level (meso/micro) and an individual level (micro).

5.4 Selecting Simulation Modelling Methods for Each Module

This step concurs with Horizontal Paradigm Linking proposed by Djanatliev and German (2015). After identifying the modules of a problem, modellers need to justify the selection of a particular simulation modelling method used for each module and whether it is the most appropriate for the job (Brailsford et al, 2013). Specifying the modelling hierarchy in a hybrid SD-ABM model, which is the hierarchical level of an SD module relative to that of an ABM module, also aids the choice of a simulation method for each module. Figure 1 represents different types of hierarchy levels for different designs of hybrid SD-ABM simulation modelling.

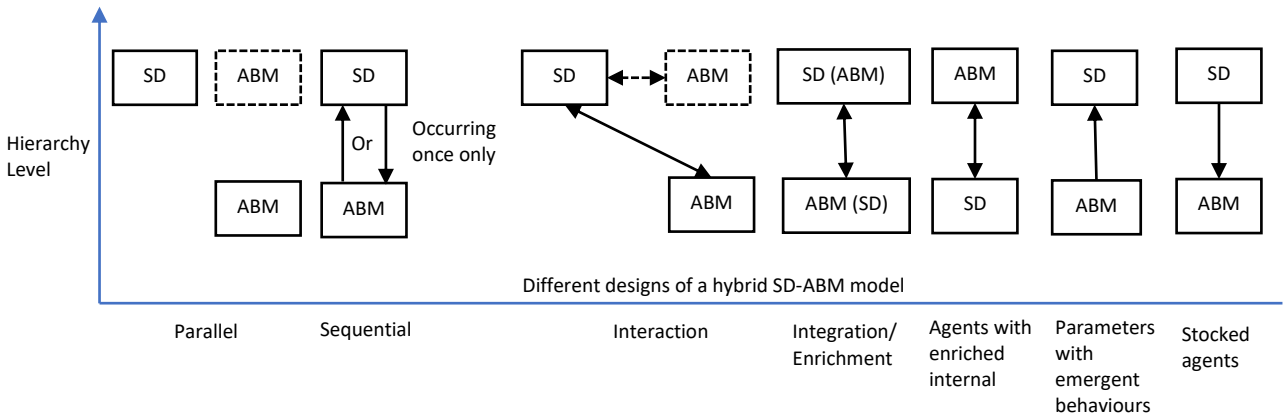


Figure 1 *The hierarchy levels and information flows for different designs of SD-ABM hybrid models*

5.5 Defining the Flow of Information between Modules

In this step, modellers decide on the paths and directions of information flow between modules in a hybrid model. Figure 1 describes the flow paths and directions of information between an SD module and an ABM module in a hybrid model by arrows. Single-headed arrows indicate that information only flows in one direction while double-headed arrows show an exchange of information in both directions between two modules. Boxes with dashed borders indicate that the hierarchical level of a module can be at any level but not occupying all levels simultaneously. In the sequential design, information is passed in one direction, either from an SD module to an ABM module or vice versa. Decisions on the flow of information between modules will support the modeller's choice of the most appropriate design for a hybrid simulation model.

5.6 Selecting a Hybrid Simulation Modelling Design

Defining the flows of information and choosing a hybrid simulation design are in essence Vertical Paradigm Linking, previously described by Djanatliev and German (2015). The flows of data from an ABM module to an SD module can be communicated in several different manners: sending the total number of agents with specific attributes in the ABM as an inflow to the SD module; the emergent property of an ABM influencing the relationship governing a stock level in the SD module; sending a population size based on an SD module stock to an ABM module within a predefined unit of time, and individual agents can be generated using distribution functions based on existing empirical data or theories to represent the necessary heterogeneity of these agents; and using the size of a stock in an SD module to affect agents' behaviours and goals as well as their environment's attributes. The hierarchy of modules and how information exchanges between them inform the selection of a design for hybridization as shown in Figure 1. We will discuss the detailed description of each hybrid SD-ABM design and examples of its application to explore the specific problems of healthcare associated infections in later work. Figure 2 shows how the design of a hybrid simulation model is chosen step by step.

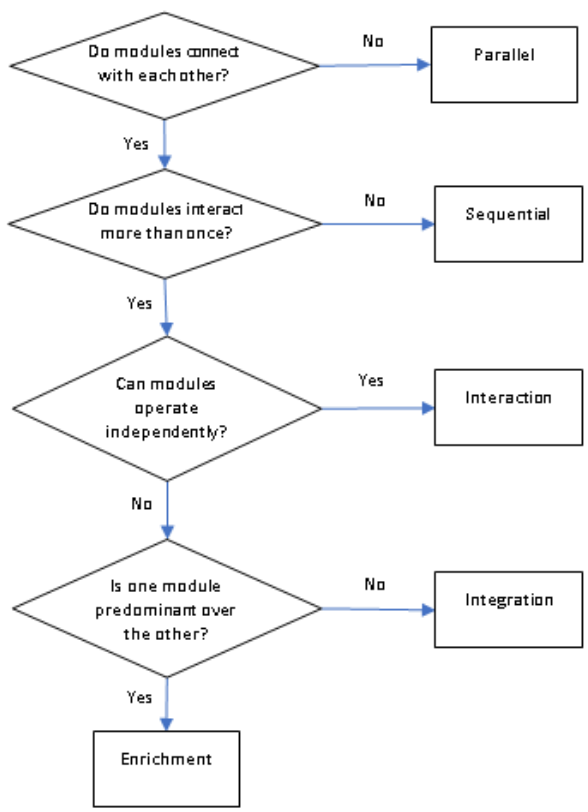


Figure 2 *Selection of a hybrid simulation modelling design*

5.7 Defining Interfaces

In the second last step, the modeller identifies clear and logical interfaces between modules. An interface that decouples the two modules defines the information passing from one module to the other, the module generating the information and the one receiving the information during the running time of the hybrid model. It is also important to determine how the output produced by a module is treated: whether the output information will become input for another module, form a part or the entire output for the hybrid model, or both. As one module in a hybrid model is represented by SD while another module is implemented in ABM, they have different levels of details. The information needs to be aggregated when moving from a lower level of detail to a higher level and disaggregated when moving in the reverse direction.

5.8 Defining Updating Rules

Updating rules will define when the information will be sent from one module to another and how new information is handled by the receiving module to maintain the logical consistency of the whole hybrid model (Onggo, 2014). Modellers also need to consider the running time of a model when defining updating rules. Although the modules in a hybrid SD-ABM model use the same time advancement method, namely fixed-time increments, they may use different time units. Additionally, modules can be run on different simulation modelling software which has its own internal time management. If SD and ABM modules use the same unit of time, updates can be easily done when the hybrid model advances its simulation time. However, this may slow down simulation run time if one runs faster than the other for same simulated time unit. One module will have to wait for the other to finish. This will be a bigger problem in large models such as models with many agents. If the modules use different units of time, updates can occur asynchronously or synchronously. Asynchronously, every time a module advances its simulation time, the module's status may alter and it will send new information to recipient modules which the interfaces define (Onggo, 2014). Synchronously, all modules in a hybrid model will pass their information to other recipient modules at predefined simulation points which can be, for example, the time step of one of the modules.

6 CONCLUSION

This proposed framework is still in its infancy but considered as a good starting point to build up a more comprehensive version as the research evolves. In future work, we will apply this framework to design a hybrid SD-ABM model to explore a problem in healthcare-associated infection prevention and control. Based on the experience from implementing the framework to build the model, we also reflect upon the framework by considering what was necessary and appropriate to facilitate the process of modelling, what was not applicable and what changes should be made to enhance the practicality of the framework.

REFERENCES

Akkermans, H. (2001). Renga: a systems approach to facilitating inter-organizational network development. *System Dynamics Review, 17*(3), 179-193. doi:10.1002/sdr.215

Begun, J. W., Zimmerman, B., & Dooley, K. (2003). Health care organizations as complex adaptive systems. *Advances Health Care Organization Theory, 253*, 253-288.

Bennett, P. G. (1985). On Linking Approaches to Decision-Aiding: Issues and Prospects. *The Journal of the Operational Research Society, 36*(8), 659-669. doi:10.2307/2582261

Bobashev, G. V., Goedecke, D. M., Feng, Y., & Epstein, J. M. (2007, December). *A Hybrid Epidemic Model: Combining The Advantages Of Agent-Based And Equation-Based Approaches.* Paper presented at the 2007 Winter Simulation Conference.

Borshchev, A., & Filippov, A. (2004). *From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools.* Paper presented at the Proceedings of the 22nd international conference of the system dynamics society, Oxford.

Brailsford, S., Desai, S., & Viana, J. (2010). *Towards the Holy Grail: combining system dynamics and discrete-event simulation in healthcare.* (Vol. 2293-2303).

Brailsford, S., Viana, J., Rossiter, S., Channon, A. A., & Lotery, A. J. (2013, 8-11 Dec). *Hybrid simulation for health and social care: The way forward, or more trouble than it's worth?* Paper presented at the Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World.

Chahal, K., & Eldabi, T. (2008, 7-10 Dec). *Applicability of hybrid simulation to different modes of governance in UK healthcare.* Paper presented at the Winter Simulation Conference.

Chahal, K., Eldabi, T., & Young, T. (2013). A conceptual framework for hybrid system dynamics and discrete event simulation for healthcare. *Journal of Enterprise Information Management, 26*(1/2), 50-74.

Djanatliev, A., & German, R. (2015, 6-9 Dec). *Towards a guide to domain-specific hybrid simulation.* Paper presented at the 2015 Winter Simulation Conference (WSC).

Gunal, M. M. (2012). A guide for building hospital simulation models. *Health Systems, 1*(1), 17-25. doi:10.1057/hs.2012.8

Kim, D.-H., & Juhn, J.-H. (1997). *System Dynamics as a Modeling Platform for Multi-Agent Systems.* Paper presented at the The 15th International Conference of the System Dynamics Society Istanbul, Turkey.

Lorenz, T., & Jost, A. (2006). *Towards an orientation framework in multi-paradigm modeling.* Paper presented at the Proceedings of the 24th International Conference of the System Dynamics society.

Martinez-Moyano, I., Sallach, D., Bragen, M., & Thimmapuram, P. R. (2007). Design for a Multilayer Model of Financial Stability: Exploring the Integration of System Dynamics and Agent-based Models.

Morel, B., & Ramanujam, R. (1999). Through the Looking Glass of Complexity: The Dynamics of Organizations as Adaptive and Evolving Systems. *Organization Science, 10*(3), 278-293. doi:10.1287/orsc.10.3.278

Morgan, J. S., Howick, S., & Belton, V. (2017). A toolkit of designs for mixing Discrete Event Simulation and System Dynamics. *European Journal of Operational Research, 257*(3), 907-918. doi:https://doi.org/10.1016/j.ejor.2016.08.016

Norling, E. (2007). *Contrasting a System Dynamics Model and an Agent-Based Model of Food Web Evolution.* Paper presented at the Multi-Agent-Based Simulation VII, Berlin, Heidelberg.

