Proceedings

Edited by
Dr Anastasia Anagnostou
Dr Katy Hoad
Dr Martin Kunc

11–13 April 2016
Ettington Chase Hotel, Warwickshire

in association with

The OR Society gratefully acknowledges the support of the following sponsors
INTRODUCTION TO THE WORKSHOP

Welcome to the 2016 Operational Research Society Simulation Workshop (SW16)

This year, we have made a few changes to the conference, which we hope will rejuvenate it. First, and most obviously, we have moved to a new venue: Ettington Chase, near Stratford-upon-Avon, which we expect to provide an excellent standard of accommodation and conference facilities. Secondly, we have included tutorials in the programme for the first time, from some prestigious names in discrete event, agent-based and system dynamics simulation. We hope that these, combined with our keynotes, contributed papers and posters, will provide a stimulating programme.

We have 28 contributed papers (including case studies), seven posters, six tutorials, one workshop and two keynote speakers on areas of simulation including simulation in disaster management, decision support and in another new development for this conference, sessions on Agent-Based Modelling and System Dynamics.

We are delighted to welcome Professor Sally Brailsford from the University of Southampton and Professor Alexander Verbraeck from Delft University of Technology as our keynote speakers. Sally will be talking about hybrid simulation, a hot topic in the simulation literature, and a talk that may well generate some debate. Alexander will be discussing Data Driven Simulation and the challenges associated with it.

In keeping with tradition, there will be a panel session on the final day of the workshop. This year we are celebrating the tenth anniversary of the Journal of Simulation (JoS), with a panel session that looks back over the past decade and forward to the next.

After it proved to be a lot of fun at SW14, this year sees the return of poster, plenary and competition to the conference. We wish our poster delegates the best of luck. Please take the time to discuss their work with them during the poster session.

From the SW16 contributed papers, we will approach some authors to suggest that they extend their work for submission to JoS, which publishes an exciting range of papers with a focus on the practice and application of simulation. OR Society members have free access to JoS online (by logging on to www.theorsociety.com and entering their user name and password). For those with institutional libraries, do remember to ask the library to subscribe to JoS.

Our thanks once again to all the contributors, the sponsors (who have also provided an exhibition of simulation products and services), the stream organisers and the reviewers. We also thank those whose hard work has made SW16 possible, especially Dr Anastasia Anagnostou, Dr Kathryn Hoad, Dr Martin Kunc (programme chairs), Dr Anastasia Gogi (poster chair) and Mrs Hilary Wilkes (administration). Finally, we are very grateful for the thoughtful advice that Professor Stewart Robinson has given to us on the organisation of the workshop.

Enjoy it!
Tom Monks and Christine Currie
Conference Chairs
CONTENTS

Tutorials

Agent-based Simulation
An Introduction to Agent-based Simulation as a Decision-Support Tool ........................................... 1
Bhakti S. S. Onggo

Conceptual Modelling
Conceptual Modelling for Simulation: A Tutorial.................................................................................. 13
Stewart Robinson

Input Analysis
Including Parameter Uncertainty in Simulation Experiments .......................................................... 25
Russell

Agent-Based Modelling
An Agent-based Model of Supply Chain Collaboration: Investigating Manufacturing Loyalty .......... 35
Niniet Arvitrida, Stewart Robinson, Antuela Tako, Duncan Robertson

Agent-based Simulation of Herding in Financial Markets .................................................................... 45
Xin Liu, Roger Brooks

Decision Support
Effective Quality Assurance of Government Models Used to support Business Critical Decisions .... 54
Alan Robinson, Paul Glover

An Agent-Based Model of Knowledge Transferral: Exploring the Need for Closure and Cognition ...... 64
Duncan Robertson, L. Alberto Franco

Disaster Management
Simulation as a Validation Tool – A Scenario Based Analysis of Disaster Haiyan ............................... 72
Sara Hasani, Emel Aktas

The impact of policy change on maritime logistics’ level of disaster preparedness and resilience .......... 81
John Kwesi Buor

Police Simulation
Maximising Police Officer Response via Informed Decision Making .................................................. 92
Johanna Leigh, Lisa Jackson, Sarah Dunnett

A Simulation of a Police Patrol Service System with Multi-grade Time-varying Incident Arrivals 103
Hanjing Zhang, Lisa Jackson, Antuela Tako, Jiyin Liu
# Simulation in Practice