Onggo, B. S. (2014, 7-10 Dec). *Elements of a hybrid simulation model: A case study of the blood supply chain in low- and middle-income countries.* Paper presented at the Proceedings of the Winter Simulation Conference

Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). *Agent-based modeling vs. equation-based modeling: A case study and users' guide.* Paper presented at the International Workshop on Multi-Agent Systems and Agent-Based Simulation.

Phelan, S. E. (1999). A Note on the Correspondence Between Complexity and Systems Theory. *Systemic Practice and Action Research, 12*(3), 237-246. doi:10.1023/A:1022495500485

Phelan, S. E. (2004). *Using integrated top-down and bottom-up dynamic modeling for triangulation and interdisciplinary theory integration.* Paper presented at the The 22nd international conference of the System Dynamics Society, Oxford.

Pidd, M. (1998). *Computer simulation in management science* (Vol. 4): Wiley Chichester.

Rahmandad, H., & Sterman, J. (2008). Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. *Management Science, 54*(5), 998-1014. doi:10.1287/mnsc.1070.0787

Roberts, M., Russell, L. B., Paltiel, A. D., Chambers, M., McEwan, P., & Krahn, M. (2012). Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force–2. *Medical Decision Making, 32*(5), 678-689.

Robinson, S. (2008). Conceptual modelling for simulation Part I: definition and requirements. *Journal of the Operational Research Society, 59*(3), 278-290. doi:10.1057/palgrave.jors.2602368

Schieritz, N., & GroBler, A. (2003, 6-9 Jan). *Emergent structures in supply chains - a study integrating agent-based and system dynamics modeling.* Paper presented at the 36th Annual Hawaii International Conference on System Sciences.

Scholl, H. J. (2001, July 23–27). *Looking across the fence: Comparing findings from SD modeling efforts with those of other modeling techniques.* Paper presented at the In Proceedings of the 19th international conference of the system dynamics society, Atlanta, USA.

Shanthikumar, J. G., & Sargent, R. G. (1983). A Unifying View of Hybrid Simulation/Analytic Models and Modeling. *Operations Research, 31*(6), 1030-1052. doi:10.1287/opre.31.6.1030

Sterman, J. (2000). *Business Dynamics, System Thinking and Modeling for a Complex World.* (Vol. 19).

Swinerd, C., & McNaught, K. R. (2012). Design classes for hybrid simulations involving agent-based and system dynamics models. *Simulation Modelling Practice and Theory, 25*, 118-133.

Swinerd, C., & McNaught, K. R. (2014). Simulating the diffusion of technological innovation with an integrated hybrid agent-based system dynamics model. *Journal of Simulation, 8*(3), 231-240.

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., & Mazzoleni, S. (2011). Theoretical considerations on the combined use of system dynamics and individual-based modeling in ecology. *Ecological Modelling, 222*(1), 210-218.

Wakeland, W., Gallaher, E., Macovsky, L., & Aktipis, C. (2004). *A comparison of system dynamics and agent-based simulation applied to the study of cellular receptor dynamics*.

Wallentin, G., & Neuwirth, C. (2017). Dynamic hybrid modelling: Switching between AB and SD designs of a predator-prey model. *Ecological Modelling, 345*, 165-175. doi:https://doi.org/10.1016/j.ecolmodel.2016.11.007

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*: The MIT Press.

AUTHOR BIOGRAPHIES

LE KHANH NGAN NGUYEN is a PhD student at the Department of Management Science at the University of Strathclyde. She received both a Bachelor of Pharmacy and Master of Public Health from the University of Sydney. Her research topic is hybrid health systems simulation modelling in controlling healthcare-associated infections.

SUSAN HOWICK is a Professor of Management Science. She gained a BSc in Mathematics from Dundee University and a MSc in Operational Research and a PhD in Management Science from Strathclyde. Her main research interest lies in taking a systems perspective when using models to support decision-makers. She is also interested in exploring approaches to integrating models that promote client value from the modelling process. Susan is currently a Vice-President of the Operational Research Society and was President of the Policy Council for the UK Chapter of the System Dynamics Society in 2007-08.

ITAMAR MEGIDDO is a Chancellor's Fellow, lecturer (assistant professor) at the Department of Management Science at the University of Strathclyde. He is also a visiting fellow at the Centre for Disease Dynamics, Economics & Policy, Washington, DC, and he is a member of the International Decision Support Initiative. His research focus is on improving evidence-based decision making and resource allocation in healthcare policy. His main body of work integrates disease (including infection and non-communicable), health system, and economic models to evaluate impact and design policy and strategy.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.043

DEVELOPING A HYBRID SIMULATION MODEL USING BOTH PARSIMONIOUS AND HIGHLY DESCRIPTIVE APPROACHES: REFLECTIONS FROM THE TRANSPORT INDUSTRY

William Jones
Kathy Kotiadis
Jesse O'Hanley

Kent Business School
University of Kent
Canterbury, Kent, CT2 7FS, UK

ABSTRACT

We put forward some initial thoughts about using both parsimonious and highly descriptive approaches to engage stakeholders during the development of a hybrid simulation study in the transport industry. The hybridisation we discuss involved combining discrete-event and agent-based simulation. We discuss how both parsimonious and highly descriptive modelling approaches, which are seemingly incompatible, were used in the development of a hybrid model to help facilitate stakeholder engagement. In our experience stakeholders with limited understanding of the system being modelled engaged with more ease when presented with highly descriptive approaches. When working with stakeholders with a better understanding, parsimonious approaches can be beneficial. We also discuss potential techniques for managing the complexity of large simulation projects by adapting ideas from software development to help modellers work with stakeholders.

1 INTRODUCTION

Developing a model of a complex system requires a significant investment of time, expertise and expense (Robinson 2004). For an organisation undertaking such an investment, often new expertise will need to be introduced into the organisation to develop the desired model. When introduced, the modeller(s) will inevitably need to engage with stakeholders across the organisation to define the system to be modelled, identify its boundaries, decompose the system into various sub-systems, processes and activities, and map their interconnections (Tako and Kotiadis 2015). In a complex organisation, these steps will inevitably involve participation of many stakeholders. This is not especially problematic when developing discrete-event simulation (DES) models given the emergence of facilitation literature such as PartiSim (Tako and Kotiadis 2015) or system dynamics (SD) given the extensive research in group model building (GMB) (Rouwette and Vennix 2011) but more ambiguous when it comes to hybrid simulation modelling given the limited literature on it.

In this study, we discuss using both parsimonious (Vandekerckhove et al. 2015) and highly descriptive (Edmonds and Moss 2005) approaches to engage stakeholders during the development of a hybrid DES and agent-based simulation (ABS) model based on our experience in the transport industry. Our primary contribution is to add our experience to the very limited pool of papers discussing the development of hybrid simulation models with stakeholders.

This study reflects on work conducted with Eurostar International Limited (EIL). EIL is the only high-speed railway company operating international train services between London and continental Europe via the Channel Tunnel. Its core destinations including Paris, Brussels, Lille and Amsterdam. Further, it operates services to Disneyland Paris and runs seasonal trains to the south of France and the French Alps.

We write the main body of this paper reflectively with a deliberately abstract style such that the reader can imagine how the ideas being discussed could apply to applications they have interest in, rather than being distracted by the specific case we reflect on. We include some footnotes to provide detail about the specific case where the extra context might be helpful.

The paper proceeds as follows. First, we provide some background and a review of the relevant academic literature (Section 2). Next, we summarise our experience of working with a large number of stakeholders and facilitating their involvement in the modelling lifecycle and put forward some techniques we used to aid the development of the hybrid model (Section 3). We go on to provide a discussion on how both parsimonious modelling (Vandekerckhove et al. 2015) and highly descriptive modelling approaches (Edmonds and Moss 2005) can be used within the same hybrid simulation model to increase engagement with and confidence in the modelling process (Section 4). Finally, we give some concluding remarks (Section 5).

2 BACKGROUND

The more complex a model becomes the more difficult it is usually to control and, in turn, gain insights about the fundamental workings of the system being modelled. A parsimonious modelling approach attempts to overcome this difficulty by designing the simplest possible model to achieve the required level of explanatory or predictive power (Vandekerckhove et al. 2015). If a parsimoniously designed model cannot produce outputs reflective of reality, it should first be asked whether the model itself is reflective of reality and or whether the modellers understanding of the system is correct before adding more complexity to the model (Edmonds 2000). This approach is often summarised by the adage "Keep It Simple Stupid" or KISS. Broadly the KISS approach recommends that a modeller should develop the simplest model possible and progressively add more complexity only when the simple model is shown to be inadequate

In contrast to KISS, Edmonds and Moss (2005), propose the "Keep it Descriptive Stupid" (KIDS) paradigm. Using a KIDS approach, a modeller should start by developing a model that is highly detailed and the most accurate reflection of the real system as possible, only simplifying this description when there is evidence and sufficient understanding to do so. When modelling a complex system, this inevitably means the initial model must be large and intricate. They argue that if simplifying from the start, some feature that is left may later turn out to be important.

Regardless of which modelling approach is adopted, the benefits of involving stakeholders in the simulation development lifecycle, in particular DES models, are well documented (Eldabi, Paul, and Young 2007; Fone et al. 2003; Jun, Jacobson, and Swisher 1999; Gunal and Pidd 2005; Lowery et al. 1994; Wilson 1981; Kotiadis et al. 2014; Robinson et al. 2014) and failing to do so can often result in findings not being accepted or acted upon (Brailsford and Vissers 2011; Fone et al. 2003; Young et al. 2009). Tako and Kotiadis (2015) present guidance for involving stakeholders throughout the lifecycle of developing a DES model. Their approach advocates engaging stakeholder through structured workshops to inform model design.

Surprisingly, there is little, if any, literature formally exploring the benefits of involving stakeholders in the lifecycle of ABM studies. However, it has been noted that ABM allows and facilitates a more direct correspondence between what can be observed by the stakeholders and what is modelled (Edmonds and Moss 2005). When applied to ABM, the descriptive nature of the KIDS approach aligns naturally with a participative, stakeholder-driven approach to model construction and validation (Barreteau, Bousquet, and Attonaty 2011). Highly descriptive ABMs have the advantage that their straight-forward correspondence with the real system provides a form of face validation (Edmonds and Moss 2005).