**A Linguistic Examination of Problem Solving Approaches in Simulation** .................................................. 113  
*Roger McHaney, Antuela Tako, Stewart Robinson*

Impact of Model Fidelity in Factory Layout Assessment Using Immersive Discrete Event Simulation . 124  
*Alessandro Petti, Windo Hutabarat, John Oyekan, Chris Turner, Ashutosh Tiwari, Neha Prajapat, Xiao-Peng Gan*

Simulating single test rule-out strategies for chest pain patients at emergency departments ............... 135  
*Sebastian Rachuba, Andrew Salmon, Zhivko Zhelev, Martin Pitt*

# Simulation Modelling Methodology

**Entering New Fields of Simulation Application – Challenges Faced in Simulation Modelling of Stroke Systems** …………………………………………………………………………………………………………………………………… 145  
*Durk-Jouke Van der Zee, Thomas Monks, Maarten M.H. Lahr, Gert-Jan Luijckx, Erik Buskens*

Black Boxes, Agency and Simulation: How Realism and Agent based Simulation Can Work Synergistically for Intervention Development…………………………………………………………………………………………………………………………… 155  
*Sean Manzi, Daniel Chalk, Mark Pearson, Jo Day, Ken Stein, Iain Lang, Martin Pitt*

**A Facilitation Workshop for the Implementation Stage: A Case Study in Health Care** ……………… 165  
*Kathy Kotiadis, Antuela Tako*

# Simulation Modelling Platforms

An Object-Functional Modelling Platform to Enable Semantic Web Analysis of Campaign Models….. 175  
*Gareth Toomey*

Towards Self-Adaptive Discrete Event Simulation (SADES)……………………………………………………… 181  
*Kathy Kotiadis*

WS-PGRADE Workflows for Cloud-Based Distributed Simulation ………………………………………………… 192  
*Nauman R. Chaudhry, Athar Nouman, Anastasia Anagnostou, Simon J. E. Taylor*

# Simulation Modelling Practices

Arguments For and Against the Use of Multiple Comparison Control in Stochastic Simulation Studies 202  
*Thomas Monks, Christine Currie, Kathryn Hoad*

Selection Procedures for Simulation: A Preliminary Investigation of Ideas from machine Learning...... 212  
*Christine Currie, Andria Ellina*
**System Dynamics**
Utilisation of Health and Social Care Services by the Over 65s Population. A System Dynamics Study

*Rudabeh Meskarian, Marion Penn, Thomas Monks, Melinda Taylor, Jonathan Klein, Sally Brailsford, Paul Benson*

Influence in Multi-Party Negotiations: a System Dynamics Approach

*Matthew Gilbert, Simon French, Jim Smith*

System Dynamics’ Potential in Developing Safety in Complex Systems – Literature Review and Pilot Study for Medication Safety in a Hospital Pharmacy

*Mohammed Ibrahim Shire, Gyuchan Thomas Jun, Stewart Robinson*

**Workshop**

The Hospital Discharge Game: A Game Theory-Inspired Workshop to Encourage Cooperation Between Health and Social Care Organisations

*Daniel Chalk, Sean Manzi*
AN INTRODUCTION TO AGENT-BASED SIMULATION AS A DECISION-SUPPORT TOOL

Bhakti Stephan Onggo
Department of Management Science
Lancaster University Management School
Lancaster University, United Kingdom
s.onggo@lancaster.ac.uk

ABSTRACT
Agent-based simulation (ABS) has been used widely in various application areas. This paper provides a tutorial on ABS. In addition to the theoretical foundations, such as the definitions of key ABS concepts and the structure of an ABS model, this tutorial also explains the technical aspect, such as how an ABS model is implemented using a software tool and how ABS is used to solve a problem. The software tool used in this tutorial is Repast. The last part of the tutorial discusses the challenges that need to be addressed in order to increase the confidence of decision-makers, who may not be familiar with computer programming, when using ABS to help them make decisions. The challenges include conceptual model representation, validation and participatory modelling.

Keywords: Agent-based simulation, Repast, conceptual modelling, validation, participatory modelling

1 INTRODUCTION
Agent-based simulation (ABS) has been applied in various areas. In business, Bonabeau (2002) categorised the application of ABS into four categories: flow management (e.g. pedestrians, traffic), markets (e.g. stock, electricity), organisational behaviour (inter- and intra-organisation such as manufacturing and supply chain) and diffusion (the spread of something in a society). Macal and North (2010) provide a different categorisation which includes epidemics, markets and socio-technical systems. Although Bonabeau (2002) and Macal and North (2010) do not mean to conduct extensive surveys, their observations are largely supported by recent surveys. Diffusion in a society is one of the main business applications of ABS, as we can see from the review by Kiesling et al. (2012). This is understandable because an ABS model can model people and how they interact within their social networks. Similarly, using ABS to model a market is sensible because a market would not exist without the interaction of trading agents. An ABS model can model how trading agents interact within their business networks by making decisions such as buy, hold or sell. An example of a review paper for the electricity market is given by Sensfuß et al. (2007). For the same reasons as in the previous two applications, the application of ABS in a supply chain or pedestrian/crowd management is also popular. A number of examples on the application of ABS in supply chains can be found in Peidro et al. (2009), who survey the quantitative models used in supply-chain planning, which includes ABS. A review on the modelling techniques used in crowd simulation which include ABS is conducted by Zhou et al. (2010).

Given the increase in the popularity of ABS, there have been a number of tutorial papers written to introduce ABS. Some tutorials focus more on the theoretical foundations of ABS (e.g. Chan et al. 2010, Macal and North 2010). Others explain both the theoretical foundations and technical knowledge (i.e. how an ABS model can be implemented using a software tool); for example, Macal and North (2007) use Ms Excel and Terna (1998) uses SWARM. This tutorial paper also explains both the theoretical and implementation aspects of ABS. The ABS concepts described in this paper are implementation-
independent, so they can be applied to other software tools. As for the software tool here, this paper uses Repast (North et al. 2013) which can be downloaded from http://repast.sourceforge.net/. The second part of the paper discusses the challenges that need to be addressed in order to increase the confidence of decision-makers, who may not be familiar with computer programming, when using ABS to help them make decisions.

To make the explanation more concrete, this paper will use Schelling’s segregation case (Schelling 1978) as an example. This case is chosen because of its simplicity although, at the same time, it is able to demonstrate the advantages of ABS. The aim of the study is to explain why segregation exists in a certain society. One possible explanation is that segregation may occur if each individual in society does not want to be in a minority in his/her neighbourhood. An ABS model is developed to prove that an individual’s decision to move or stay in his/her neighbourhood is able to generate the pattern (i.e. segregation) observed at the population level. All the computer code used in this paper is taken from the sample code provided by Repast.

The remainder of this paper is structured as follows. Section 2 explains ABS and its main components (agents, environment and behaviours/interactions). The approach used in this tutorial paper is to explain a concept followed by its implementation in Repast. The objective is to help the reader understand the concept better, not only by explanation of the concept but also by looking at the implementation of it. Section 3 highlights the challenges that we need to address to encourage decision-makers to use ABS as a decision-support tool. The challenges include conceptual model representation, model validation and participatory modelling.

2 AGENT-BASED SIMULATION

There are a number of terminologies that are similar to ABS, such as multi-agent systems (MAS), agent-directed simulation and distributed agents. Some articles use different terminology, such as MAS and ABS, to refer to the same thing. On the other hand, different articles may use the same terminology, such as ABS, to refer to two different things. This is understandable given the fact that these related techniques have been applied in various application areas relatively independently.

This paper adopts the following definition: “An ABS model is a simulation model that is formed by a set of autonomous agents that interact with their environment and other agents through a set of internal rules to achieve their objectives.” ABS is one of the paradigms in simulation modelling (others include discrete-event simulation (DES) and system dynamics (SD)). ABS modellers see the world as a set of interacting agents inhabiting in an environment. Their view is different from that of DES modellers who see the world as a set of interacting queues and activities, or the view of SD modellers who see the world as a set of interconnected flows and stocks. The agents in ABS are autonomous, meaning that we do not model a central controller who coordinates the behaviour of agents. Hence, agents are capable of making independent decisions. For example, if we simulate the movements of individuals in a city, each agent (i.e. individual) makes an independent decision whether or not to move to a new neighbourhood or stay in the current one. Agents interact with other agents directly or indirectly (through their environment). These interactions are driven by rules that are internal (or local) to each agent. These rules are followed because each agent wants to achieve its objectives (for example, an agent wants to be in a majority in his/her neighbourhood). The interactions between agents may generate a certain pattern that appears at the population level (known as emergence behaviour). For example, the interaction between agents may generate a segregation pattern. This segregation pattern emerges from the interaction of agents, rather than a central controller in the model. Arguably, the main advantage of ABS comes from its ability to generate emergence behaviour from local interactions between heterogeneous agents. Given the definition above, it is clear that the main components of an ABS model are a set of agents in an environment and their behaviours/interactions.
2.1 Agents