To harness the various benefits of different modelling approaches, hybrid simulation models have gained in popularity in recent years (Mustafee et al. 2017, Brailsford 2015, Brailsford et al. 2018). These are conceptual models, implemented in specialised software, that combine more than one simulation paradigm. Siebers et al. (2010) notes that ABM is well suited "*when the goal is modelling the behaviours of individuals in a diverse population*". DES, on the other hand, is known to be able to accurately represent a system

involving stochastic events and processes governed by known input parameters. Consequently, models combining ABM with DES are particularly useful for representing complex organisations, often in service industry settings (Brailsford 2014), where several, seemingly autonomous entities operate according to their own set of events and processes and where their interactions cause complex system behaviours to emerge.

Summarising the causal factors of low stakeholder engagement, Jahangirian et al. (2015) identify "*difficulty with understanding and working with simulation tools, techniques and models*" as a key issue. When building a hybrid simulation model of a large organisation or other complex systems, inevitably not all stakeholders can be involved during all stages of the development lifecycle. It is likely that at each stage, the modeller will need to engage different stakeholders or re-engage stakeholders who have been out of touch for a period of time. As such, there is a need to design models and present model design choice to stakeholders in a way that easily enables them to engage and can help bridge the "*communication gap*" (Jahangirian et al. 2015) between modeller and stakeholders.

With this in mind, there is a clear need to understand how best to present complicated ideas and models to stakeholders in order to encourage involvement and gain trust. In the following section, we reflect on our experience of developing a hybrid simulation model for a large public transport organisation. We consider the stakeholders' views on the different modelling methodologies that were employed and discuss techniques used to aid working with stakeholders.

3 MANAGING THE COMPLEXITY OF ENGAGING STAKEHOLDER IN COMPLEX MODEL'S

We present here the experience of a modeller engaging with stakeholders in the development of a hybrid simulation model. We recount the level of stakeholders' engagement with each modelling approach and the stakeholders' beliefs about what could be achieved from modelling. Domain-driven design, encapsulation and test driven development are all techniques used by software developers when collaborating on complex software projects. We take these methods and discuss how they can be re-purposed as tools for managing the implementation of hybrid simulation modelling project that require engagement with many stakeholders and stakeholder buy-in.

The hybrid model we reflect on used an ABM and DES approach to represent the organisation's operations. The ABM structure represents the many sub-systems of the organisation as autonomous agents[1]. Within most of these sub-system agents, a DES model is defined that captures the processes and events of that sub-system[2]. Other agents provide data or make policy decisions regarding the operation of the system (see Fig. 1). In parallel to the simulation, other analytic techniques common to operations research were also introduced to the organisation. The organisation had no prior experience of these types of simulation or operations research techniques.

3.1 Stakeholder Perspectives

During the development of the hybrid simulation model the modeller was required to engage, broadly speaking, with two distinct groups of stakeholders. The first group of stakeholders were experts in specific areas (i.e., one sub-system of the organisation). The modeller needed to engage with them to develop simulations that represented the sub-system of interest to a level of detail acceptable to the stakeholders. This had to be repeated, in turn, for each sub-system. Further, when dealing with these stakeholders, the modeller needed to satisfy them that the system-level model containing their sub-system of interest was also accurately represented. These stakeholders did not necessarily have an expert understanding of the internal working of other components of the system. Additionally, the modeller was required to liaise with

[1]EIL is made up of many geographically separate sub-systems such as stations, depots, trains, the control room, etc which operate autonomously. All of these must operate together seamlessly in order for the service to be successfully delivered. Each one of these individual sub-systems is captured as an individual agent within the model.

[2]For example, when a train arrives in a station in the model this triggers the start of a DES model capturing the activities involved in the turn around of a train as can be observed in EIL's stations, e.g., unloading of passengers, followed by cleaning and restocking, followed by security sweep, followed by loading of new passengers.

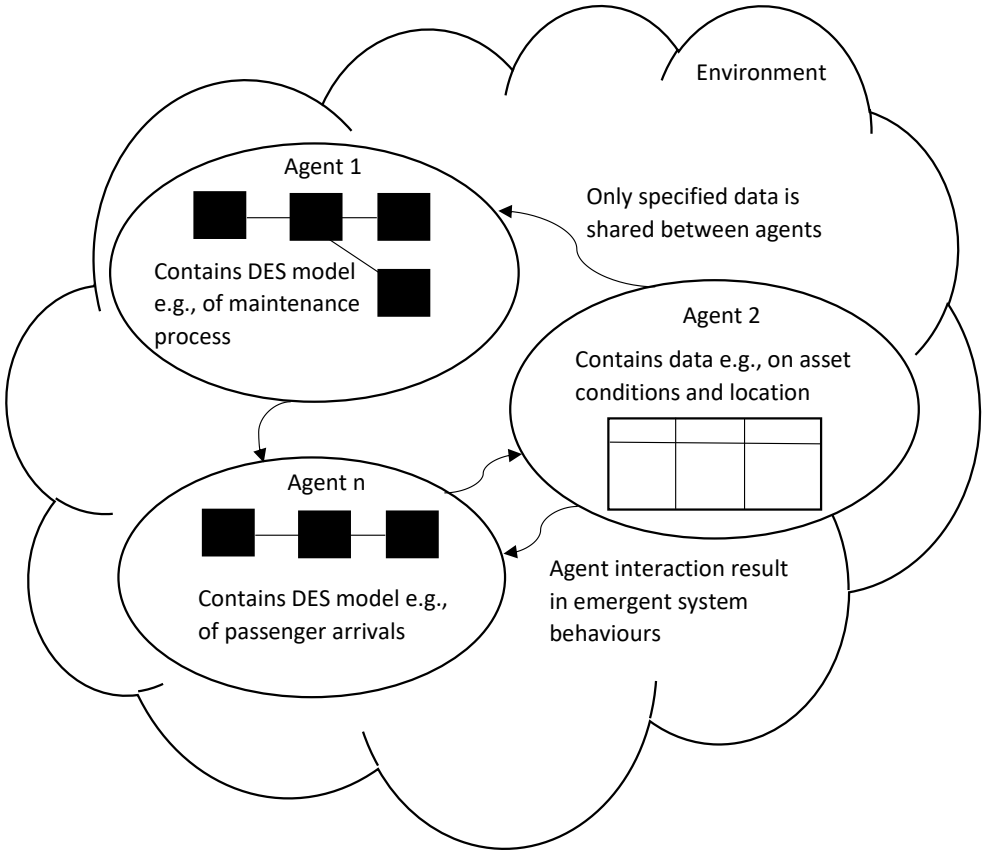


Figure 1: A highly simplified illustration of the developed hybrid simulation model structure.

a second group of stakeholders. Their interest differed from the first group. They were interested in the combined operation and behaviours of all the interacting sub-systems, more than any specific sub-system. This situation is likely common in that these two groups of stakeholders will be encountered in many types of organisations. The first group of stakeholders are usually those working in or managing a specific area or process, while the second group mostly include the organisation's senior management. These groups of stakeholders have different involvement with the system and, hence, different perspectives on its operation. However, based on personal experience, both share a similar understanding of how modelling is capable of capturing the main elements of a system and how it can benefit their own set of interests. A more detailed comparison of stakeholder's views on the modelling methods that were used is shown in Tab. 1. We acknowledge that the modeller's views and explanation may have influenced those of the stakeholders[3].

One key insight gleaned during initial scoping work was that stakeholders' ability to engage with a particular level of the model (system versus sub-system level) was dependant on their expertise in the system level being modelled. Of the two modelling approaches, a highly descriptive approach was found to be easier for stakeholders to engage with when they are non-experts due to the clear correspondence between the observable system and the model (Edmonds and Moss 2005). When stakeholders had expertise in a certain sub-system, however, they were much more comfortable to see the sub-system modelled using

[3]EIL is a geographically distributed organisation. Each sub-system e.g., the Paris terminal or London maintenance depot, will be managed and operate by a group of stakeholders who, due to their vast experience operating and running that sub-system day to day, have an expert knowledge of how it works. Those particular stakeholders, however, do not necessarily have an expert knowledge of how other sub-systems operate or the various sub-systems interact. Other groups of stakeholders (i.e., senior leadership) within EIL to try to look at the organisation holistically and ensure all sub-systems operate together seamlessly such that the service can run smoothly.

Table 1: Comparison of stakeholders views of modelling methods.

Comparison Measure	*Analytical Modelling*	*Discrete Event Modelling*	*Agent Based Modelling*
Ontology: To do with the stakeholders assumptions about how the system being modelled is made up and what the stakeholders assumes can exist. (i.e., how a valid representation of the system can be achieve).	A mathematical function describing the relationship between an input and an output.	The system can be defined by a series of events / process that occur in a know / definable order. An abstraction of the systems events / processes bounded by known parameters.	The system consists of several independent components or sub systems, however, their operations may impact on each other. A descriptive representative of each component of the real system.
Epistemology: To do with the stakeholders beliefs about how the model could be used to their benefit. Is the method concerned with finding out about objective facts and data or ideas and phenomena that have no external reality; i.e., phenomena that can be interpreted.	For a given set of system inputs the expected system outputs can be generated, however, they may be limited in their utility, and accuracy as the results are likely to provide theoretical bounds, that may be significantly different from what is achievable in reality.	Experimentation can recreate the events and processes of the real system and generate insight into how to improve them. The method tests a clearly defined set of parameters and the data generated is accepted as accurate.	Experimentation can recreate phenomena seen to occur in the real system, providing a method of investigating these and generating understanding, hence, improving decision making. Emergent system characteristics depend on the agents behaviours, hence, if each agent is modelled appropriately the system characteristics should accurately reflect the real world.
Axiology: What is the stakeholders understanding of the purpose or use of the models? Is the intention to explain or predict the real system, or to understand it?	Can be used to produce statistical information concerned with overall system performance. The data generated provides insight from which further theories can be extracted.	The data generated should be comparable to data generated by a physical implementation. Parameters in the model can be modified, as they could be in the real system and, hence, the impact can be observed.	A model that accurately describes the system can produce accurate information. The results generated aim to identify phenomena and provide understanding of their origins. A validated model could predict the impact of changing system inputs.

a more abstract or parsimonious modelling approach when these were developed in a stakeholder-driven manner.

Based on this, it was ultimately decided to build a hybrid ABM and DES simulation model. The advantage of the hybrid model was that it combined within the same model both a highly descriptive modelling approach (Edmonds and Moss 2005), mainly through ABM, and a parsimonious modelling approach (Vandekerckhove et al. 2015), mainly through DES. Agents within the ABM were defined for each sub-systems of the organisation. Various DES models were then built into the agents by working closely with key stakeholders who had expertise in a given sub-system. In many cases, this was done through a series of structured workshops as recommended by Tako and Kotiadis (2015). Typically, DES sub-models were developed parsimoniously by focusing on the key events and processes of the sub-system as defined by the stakeholders. These stakeholders had expert knowledge of these systems and understood how events and processes could be abstracted to relatively simple models[4]. Conversely, the ABM structure containing these agents was highly descriptive of the organisational structure. Note that due to the size of the organisation, it was simply not possible to involve all stakeholders in all phases of model development. However, due to the highly descriptive ABM structure, regardless of one's specific area of expertise, all stakeholders could see correspondence between the model and the organisation and so were satisfied that it was a reflective and valid model[5]. On several occasions, stakeholders enquired about a sub-system outside their specific expertise. The modeller was happy to show this to them and explain that that part of the model had been developed with experts in that sub-system. Stakeholders were mostly happy to accept this. If they did make any comments, these were raised with the sub-system experts and, if necessary, appropriate changes made.

3.2 Techniques for Managing Complexity

Throughout the model development process we used techniques from software development, adapting them slightly, to help manage the complexity of the model building process. Three key techniques used were domain-driven design, encapsulation and test driven development. We discuss each of these and reflect on their benefits to the model development process as follows.

Domain-driven design (Evans 2004) is common practice in software engineering and aims to design software in such a way that it is clear what its purpose is and help manage complexity for developers. This is achieved by focusing software development projects on the core domain (defined sphere of knowledge) and domain logic and by basing software designs on a conceptual model of the true domain that has been devised by technical and domain experts working in collaboration address specific domain problems. Here, software engineers are collaborating with stakeholders to develop the best product (e.g., collaborating with accountants to develop accountancy software). This is similar to modellers collaborating with stakeholders to develop a simulation, however, the significant difference with a simulation study is the need for validation by stakeholders and their acknowledgement that the underpinning conceptual model and subsequent simulation model implementation provide an accurate reflection of reality to ensure they are willing to accept and act on experimentation findings. This need for stakeholder validation of simulations underpinning a conceptual model does not exist in other forms of software engineering.

[4]For example, due to the expertise in the events and processes involved, the team running EIL's London maintenance depot could understand how the depots operation could be abstracted to a simple DES model. In the model a train would arrive with maintenance requirements and leave when they were fixed. The time it spent in the depot would depend on the type of maintenance required, the available resources in the depot (i.e., engineers) and other maintenance demands at that time (i.e., other trains in the depot). The depot had a maximum capacity for the number of trains that could be accommodated, if that was reached no more trains could enter. They would have to either go to another depot, or if that was not possible, queue until capacity became available.

[5]Due to the highly descriptive nature all EIL stakeholders could see how the model directly mapped to parts of the organisation. They could see that each sub-system was contained within the model.

Closely related is the idea of encapsulation (Horstmann and Cornell 2002). Encapsulation, a fundamental concept in object-oriented programming, proposes combining data and methods that impact / use that data within a single distinct unit (e.g., a Java class). This concept separates the internal workings of a defined object from the rest of the programme. Only defined inputs or outputs are visible outside of the unit.

A number of ideas can be borrowed from domain-driven design and encapsulation concepts in the field of software engineering to help simulation modellers better engage with stakeholders. The structure of the hybrid simulation approach used in this study was designed to be easily understandable by stakeholders. The domain-driven design approach to developing software that makes clear what its purpose is fits well with the KIDS approach proposed by Edmonds and Moss (2005). As Edmonds and Moss (2005) note, this highly descriptive modelling approach enables a "*direct correspondence between what is observed and what is modelled*", which helps with face validation of the model. The ABM structure of the hybrid simulation model developed represents, as accurately as possible, the internal structure of the transport industry organisation. Further, stakeholders could relate to the different components of the system being encapsulated in the hybrid model that was developed, agents often performed complex operations (e.g., by running DES sub-models), but only certain, explicitly defined information was shared between agents. This reflected the accepted reality of the organisation. Sub-systems operate largely autonomously but share relevant information (e.g., information is shared with other parts of the organisation regarding when maintenance work on a particular asset is due to be completed, but specific details of all the maintenance process are not shared, as they were not relevant to other parts of the business). In several instances, the model could have been simplified by breaking from the domain-driven design and encapsulation principles. However, consistently following these principles to maintain a descriptive similarity between the model structure and the real system and encapsulating data in a manner reflective of the organisation's operation facilitated stakeholders' understanding and aided their engagement in the modelling process.

Domain-driven design and encapsulation have a further advantage, which links to a third concept from software engineering, test-driven development. Model development is a highly iterative process (Willemain 1995, Balci 1994, Robinson 2013). When developing a complex model of a large organisation, inevitably several versions of the model will be required and it is highly likely that the underpinning conceptual model will evolve as the modeller continually engages with more, newly introduced stakeholders. Test-driven development (Astels 2003, Beck 2003) is a process for developing software by proposing very specific test cases then adapting the software so that it can 'pass' these tests (i.e., fulfil specified criteria). This is typically done in relatively short cycles and does not allow any additions to the software that are not proven to meet requirements. This process is used by software developers, who both pose and complete the tests, to ensure changes to the software work as intended. Simulation development of a large organisation can similarly develop in a test-driven way. When working with stakeholders or introducing new stakeholders during the development of a model, inevitably new requirements for the model or scenarios to simulate will emerge. During the development of the hybrid simulation model we developed, when this situation occurred, collaboratively the modeller and the stakeholders posed a 'test', (i.e., a scenario observed in the organisation's operation to be replicated in simulation). The model was then adapted to replicate this. This test-driven development is likely easier to do if the model has been developed considering domain-driven design and encapsulation. With these concepts appropriately implemented, changes can be made within a specific part of the model to satisfy new requirements of stakeholders without having to make a fundamental change to the conceptual model. This approach acknowledges that initial engagement with stakeholders by the modeller will fail to capture all the relevant information, something that is inevitable in a large organisation. Designing the model code in this manner enables iterative development as the conceptual model evolves.

4 DISCUSSION

Edmonds and Moss (2005), propose the KIDS approach to model development as a counter to the widely used KISS approach. It is generally believed that these represent incompatible perspectives on how models

should be developed. On the contrary, we incorporated both of these approaches during the development of a single, hybrid simulation model, switching between the two to encourage engagement in the modelling process and ensure stakeholders are happy to accept the final model as valid.

In our case study, due to the size of the organisation and scope of the project, it was impossible to engage all relevant stakeholders throughout the development lifecycle of the hybrid simulation model. Further, it was inevitable that the final model was going to be complicated and difficult for non-experts to engage with. With this in mind, our model was structured to maximise stakeholder engagement. The model design considered stakeholders ontological, epistemological, and axiological perspectives (see Tab. 1). Stakeholders could clearly identify the mostly autonomous sub-systems of the large organisation. However, few stakeholders within the organisation had a clear understanding of how the multitude of interactions of the many component sub-systems or why these exist within the organisation's overall operation. This reflects the ontological perspective of agent-based modelling (Macal and North 2008) that was explained to and accepted by the stakeholders (see Tab. 1) and, hence, why a highly descriptive agent-based model structure was used. Operations of the various sub-systems were contained within the agents of the model structure and captured using a DES approach. Here, DES models were developed collaboratively with stakeholder through structured workshops following the approach of Tako and Kotiadis (2015). The DES approach captured these sub-systems as a series of time-dynamical events and processes. Stakeholders involved in the development of these DES models were experts in the individual sub-systems. They were happy to accept this abstract ontological view of the system due to their expert knowledge of the processes and events involved.

Of course, here we are reflecting on just one example of a hybrid simulation study. The hybrid highly descriptive and parsimonious approach used supported by the techniques from software development discussed (domain-driven design, encapsulation, test-driven development), help successfully deliver a modelling study the organisation was satisfied with and able to realise significant benefits from. Different groups of stakeholders from other industries will have unique problems, different worldviews and other preconceptions. The approach we have discussed may not be suitable for their problems. If that is the case, we hope at least this paper will provide other simulation practitioners with ideas they can adapt.

Crooks, Castle, and Batty (2008) note that models should be based on theory and that the traditional role of a model is to represent theory into a form whereby it can be tested and refined. In effect, a computer simulation model provides a laboratory for virtual experimentation. This is typically encapsulated in a parsimonious KISS modelling (Vandekerckhove et al. 2015) approach. However, this traditional scientific method is not always followed, particularly with agent-based models which are often used to develop theory. This is in line with the KIDS approach Edmonds and Moss (2005) propose. It should be acknowledged that any model attempting to be highly descriptive is inevitably forced to make simplifications, many of which will inevitably be hidden within model design assumptions and the software implementation (Crooks, Castle, and Batty 2008).

5 CONCLUSION

When developing complex simulation models of large organisations there is a need to design and present the models to stakeholders in a way that encourages their engagement. We present here an example of successfully developing a hybrid simulation model for a large transport industry organisation that helped to facilitate stakeholder engagement. We found that stakeholders were able to engage with models developed parsimoniously when they were experts in the sub-system being modelled. When stakeholders had less expertise in the system, a highly descriptive modelling approach was easier for them to engage with and found to promote validity of the model. In the hybrid simulation developed, both the parsimonious and highly descriptive approaches where used to better reflect the variability of stakeholders' expertise. We also explain how methods from software engineering (domain driven design, encapsulation, and test-driven development) were used to support working with stakeholders and aid the development of a highly complex hybrid simulation model.

REFERENCES

Astels, D. 2003. *Test driven development: A practical guide*. Prentice Hall Professional Technical Reference.

Balci, O. 1994. "Validation, verification, and testing techniques throughout the life cycle of a simulation study". *Annals of Operations Research* 53 (1): 121–173.

Barreteau, O., F. Bousquet, and J.-M. Attonaty. 2011. "Role-Playing Games for Opening the Black Box of Multi-Agent Systems : Method and Lessons of Its Application to Senegal ...". *Journal of Artificial Societies and Social Simulation* 4 (2).

Beck, K. 2003. *Test-driven development: by example*. Addison-Wesley Professional.

Brailsford, S. 2014. "Modeling Human Behaviour - An (ID)entity Crisis?". In *Proceedings of the Winter Simulation Conference, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller*, 1539–1548. Piscataway, New Jersey: IEEE.

Brailsford, S. 2015. "Hybrid simulation in healthcare: New concepts and new tools". In *Proceedings of the Winter Simulation Conference, edited by L. Yilmaz, W K V. Chan, I Moon, T. M K. Roeder, C. Macal, and M D. Rossetti*, 1645–1653. Piscataway, New Jersey: IEEE.

Brailsford, S., and J. Vissers. 2011. "OR in healthcare: A European perspective". *European Journal of Operational Research* 212 (2): 223–234.

Brailsford, S. C., T. Eldabi, M. Kunc, N. Mustafee, and A. F. Osorio. 2018. "Hybrid simulation modelling in operational research: A state-of-the-art review". *European Journal of Operational Research* 278 (3): 721–737.

Crooks, A., C. Castle, and M. Batty. 2008, nov. "Key challenges in agent-based modelling for geo-spatial simulation". *Computers, Environment and Urban Systems* 32 (6): 417–430.