There is no consensus in the ABS literature on the definition of an agent (North and Macal 2007). Instead, we find a spectrum of complexity in the definition of an agent. At one extreme, the composition of an ABS model can be as simple as a set of homogeneous agents with simple attributes (such as speed and direction) and simple behaviours (such as move and accelerate). This type of agent is referred to as a pseudo-agent by North and Macal (2007). At the other extreme, an ABS model can be formed by a set of heterogeneous agents with various complex attributes (such as memory) and abilities (such as communication, perception, planning and learning).

Consistent with the ABS definition used in this paper, an agent is defined as “an entity that can make an independent decision in order to achieve its objectives”. An agent can be human or non-human. When we develop an ABS model, it is important to understand the difference between agent and agent type. An agent type is a modelling construct that represents all agents that can be identified by using the same set of characteristics and are able to perform the same set of activities. In a hospital model, the types of agent include patient, clerk and doctor. A patient is an agent type because all patients can be identified by using the same set of characteristics (such as patient ID/social-security number, name and severity level) and are able to perform the same activities (such as arrive at the clinic, complete a form). From an agent type, we can generate one or more agents of the same type. In other words, an agent is an instance of agent type. For example, from the agent type patient, we can generate two patients called Joe and Jane. Both Joe and Jane have a unique ID. They also need to perform the same activities, such as arrive at the hospital, but their arrival and method of transport can be different.
In Repast, we need to define an agent type as a class. The following code shows how to define an agent type used in the segregation model. Line 01 defines a class (i.e. agent type) called Agent. We can change this name to another name, for example Individual. This agent type has a number of characteristics/properties as defined in lines 02–06. Line 07 is a constant to indicate the number of colours that may represent ethnicity or any group to which the agent belongs (Figure 1 shows that each agent belongs to one of three colours). Line 08 is the constructor for the agent type. This is where we should perform the initialization. Lines 09–10 define a one-agent behaviour called step, which is executed in every time step from time step 0. This synchronous time-advance mechanism is common to ABS where each agent performs an activity in every time step. In addition to this synchronous time-advance mechanism, Repast provides a very time-flexible scheduler that allows us to have advanced control over when an activity should be performed by an agent. Line 11 defines another behaviour called die, which is only executed when this method is called.

```java
01. public class Agent {
02.   private int maxAge, currentAge, type;
03.   private double percentLikeNeighbors;
04.   private String id;
05.   private double numLikeMe;
06.   private double neighborCount;
07.   private int numberOfAgentTypes = 3;
08.   public Agent(String id) { /* TO BE ADDED */ }
09.   @ScheduledMethod(start=0,interval=1)
10.   public void step() { /* TO BE ADDED */ }
11.   public void die(){ /* TO BE ADDED */ }
12. }
```

The detailed code for the agent’s initialisation is shown below. When an agent is created, it will be assigned a unique ID (lines 01–02). Line 03 is standard code to retrieve simulation parameters from the user interface (Figure 1). Each parameter shown in the user interface (such as “Min Death Age” and “Max Death Age”) is linked to the code using a unique identifier. For example, the parameter “Min Death Age” is linked to the code using the identifier “minDeathAge” (line 04). Subsequently, we store the parameter value (80 in Figure 1) as the variable minDeathAge. The variables minDeathAge and maxDeathAge are used to generate the maximum age of an agent based on a uniform distribution (line 07). Line 06 assigns the colour of the agent that will be used to decide whether the agent will move to a new neighbourhood or stay in the current one. Line 08 retrieves the minimum percentage of neighbours with the same colour that will make the agent stay in the current neighbourhood. This means that all agents have the same threshold (homogeneous). We can make the agents heterogeneous by sampling the threshold values using a distribution function, such as the uniform distribution used in line 07.

```java
01. public Agent(String id) {
02.   this.id = id;
03.   Parameters p = RunEnvironment.getInstance().getParameters();
04.   int minDeathAge = (Integer)p.getValue("minDeathAge");
05.   int maxDeathAge = (Integer)p.getValue("maxDeathAge");
06.   this.type = RandomHelper.nextIntFromTo(0, numberOfAgentTypes-1);
07.   this.maxAge = RandomHelper.nextIntFromTo(minDeathAge, maxDeathAge);
08.   this.percentLikeNeighbors = (Double)p.getValue("percentLikeNeighbors");
09. }
```

### 2.2 Environment

The second component in an ABS model is the environment where the agents live. The environment can be spatial or relational. The spatial environment refers to the physical location where an agent lives, such as a region, city or coordinates. The relational environment refers to the connections between agents, e.g. social network, communication network or friendship network.

The environment is an important part of an ABS model because it may affect agents and their behaviours. In the segregation example, an individual makes a decision to move or to stay based on the state of his/her neighbourhood. Furthermore, the environment can be used as a medium for the interactions between agents. In this case, an agent may change the environment, which consequently affects other agents. In the segregation example, when an individual moves, s/he changes the state of the old and new neighbourhoods. These changes will affect the decisions made by other individuals in the old and new neighbourhoods. Often, we need to model an environment to be dynamic, i.e. the environment changes even when all agents do nothing to it. In other words, an environment can have behaviours too. Hence, in the implementation, it can be implemented as an agent. All of these show that the environment is an important part of an ABS model.

**Figure 2** Repast’s projections (from top left clockwise: grid, continuous space, network on top of grid and space, GIS) – all taken from samples provided by Repast

Repast implements the environment in an ABS by adding a *Projection* to a context. Hence, a context must exist before a projection is added. A projection imposes a certain structure on agents in a context. The
current version of Repast supports the following structures: Grid, Continuous Space, Geographical Information System (GIS) and Network, as shown in Figure 2.

A grid projection creates cells that are structured in one or more dimensions. For example, a 2D grid creates a chessboard-like environment for the agents. In this projection, each agent lives in a cell. This is a structure used in many popular ABS models including Schelling’s segregation model. The code snippet below shows how to add a grid projection to a context. The approach follows the factory design pattern (one of the commonly used design patterns in software development). First, we need to create a factory (line 01). Then, we create a grid projection using the factory (line 02). To create a grid projection, we need to give a name to the grid projection, the context and the grid parameters. A description of all the parameters can be read from the Repast API document (it comes with the software). In the example below, the grid will have wrap-around borders (like a toroid), the agents will be added to the grid manually (hence, agents will not be allocated to cells until we do it ourselves), multiple agents can live in the same cell in the grid, and the size of the grid is 100×100 cells.

```java
01. GridFactory gridFactory = GridFactoryFinder.createGridFactory(null);
02. grid = gridFactory.createGrid("a grid name", context,
    new GridBuilderParameters<Object>(
        new WrapAroundBorders(),
        new SimpleGridAdder<Object>(),
        true, 100, 100
    )
);
```

In an n-dimensional continuous-space projection, the location of an agent is represented by a vector \((x_1, x_2, \ldots, x_n)\). The approach to add a continuous-space projection to a context is the same as in grid projection. The example below creates a continuous space projection in which agents will be added to the space in random locations, the space forms a toroid and the size of the space is 100×100.

```java
01. ContinuousSpaceFactory spaceFactory =
    ContinuousSpaceFactoryFinder.createContinuousSpaceFactory(null);
02. space = spaceFactory.createContinuousSpace("a space name", context,
    new RandomCartesianAdder<Object>(),
    new repast.simphony.space.continuous.WrapAroundBorders(), 100, 100
);
```

A GIS projection enables us to model a more realistic-looking geographical area in which agents live. In this projection, each agent is associated with a spatial geometry, such as a polygon or point. This projection by default uses a longitude-latitude coordinate reference system. The approach to create this projection is the same as in grid projection, as shown in the snippet below.

```java
01. GeographyParameters geoParams = new GeographyParameters();
02. Geography geography =
    GeographyFactoryFinder.createGeographyFactory(null).createGeography(
        "Geography", context, geoParams);
```