Edmonds, B. 2000. "Complexity and scientific modelling". *Foundations of Science* 5 (3): 379–390.

Edmonds, B., and S. Moss. 2005. *Multi-Agent and Multi-Agent-Based Simulation*, Volume 3415 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Eldabi, T., R. J. Paul, and T. Young. 2007. "Simulation modelling in healthcare: reviewing legacies and investigating futures". *Journal of the Operational Research Society* 58 (2): 262–270.

Evans, E. 2004. *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley Professional.

Fone, D., S. Hollinghurst, M. Temple, A. Round, N. Lester, A. Weightman, K. Roberts, E. Coyle, G. Bevan, and S. Palmer. 2003. "Systematic review of the use and value of computer simulation modelling in population health and health care delivery". *Journal of Public Health* 25 (4): 325–335.

Gunal, M., and M. Pidd. 2005. "Simulation Modelling for Performance Measurement in Healthcare". In *Proceedings of the Winter Simulation Conference, edited by M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines*, 2663–2668. Piscataway, New Jersey: IEEE.

Horstmann, C. S., and G. Cornell. 2002. *Core Java 2: Volume I, Fundamentals*. Pearson Education.

Jahangirian, M., S. Borsci, S. G. S. Shah, and S. J. E. Taylor. 2015. "Causal factors of low stakeholder engagement: a survey of expert opinions in the context of healthcare simulation projects". *SIMULATION* 91 (6): 511–526.

Jun, J. B., S. H. Jacobson, and J. R. Swisher. 1999. "Application of discrete-event simulation in health care clinics: A survey". *Journal of the Operational Research Society* 50 (2): 109–123.

Kotiadis, K., A. A. Tako, and C. Vasilakis. 2014. "A participative and facilitative conceptual modelling framework for discrete event simulation studies in healthcare". *Journal of the Operational Research Society* 65 (2): 197–213.

Lowery, J., B. Hakes, W. Lilegdon, L. Keller, K. Mabrouk, and F. McGuire. 1994. "Barriers to implementing simulation in health care". In *Proceedings of the Winter Simulation Conference, edited by J.D. Tew, S. Manivannan, D.A. Sadowski, and A.F. Seila*, 868–875. Piscataway, New Jersey: IEEE.

Macal, C. M., and M. J. North. 2008, dec. "Agent-based modeling and simulation: ABMS examples". In *2008 Winter Simulation Conference*, 101–112: IEEE.

Mustafee, N., S. Brailsford, A. Djanatliev, T. Eldabi, M. Kunc, and A. Tolk. 2017. "Purpose and Benefits of Hybrid Simulation: Contributing to the Convergence of Its Definition". In *Proceedings of Winter Simulation Conference, edited by W.K.V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page*, 1–15. Piscataway, New Jersey: IEEE.

Robinson, S. 2004. *Simulation: The practice of model development and use*. Chichester: John Wiley & Sons.

Robinson, S. 2013. "Conceptual modeling for simulation". In *Proceedings of the Winter Simulation Conference, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl*, 377–388. Piscataway, New Jersey: IEEE.

Robinson, S., C. Worthington, N. Burgess, and Z. J. Radnor. 2014. "Facilitated modelling with discrete-event simulation: Reality or myth?". *European Journal of Operational Research* 234 (1): 231–240.

Rouwette, E. A. J. A., and J. A. M. Vennix. 2011. *Group Model Building*, 484–496. New York, NY: Springer New York.

Siebers, P. O., C. M. Macal, J. Garnett, D. Buxton, and M. Pidd. 2010. "Discrete-event simulation is dead, long live agent-based simulation!". *Journal of Simulation* 4 (3): 204–210.

Tako, A. A., and K. Kotiadis. 2015. "PartiSim: A multi-methodology framework to support facilitated simulation modelling in healthcare". *European Journal of Operational Research* 244 (2): 555–564.

Vandekerckhove, J., D. Matzke, and E.-J. Wagenmakers. 2015, dec. *Model Comparison and the Principle of Parsimony*, Volume 1. Oxford University Press.

Willemain, T. R. 1995. "Model Formulation: What Experts Think About and When". *Operations Research* 43 (6): 916–932.

Wilson, J. C. T. 1981. "Implementation of Computer Simulation Projects in Health Care". *Journal of the Operational Research Society* 32 (9): 825–832.

Young, T., J. Eatock, M. Jahangirian, A. Naseer, and R. Lilford. 2009. "Three Critical Challenges for Modeling and Simulation in Healthcare". In *Proceedings of the Winter Simulation Conference, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls*, 1823–1830. Piscataway, New Jersey: IEEE.

AUTHOR BIOGRAPHIES

WILLIAM JONES was a Knowledge Transfer Partnership (KTP) Associate at Kent Business School, University of Kent in partnership with Eurostar International Limited. He received a B.Sc. Eng. in Mechanical Engineering from Cardiff University, in 2013 and an EngD. from the University of Bristol, in 2019. He is now a postdoc at the University of Sydney. His research interests include modelling and simulation, the process of model development and automation. His email address is william.jones@sydney.edu.au.

KATHY KOTIADIS is a Reader (Associate Professor) in Management Science / Operational Research at Kent Business School, University of Kent. She has a B.Sc. (Hons) in Management Science and a PhD. in Operational Research both from the University of Kent. Her main research interests include discrete event simulation modelling applied to health care and simulation model development using problem structuring methods. In 2009, she was awarded the KD Tocher Medal by the UK OR society for the best simulation paper published in the Journal of Simulation in 2007-08. Her email address is k.kotiadis@kent.ac.uk.

JESSE O'HANLEY is a Professor of Environmental Systems Management at Kent Business School. His research focuses on the development and application of optimisation and simulation techniques for environmental management, facility location, transportation and logistics. He has authored dozens of articles and other publications covering a wide range of theoretical and applied topics. He is the 2015 winner of the prestigious EURO Excellence in Practice Award in recognition for the development of a state-of-the-art decision support for river connectivity restoration. His email address is j.ohanley@kent.ac.uk.

Proceedings of the Operational Research Society Simulation Workshop 2021 (SW21)
M. Fakhimi, D. Robertson, and T. Boness, eds.
DOI: https://doi.org/10.36819/SW21.044

ASSET AND LIABILITY MANAGEMENT IN INSURANCE FIRMS: A BIASED-RANDOMISED APPROACH COMBINING HEURISTICS WITH MONTE-CARLO SIMULATION

Mr. Armando Nieto

Divina Pastora Seguros
Calle Xativa 23
46004 Valencia, SPAIN
armando.nieto@divinapastora.com

Dr. Angel A. Juan
Dr. Christopher Bayliss

Universitat Oberta de Catalunya – IN3
Av. Carl Friedrich Gauss 5
08860 Castelldefels, SPAIN
{ajuanp, cbayliss}@uoc.edu

Dr. Renatas Kizys

University of Southampton
University Road
Southampton SO17 1BJ, UK
r.kizys@soton.ac.uk

ABSTRACT

The management of assets and liabilities is of critical importance for insurance companies and banks. Complex decisions need to be made regarding how to assign assets to liabilities such in a way that the overall benefit is maximised over a multi-period horizon. At the same time, the risk of not being able to cover the liabilities at any given period must be kept under a certain threshold level. This optimisation problem is known in the literature as the asset and liability management (ALM) problem. In this work, we propose a biased-randomised algorithm to solve a real-life instance of the ALM problem. Firstly, we outline a greedy heuristic. Secondly, we transform it into a probabilistic algorithm by employing Monte-Carlo simulation and biased-randomisation techniques. According to our computational tests, the probabilistic algorithm is able to generate, in short computing times, solutions that outperform by far the ones currently practised in the sector.

Keywords: Heuristics, Asset and Liability Management, Biased Randomised Algorithm, Monte Carlo

1 INTRODUCTION

Financial institutions have to face some critical risk-management processes (Cornett and Saunders 2003). Among such processes, asset and liability management (ALM) is of paramount importance due to its potential consequences. ALM consists of a range of techniques necessary to invest adequately, so that the firm's long-term liabilities are met (Ziemba et al. 1998). For an insurance company, a liability constitutes the legal responsibility to repay the insurance contributions that the customer has been making over an agreed length of time, which are increased by the interest rate. This is a typical transaction of pension or life insurance intended to secure retirement income, which gives rise to a three-tier financial problem. First, the insurance company receives the customer's premium. Second, the company invests this premium in the long term, so that the financial benefit envisaged in the insurance policy is secured. Third, in the event of the customer's retirement or death, the insurance company needs to have sufficient funds to meet

its liability to the customer. While the aforementioned financial problem unfolds, the insurance company is confronted with a range of risks, which arise either from its role as a financial intermediary or due to adverse regulatory as well economic and social policies. If the insurer's obligation to the customer is not honoured, its default becomes a likely scenario. A default can be very costly for the firm, since it can inflict a loss of credibility and reputation. On the one hand, it can face a legal action from its creditors. As a result the insurer may be forced to pay hefty fines by the regulatory body. On the other hand, the firm's market share may diminish as its customers may switch to other insurers.

It is thus not surprising that the ALM problem has been widely studied in the literature. As interest rates vary over time, the present value of both assets and liabilities responds to such variation. Consequently, optimal and smart asset management solutions become critical to the insurer, who seeks to ensure that the liabilities can be met at the time when they are required, while at the same time, the value of the firm is maximised. In practical applications, one of most popular solutions to this asset management problem is the so-called cash-flow matching (Iyengar and Ma 2009), whose main objective is to ensure the timely payment of the liabilities. In some European countries, the legislation does not envisage any specific mechanism to ensure that the firm's obligations are met. Instead, capital is regulated by targeting the value of the reserves that the company needs to build on its balance sheet. In general, regulations impose a specific interest rate to calculate the provisions of the firm's liabilities over the short and medium term. Sufficient provisions are required to achieve the solvency of the firm. Furthermore, if the firm's manager can prove that its assets are adequate to cover its liabilities in the long term, the firm is granted permission to use a higher interest rate in its provisions. This allows its capital value on the balance sheet to be lower.

Heuristic and metaheuristic algorithms have become a new standard when dealing with complex and large-scale portfolio optimisation and risk management problems (Doering et al. 2019). Hence, in this paper we propose a heuristic-based algorithm to find out which assets of a firm's portfolio can be efficiently used to reduce the risk of default liability while minimising the monetary cost for the company. Our approach combines Monte Carlo simulation (MCS) with a greedy heuristic. This combination results in a biased-randomised probabilistic algorithm. Biased-randomised algorithms make use of random sampling from a skewed probability distribution (e.g., a geometric one) in order to 'inject' some non-uniform (oriented) randomness into a greedy heuristic. That way, the latter is transformed into a more efficient probabilistic algorithm without losing the logic behind the heuristic (Grasas et al. 2017). The rest of the paper is structured as follows. Section 2 provides a brief literature review on ALM, while Section 3 reviews biased-randomised algorithms using MCS. Section 4 discusses the typical cash-flow behaviour in both assets and liabilities. Section 5 outlines the optimisation problem. Then, Section 6 proposes a greedy heuristic as an initial solving method, while Section 7 extends the aforementioned heuristic into a probabilistic algorithm. A series of computational experiments, based on real-life data, are carried out in Section 8. Finally, Section 9 concludes.