A network projection is used to create a relationship network between the agents. The relationship can be physical (such as a road network or a water-distribution network) or non-physical (such as a social network or a professional network). The approach to create a network projection is slightly different from the other projections. A network is created using NetworkBuilder, as shown in lines 01–02 below. The last parameter in line 01 specifies whether or not the network is directed. The code snippet below creates a vertex for each agent in the context. However, the edges between vertices (i.e. relationships between agents) are not created. We need to add relationships using an addEdge method later. It is possible to create a network with edges from an existing file or from an algorithm. In this case, we can insert a new line between lines 01 and 02 and call the method load to load a network from a file, or use NetworkGenerator to generate
Ongoing

a network from an algorithm such as small-world or random. We can also implement our own network-generator algorithm.

```java
01. NetworkBuilder<Object> netBuilder =
    new NetworkBuilder<Object>("a network name", context, true);
02. netBuilder.buildNetwork();
```

Repast allows us to associate multiple projections with a context. This means we can build an ABS model with several layers of environment. For example, in a communications-network simulation, we can define three layers of environment, such as a grid, to model all the cell stations that serve the agents’ mobile phones, a continuous space to model the physical location of agents and a network to represent the current active mobile communications between agents.

### 2.3 Behaviours

One of the main advantages of ABS is its ability to use agents’ behaviours and interactions to generate (and explain) a pattern or behaviour observed at the system level. The behaviour of an agent includes an update to its internal properties and an interaction with other agents or the environment. This behaviour is often implemented as a computer code. Some tools (including Repast) provide a facility to implement behaviour using a graphical notation, such as UML State Chart. The behaviour of individuals in the segregation model is implemented in the method `step` (we can change the name to another). The detailed code is shown below.

```java
01. @ScheduledMethod(start=0, interval=1)
02. public void step() {
03.   Context<Agent> context = (Context)ContextUtils.getContext(this);
04.   Grid<Agent> grid = (Grid)context.getProjection("Grid");
05.   boolean lookingForNewSite = true;
06.   while(lookingForNewSite) {
07.     VNQuery<Agent> query = new VNQuery<Agent>(grid, this);
08.     numLikeMe = 0;
09.     neighborCount = 0;
10.    for (Agent agent : query.query()) {
11.      if (agent.getType() == this.getType()) numLikeMe++;
12.      neighborCount++;
13.    }
14.    if (numLikeMe / neighborCount >= percentLikeNeighbors){
15.      lookingForNewSite = false;
16.    }
17.    else {
18.      int width = grid.getDimensions().getWidth();
19.      int height = grid.getDimensions().getHeight();
20.     int x = RandomHelper.nextIntFromTo(1, width-1);
21.     int y = RandomHelper.nextIntFromTo(1, height-1);
22.    while(grid.getObjectAt(x,y) != null) {
23.      x = RandomHelper.nextIntFromTo(1, width-1);
24.      y = RandomHelper.nextIntFromTo(1, height-1);
25.    }
26.     grid.moveTo(this, x,y);
27.    }
28.  }
29.  currentAge++;
30.  if (currentAge >= maxAge) this.die();
31. }
```

Line 03 retrieves the context of the individual. This context is used in line 04 to retrieve the grid projection. The loop from line 06 to 28 is repeated until the individual is happy with the state of his/her neighbourhood. Line 07 retrieves all the neighbours around the individual. The loop from line 10 to 13
counts the number of neighbours with the same colour. Line 14 checks if the proportion of neighbours with
the same colour is above a threshold value. If yes, the individual will stay in the current cell (line 15). Otherwise, the individual will move to a random cell that is empty (lines 18–26). Line 29 increases the age
of the individual. Finally, line 30 decides if the individual has to leave the context (i.e. die).