2 RECENT WORK ON ASSET AND LIABILITY MANAGEMENT

The scientific literature on ALM is quite extensive and covers several decades. Due to space limitations, we focus on research published over the last two decades. Stochastic programming models have been widely used to improve financial operations and risk management. Hence, building on multi-stage stochastic programming to model a pension fund, Kouwenberg (2001) develop scenario-generation methods for the ALM. Gondzio and Kouwenberg (2001) combine decomposition methods and high-performance computing to cope with large-scale instances of the problem. They simulate over 4 million scenarios, 12 million constraints, and 24 million variables to study a pension fund. Dempster et al. (2003) combines dynamic stochastic optimisation with Monte Carlo simulation to analyse an ALM problem involving global asset classes and contribution pension plans. Arguably, their approach can also be used to manage financial planning problems related to insurance firms, risk capital allocation, and corporate investment, among others. Additional applications and case studies on ALM can be found in Zenios and Ziemba (2007). Also, Kouwenberg and Zenios (2008) review stochastic programming models for ALM. Among other issues, they

analyse the performance of these models when applied to pension funds, discussing both their advantages and limitations.

More recently, Ferstl and Weissensteiner (2011) consider a multi-stage ALM under time-varying investment opportunities. To minimise the conditional value at risk of shareholder value, the authors utilise stochastic linear programming and a decomposition of the benefits in dynamic re-allocation. Examples of specialised books dedicated to ALM are Bauer et al. (2006), Adam (2008), Mitra and Schwaiger (2011), and Choudhry (2011). Gülpinar and Pachamanova (2013) present an ALM model based on robust optimisation techniques. Their model incorporates a time-varying aspect of investment opportunities. These authors perform a series of computational studies with real market data in order to compare the performance of their approach to that of classical stochastic programming. More recent approaches to ALM focus on the mean–variance ALM with constant elasticity of variance (Zhang and Chen 2016), random coefficients (Wei and Wang 2017), or stochastic volatility (Li et al. 2018). Fernández et al. (2018) introduce a stochastic ALM model for a life insurance company. They use GPUs to run Monte Carlo simulations in parallel. Dutta et al. (2019) employ big data analytics and stochastic linear programming in ALM under uncertainty scenarios. The authors study the relevance of employing a large number of scenarios in solving the stochastic ALM problem. Finally, Li et al. (2019) use a multi-period mean-variance model to analyse an ALM problem with probability constraints. In their model, investors seek to control for the probability of bankruptcy, while the process is influenced by uncertainty in the cash flows.

3 RECENT WORK ON BIASED-RANDOMISED ALGORITHMS

Different examples on the use of Monte Carlo simulation methods to guide the search of heuristic-based algorithms can be found in the literature (Faulin and Juan 2008, Faulin et al. 2008, Juan et al. 2009). One particular case is that of biased-randomisation (BR) techniques. As described in detail by Grasas et al. (2017), BR techniques make use of Monte-Carlo simulation and skewed probability distributions in order to transform a greedy heuristic into a probabilistic algorithm without losing the logic behind the heuristic. This transformation is achieved after sorting each constructive movement by a given criterion and then assigning diminishing probabilities of being selected as the movement becomes less promising. In practice, the use of randomised greediness here allows for a fuller exploration of the solution space, but with the advantage that the effective logic behind the greedy heuristic is retained (Figure 1).

BR techniques have been successfully used during the last years to solve different rich and realistic variants of vehicle routing problems (Dominguez et al. 2016, Calvet et al. 2016), permutation flow-shop problems (Martin et al. 2016, Gonzalez-Neira et al. 2017), location routing problems (Quintero-Araujo et al. 2017), facility location problems (De Armas et al. 2017), waste collection problems (Gruler et al. 2017), horizontal cooperation problems (Quintero-Araujo et al. 2019), and constrained portfolio optimisation problems (Kizys et al. 2019).

4 CASH FLOWS OF LIABILITIES AND ASSETS

Under an insurance policy, the insurer is liable to pay whenever the event described in the contract takes place. This is a 'must' obligation that the insurer has to honour. Otherwise, the company would face a hefty monetary fine, its reputation would be severely damaged, and its administrators could be taken to court. The insurer's liabilities comprise all policies subscribed by its customers. This aggregation results in an irregular and difficult-to-predict cash flow structure. Indeed, each policy has a different maturity and size, and is bound to a set of conditions. Being based on real-life data, Figure 2 shows a typical example of how liabilities are distributed over a period of 30 years. Figure 2 unveils a long term liability schedule, which sheds light on frequent cash flows arising from transactions in each time period. To complicate things further, these liabilities are not static, since a common policy can end in different ways: *(i)* when a customer decides to cancel it; *(ii)* when the policy reaches its maturity date; or *(iii)* when the customer dies.

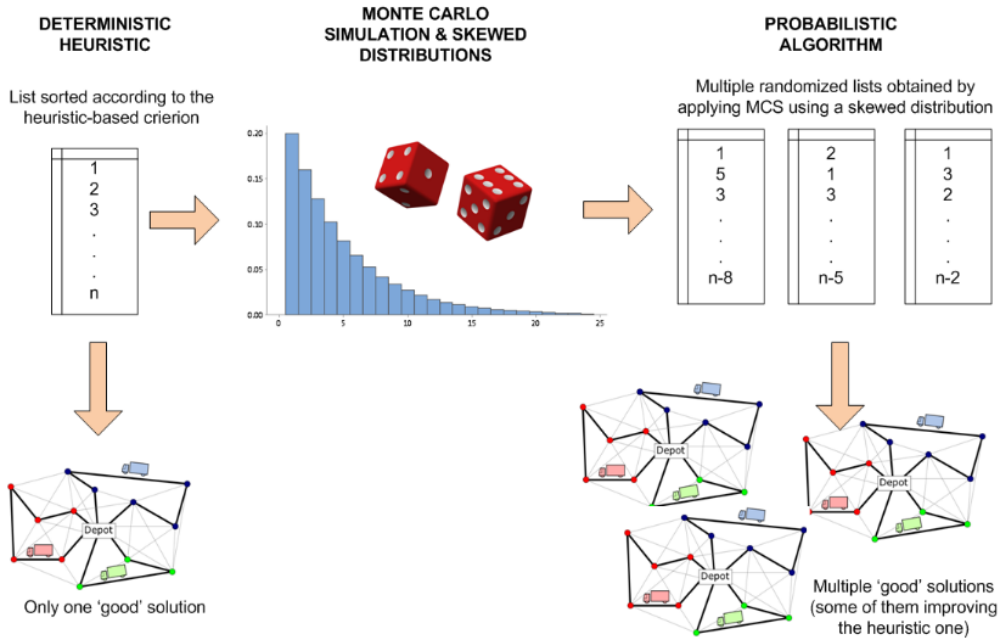


Figure 1: Schematic representation of the biased-randomisation process.

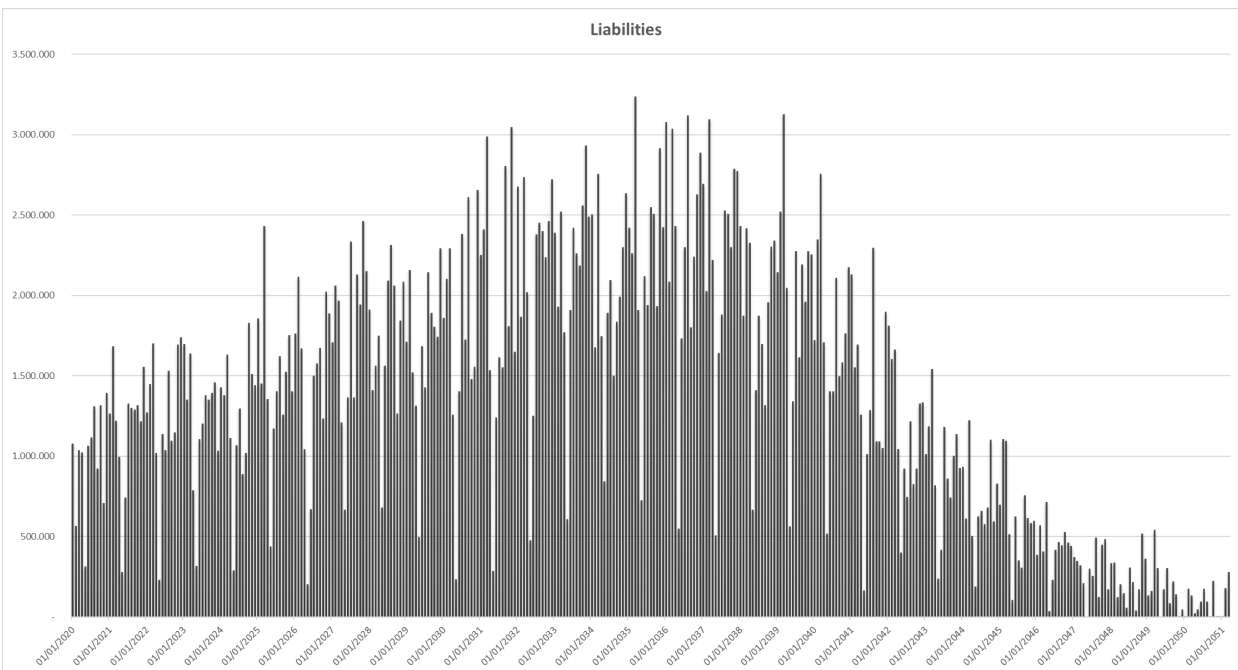


Figure 2: Liability cash-flow profile of a hypothetical portfolio.

On the flipside of the insurer's balance sheet, the manager is tasked to select a set of assets to cover the liabilities in each period. Because of the opportunity cost of these assets, the total value of these selected values should be just the necessary one, since these assets remain 'frozen' and cannot be used for any other purpose. In other words, once the assets that will cover the firm's liabilities have been selected, they

cannot be used in any other transaction. Therefore, this results in an optimisation problem, in which a set of minimum-value assets has to be determined to cover the firm's liabilities. If liabilities are assumed to be static (deterministic), assets can be optimally selected in advance. Corporate and government bonds are the predominant asset classes in the insurance market, since returns on a bond market investment can be accurately predicted in advance. The static assumption makes it simpler to predict the value of assets, as opposed to the value of liabilities. It is also worth noting that assets feature a significantly shorter span time than liabilities. For instance, while insurance contracts cover the customer's retirement or full life – which can span over 45 years – typical maturities of bond market instruments do not extend beyond 30 years. This generates a maturity mismatch between assets and liabilities. In addition, while liability cash flows might arise at any moment in time, the cash-flow structure of assets is more concentrated around some particular time periods. Figure 3 shows a typical asset portfolio associated with an insurance company. If we compare this structure with the previous one for liabilities, we can observe remarkable differences that suggest a non-trivial matching problem.

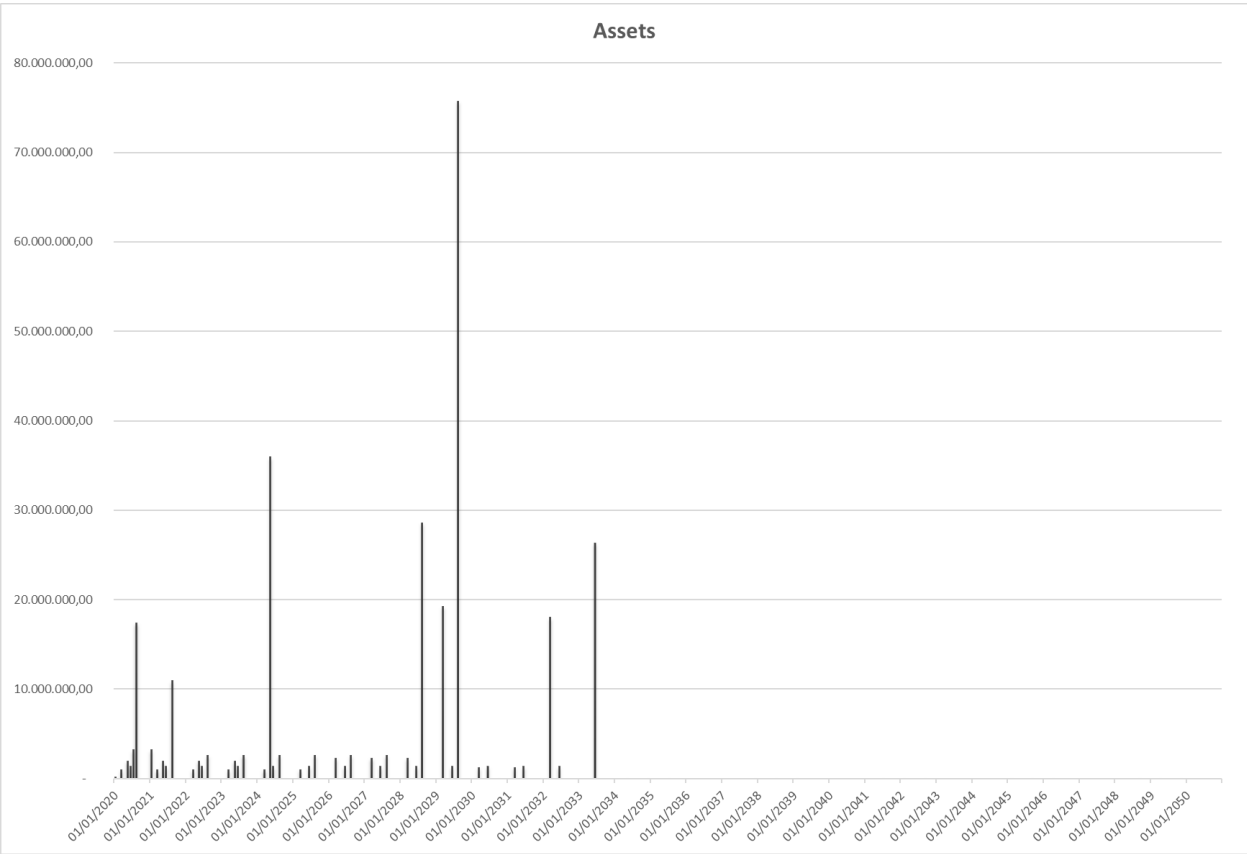


Figure 3: Asset cash-flow profile in a typical portfolio.

5 THE FINANCIAL BALANCE PROBLEM

The financial balance problem consists of choosing a portfolio of assets and blocking them just to match the liabilities. Thus, the insurer needs to manage the income arising from assets, invest it in the short term, and use the investment to pay the liabilities claimed by its customers. If its assets do not generate sufficient cash flows, then the insurance company needs to borrow, which inflicts a penalty cost. Accordingly, it is possible to formulate the financial balance as follows, where $t = 0, 1, 2, \ldots, T$ represents the time period:

$$S_0 = A_0 - L_0 \quad \text{and} \quad S_t = \begin{cases} S_{t-1} r_{t-1,i} + A_t - L_t & \text{if } S_{t-1} \geq 0 \\ S_{t-1}(r_{t-1,t} + \delta) + A_t - L_t & \text{if } S_{t-1} < 0 \end{cases} \quad \forall 1 \leq t \leq T \tag{1}$$

In Equation (1) S_0 (S_t) is the capital in period 0 (t), A_0 (A_t) denotes the value of assets in period 0 (t), L_0 (L_t) denotes the value of liabilities in period 0 (t), and $r_{t-1,t}$ is the interest rate used to capitalise the resources from period t to $t-1$. Finally, δ represent the bid-ask spread on the interest rate. it is worth noting that Equation (1) is recursive, and the balance sheet sign determines if the bid or the ask interest rate is used to capitalise the resources until the next term. If the balance is negative, the company will need to borrow. As a result, it will need to pay the ask interest rate on the credit line, which will require more capital. Notably, the selected assets have an effect on the balance sign, creating a binary tree of 2^T nodes. Moreover, if the balance falls negative its size is restricted by the credit limit.

Based on the aforementioned, we are now in a position to outline an optimisation program that solves for the optimal choice of the assets and the associate weights to match the liabilities. On an individual basis, let A_t^j be the portfolio of the firm's assets, where super-index j refers to a particular asset, and the sub-index t refers to the cash flow of asset A^j in period t. Let L_t be the cash flow associated with liabilities in period t.

The goal is to select a portion α^j of each asset A^j with the following goal:

$$\min \sum \alpha^j PV(A^j) \tag{2}$$

where PV is the present value, which is computed using the term structure of interest rates. Moreover, the selection of assets is subject to the following constraints, where τ refers to the maximum credit line of the firm:

$$S_0 = A_0 - L_0 \tag{3}$$

$$S_n \geq 0 \tag{4}$$

$$\forall t \geq 1 \quad S_t >= -\tau \tag{5}$$

6 A GREEDY HEURISTIC

In this section we propose a greedy heuristic that finds a selection of our assets, α^j. In the next section, this heuristic is extended into a biased-randomised algorithm, which allows to improve the solutions provided by the greedy heuristic. The heuristic constructs a feasible solution, one step at a time, by always choosing the 'best-next-move' in the short run (i.e., without taking into account the possible long-run implications of this selection). For that, we consider that the liability cash flow can be estimated by aggregating individual cash flows in each period of time. Then, we are interested in solving a simplified matching problem, which considers just the cash flow associated with one of these liabilities; the specific liability is randomly selected. Once the chosen liability has been matched by a set of assets, a new liability cash flow is randomly chosen and new assets (from the remaining ones) are drawn to cover it. This process is re-iterated until all the liabilities have been covered by asset cash flows (Algorithm 1).

Notice that the first step in Algorithm 1 is to decide an order for the list of liability cash flows. A natural order is the one given by the maturity date, so that the next-in-time cash flow that will have to be payed is introduced first, the second one is next, and so on. The selection of the best asset is quite simple, since we have to match only one liability cash flow at a time. Hence, we only have to iterate over the remaining assets to get the minimum fraction needed to match our new liability. Only assets with a value larger than the current liability value are considered.

Algorithm 1 Greedy heuristic

```
Order liabilities by maturity date.
for each liability k do
  Insert k
  repeat
    for each asset j do
      Calc asset fraction needed to match k
      Select best asset so far, j*
    end for
  until Liability k is matched
end for
```

7 A BIASED-RANDOMISED ALGORITHM

By examining Algorithm 1, one can notice the following: once the order of the liabilities to be matched has been fixed, the solution (set of assets chosen to cover the liabilities) is unique. This suggests than one way to generate different solutions is by introducing a biased-randomised process when sorting the liabilities. To this end, we make use of a skewed probability distribution (the geometric one in our case) to re-order the liabilities list, hence using Monte Carlo simulation to generate a differently ordered lists in each run of the algorithm. The geometric distribution only requires a parameter, $p \in (0,1)$. As p converges to 1, the list tends to be sorted following the greedy criterion employed by the initial heuristic (i.e., by maturity date). On the contrary, as p converges to 0, the list tends to follow a uniformly random order. The values in between are the interesting ones, since they represent a compromise between a greedy and a uniform random order. Figure 4 shows the Java code employed to generate the biased-randomisation effect.

```
indexMax = L;
Random rnd = new Random();
for (int i = 0; i < L; i++)
{
    index = ((int)(Math.Log(rnd.NextDouble()) / Math.Log(1 - GeomPar))) % indexMax;
    LiabilityList[i] = RevertedLiability[indexMax - 1 - index];
    RevertedLiability.RemoveAt(indexMax - 1 - index);
    indexMax--;
}
```

Figure 4: Code for the biased-randomised selection of liabilities.

8 COMPUTATIONAL EXPERIMENTS

In order to test our method, we have considered data from a real-life insurance firm. This firm holds 21 assets, which are predominantly government bonds and interest rate swaps. We have considered a discount rate of 1.09%, an interest rate to capitalise resources of 0.5%, and a time span of 33 years. For the geometric distribution, a parameter $p = 0.8$ has been selected after a quick trial-and-error process. Also, we have used 100 iterations, a maximum credit of 1 million euros, and assumed that the credit line carries a 5% interest rate. The origin of the liabilities are pensions, and their present value is 442 million euros.

The assets selected by the actuarial team add up to 490 million euros. Running our algorithm for a few seconds, we found a solution with an associated value of 450 million euros, which represents an 8% savings with respect the solution provided by the actuarial team. As shown in Figure 5, the solution

structure is not trivial, so it is not surprising that it could not be found without the help of an algorithm as the one proposed here.

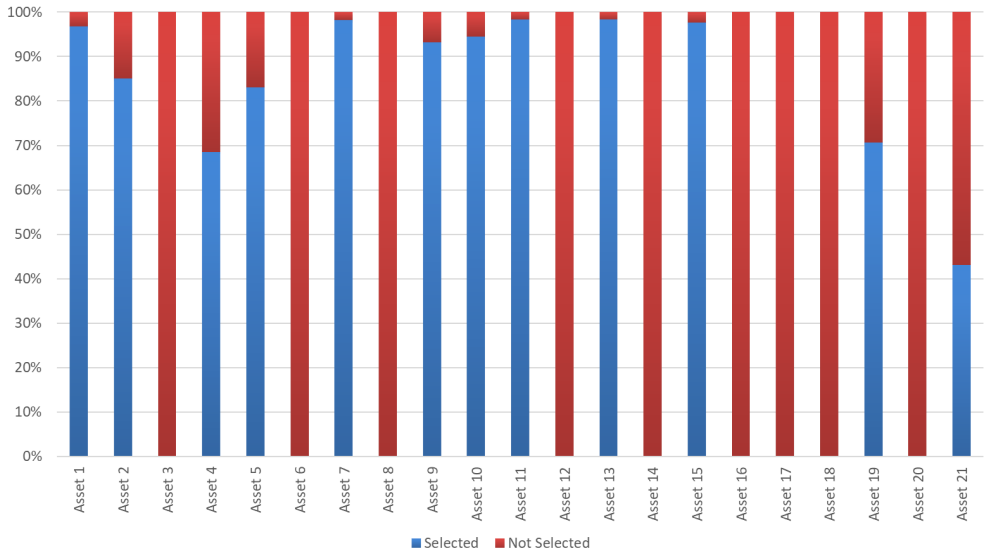


Figure 5: A solution showing the selection of assets and percentages.

Testing different values for the parameter p does not seem to provide significantly better results. The fastest result is found if the original liability order is based on the present value of each liability cash flow. This makes sense, as we first match the largest liability values with the best possible asset. Using this initial order criteria, only 100 iterations are necessary to get a high-quality solution.

9 CONCLUSIONS

This paper proposes a solving approach for the asset and liability management problem. Our algorithm makes use of Monte Carlo simulation to transform a greedy heuristic into a probabilistic algorithm. The resulting biased-randomised algorithms is a fast and easy-to-implement method for selecting the minimum amount of assets to cover a portfolio of liabilities. Our method is flexible and it can be easily extended to new constraints, either if they provide from a specific regulation or from the firm's strategy. Our approach can be used in a real-life situation by iteratively applying it to a set of liabilities. According to our computational experiments, the savings it generates can be considerable. Considering that the insurance market is strongly regulated, having an efficient, flexible, and easy-to-implement method to select the proper assets inside a firm's portfolio is extraordinarily important.

As future work, we plan to: *(i)* extend our probabilistic algorithm into a full metaheuristic one; and *(ii)* test the algorithm in more benchmark data sets –some of them using real-life data.

ACKNOWLEDGMENTS

This work has been partially funded by the agreement between the Divina Pastora Seguros and the Universitat Oberta de Catalunya.

REFERENCES

Adam, A. 2008. *Handbook of Asset and Liability Management: From models to optimal return strategies*. John Wiley & Sons.

Bauer, R., R. Hoevenaars, and T. Steenkamp. 2006. *Asset liability management*. Oxford: Oxford University Press.

Calvet, L., A. Ferrer, M. I. Gomes, A. A. Juan, and D. Masip. 2016. "Combining statistical learning with metaheuristics for the multi-depot vehicle routing problem with market segmentation". *Computers & Industrial Engineering* 94:93–104.

Choudhry, M. 2011. *Bank asset and liability management: strategy, trading, analysis*. John Wiley & Sons.

Cornett, M. M., and A. Saunders. 2003. *Financial institutions management: A risk management approach*. McGraw-Hill/Irwin.

De Armas, J., A. A. Juan, J. M. Marquès, and J. P. Pedroso. 2017. "Solving the deterministic and stochastic uncapacitated facility location problem: from a heuristic to a simheuristic". *Journal of the Operational Research Society* 68 (10): 1161–1176.

Dempster, M. A., M. Germano, E. A. Medova, and M. Villaverde. 2003. "Global asset liability management". *British Actuarial Journal* 9 (1): 137–195.

Doering, J., R. Kizys, A. A. Juan, A. Fito, and O. Polat. 2019. "Metaheuristics for rich portfolio optimisation and risk management: Current state and future trends". *Operations Research Perspectives* 6:100121.

Dominguez, O., D. Guimarans, A. A. Juan, and I. de la Nuez. 2016. "A Biased-randomised large neighbourhood search for the two-dimensional vehicle routing problem with backhauls". *European Journal of Operational Research* 255 (2): 442–462.

Dutta, G., H. V. Rao, S. Basu, and M. K. Tiwari. 2019. "Asset liability management model with decision support system for life insurance companies: Computational results". *Computers & Industrial Engineering* 128:985–998.

Faulin, J., M. Gilibert, A. A. Juan, X. Vilajosana, and R. Ruiz. 2008. "SR-1: A simulation-based algorithm for the capacitated vehicle routing problem". In *Proceedings of the 2008 Winter Simulation Conference*, 2708–2716. IEEE.

Faulin, J., and A. A. Juan. 2008. "The ALGACEA-1 method for the capacitated vehicle routing problem". *International Transactions in Operational Research* 15 (5): 599–621.

Fernández, J. L., A. M. Ferreiro-Ferreiro, J. A. García-Rodríguez, and C. Vázquez. 2018. "GPU parallel implementation for asset-liability management in insurance companies". *Journal of Computational Science* 24:232–254.

Ferstl, R., and A. Weissensteiner. 2011. "Asset-liability management under time-varying investment opportunities". *Journal of Banking & Finance* 35 (1): 182–192.

Gondzio, J., and R. Kouwenberg. 2001. "High-performance computing for asset-liability management". *Operations Research* 49 (6): 879–891.

Gonzalez-Neira, E. M., D. Ferone, S. Hatami, and A. A. Juan. 2017. "A biased-randomized simheuristic for the distributed assembly permutation flowshop problem with stochastic processing times". *Simulation Modelling Practice and Theory* 79:23–36.

Grasas, A., A. A. Juan, J. Faulin, J. de Armas, and H. Ramalhinho. 2017. "Biased randomization of heuristics using skewed probability distributions: A survey and some applications". *Computers & Industrial Engineering* 110:216–228.

Gruler, A., C. Fikar, A. A. Juan, P. Hirsch, and C. Contreras-Bolton. 2017. "Supporting multi-depot and stochastic waste collection management in clustered urban areas via simulation–optimization". *Journal of Simulation* 11 (1): 11–19.

Gülpinar, N., and D. Pachamanova. 2013. "A robust optimization approach to asset-liability management under time-varying investment opportunities". *Journal of Banking & Finance* 37 (6): 2031–2041.

Iyengar, G., and A. K. C. Ma. 2009. "Cash flow matching: a risk management approach". *North American Actuarial Journal* 13 (3): 370–378.

Juan, A. A., J. Faulin, R. Ruiz, B. Barrios, M. Gilibert, and X. Vilajosana. 2009. "Using oriented random search to provide a set of alternative solutions to the capacitated vehicle routing problem". In *Operations*

Research and Cyber-Infrastructure, edited by J. W. Chinneck, B. Kristjansson, and M. J. Saltzman, 331–345. Boston, MA: Springer US.

Kizys, R., A. A. Juan, B. Sawik, and L. Calvet. 2019. "A Biased-randomized iterated local search algorithm for rich portfolio optimization". *Applied Sciences - Basel* 9 (17).

Kouwenberg, R. 2001. "Scenario generation and stochastic programming models for asset liability management". *European Journal of Operational Research* 134 (2): 279–292.

Kouwenberg, R., and S. A. Zenios. 2008. "Stochastic programming models for asset liability management". In *Handbook of Asset and Liability Management*, 253–303. Elsevier.

Li, D., Y. Shen, and Y. Zeng. 2018. "Dynamic derivative-based investment strategy for mean–variance asset–liability management with stochastic volatility". *Insurance: Mathematics and Economics* 78:72–86.

Li, X., X. Wu, and H. Yao. 2019. "Multi-period asset-liability management with cash flows and probability constraints: A mean-field formulation approach". *Journal of the Operational Research Society*:1–18.

Martin, S., D. Ouelhadj, P. Beullens, E. Ozcan, A. A. Juan, and E. K. Burke. 2016. "A multi-agent based cooperative approach to scheduling and routing". *European Journal of Operational Research* 254 (1): 169–178.

Mitra, G., and K. Schwaiger. 2011. *Asset and Liability Management Handbook*. Springer.

Quintero-Araujo, C. L., J. P. Caballero-Villalobos, A. A. Juan, and J. R. Montoya-Torres. 2017. "A biased-randomized metaheuristic for the capacitated location routing problem". *International Transactions in Operational Research* 24 (5): 1079–1098.

Quintero-Araujo, C. L., A. Gruler, A. A. Juan, and J. Faulin. 2019. "Using horizontal cooperation concepts in integrated routing and facility-location decisions". *International Transactions in Operational Research* 26 (2): 551–576.

Wei, J., and T. Wang. 2017. "Time-consistent mean–variance asset–liability management with random coefficients". *Insurance: Mathematics and Economics* 77:84–96.

Zenios, S. A., and W. T. Ziemba. 2007. *Handbook of Asset and Liability Management: Applications and case studies*, Volume 2. Elsevier.

Zhang, M., and P. Chen. 2016. "Mean–variance asset–liability management under constant elasticity of variance process". *Insurance: Mathematics and Economics* 70:11–18.

Ziemba, W. T., J. M. Mulvey, H. Moffatt et al. 1998. *Worldwide Asset and Liability Modeling*, Volume 10. Cambridge University Press.

AUTHOR BIOGRAPHIES

ARMANDO NIETO is the CEO of Divina Pastora Seguros, one of the largest insurance firms in Spain. He holds a BSc and MSc in Physics from the Universidad de Valencia, and is currently completing his PhD in Computational Finance at the Universitat Oberta de Catalunya. His e-mail address is: anietoran@uoc.edu.

ANGEL A. JUAN is a Full Professor of Operations Research in the Computer Science Dept. at the Universitat Oberta de Catalunya (Barcelona, Spain). Dr. Juan holds a PhD in Industrial Engineering and an MSc in Mathematics. He has published about 80 articles in JCR-indexed journals. His website address is http://ajuanp.wordpress.com and his email address is ajuanp@uoc.edu.

RENATAS KIZYS is a Associate Professor in Finance at the University of Southampton. He holds a MA in Economics (Vilnius University, Lithuania) and a PhD in Economics (University of Alicante, Spain). His research interests centre on Financial Economics. His e-mail address is r.kizys@soton.ac.uk.

CHRISTOPHER BAYLISS is a post-doctoral researcher in the IN3 – Universitat Oberta de Catalunya. He got a PhD in Computational Mathematics at the University of Nottingham. His main research interests include metaheuristics and simulation-optimisation. His email address is cbayliss@uoc.edu.

#SW21

www.theorsociety.com

@theorsociety

 | ISBN: 978-0-903440-66-0