The code for the method die is shown below. We retrieve the context (line 02) so that the agent can be
removed from the context (line 03). A new agent is created using the same ID as the dead agent (line 04)
and added to the context (line 05). This method shows that an agent can be removed and added dynamically
during a simulation run.

```java
01. public void die() {
02.   Context<Agent> context = (Context)ContextUtils.getContext(this);
03.   context.remove(this);
04.   Agent child = new Agent(this.id);
05.   context.add(child);
06. }
```

### 2.4 Building the model

In Repast, an ABS model is implemented as a `Context` created by `ContextBuilder` using the method `build`. When we implement the method `build`, we can take the following actions: (1) create, initialise and add agents to the model, (2) define the environment where the agents live and (3) read the parameters needed for the simulation. The following code shows how we build the context for the segregation model.

Line 01 shows how we define a class that implements `ContextBuilder`. In the implementation, we need to overwrite the method `build` as shown in line 02. Lines 03–06 show an example of how to read parameters for the simulation. In this example, we read the number of agents, the world height and the world width. The values of these parameters are assigned to the variables `numAgents`, `h` and `w`, respectively. Line 07 adds a grid projection to the context. Hence, we are defining a grid environment for the agents in the context. The grid has wrap-around borders and when an agent is added to the grid, the agent will be put in a random cell that is empty (because the third parameter is false, one cell cannot accommodate more than one agent). The loop from line 08 to 11 adds a number of agents as specified in the simulation parameters. An agent is created from the agent type in line 09. The new agent is added to the context in line 10. Finally, the method `build` needs to return the model (i.e. the context) as shown in line 12. At this stage, we have a complete ABS model. Of course, we still need to specify how the agents and their environment will be displayed during runtime. Repast provides a wizard to define this, as explained in Repast’s ‘getting started’ document (it comes with the software). This example shows the typical actions that we take inside the method `build`, i.e. reading simulation parameters and adding the agents and environment to the model. It should be noted that we can add and remove agents from their context dynamically during a simulation. Hence, the location of the code to add agents is not limited to this method. A context in Repast has other features that may be useful for a more complex model. For example, a context can have sub-contexts to form a hierarchy of contexts.

```java
01. public class SchellingModel implements ContextBuilder<Object> {
02.   public Context<Object> build(Context<Object> context) {
03.     Parameters p = RunEnvironment.getInstance().getParameters();
04.     int numAgents = (Integer)p.getValue("initialNumAgents");
05.     int h = (Integer)p.getValue("worldHeight");
06.     int w = (Integer)p.getValue("worldWidth");
07.     GridFactoryFinder.createGridFactory(null).createGrid("Grid",
08.         context,
09.         new GridBuilderParameters<Object>(
10.             new WrapAroundBorders(), new RandomGridAdder<Object>(), false, w, h
11.         )
12.     );
13.     for(int i=0; i<numAgents; i++) {
14.         Agent agent = new Agent("Agent-"+i);
15.     }
```
2.5 How can ABS help in making better decisions?

Simulation-modelling paradigms such as DES and SD have been used to help people make better decisions. ABS can help people make better decisions in the same way as DES and SD. Macal and North (2010) provide a list of situations where ABS is useful. There is one situation in which ABS is particularly useful, i.e. when we need to explain how the interactions between independent decisions made by agents can lead to an observable pattern at the system level. For example, a policymaker may notice that there is some segregation in a city and try to explain why it happens. Based on her knowledge of the people living in the city, she provides a possible explanation. Her explanation is that if all the individuals in a city do not want to be a minority in their neighbourhood, the city will be segregated. In order to support this explanation, the above segregation model was built. When an experiment is done using the model, the model consistently generates segregation in the city. Hence, the model proves that her explanation is plausible. Of course, there might be other explanations. This is where empirical work is needed to test which explanation is better.

The above reasoning method is called abduction. We start with a result $Q$ (e.g. there is segregation in my city) and we construct an explanation or rule $P \rightarrow Q$ (if all individuals in a city do not want to be in a minority in their neighbourhood, the city will be segregated). An ABS model is developed to implement the rule and executed to see if the rule ($P \rightarrow Q$) can generate the result ($Q$). If it is true, then we can conclude that the case ($P$) is one possible explanation. This reasoning is not as strong as induction and deduction. However, this method is common when there is limited information available to us.

3 CHALLENGES FOR ABS AS A DECISION SUPPORT TOOL

This tutorial shows that ABS can be used as a decision-support tool. Despite its benefits, we can see that there are challenges that may hinder the use of ABS as a decision-support tool. I conclude this tutorial by highlighting three such challenges, namely, a good conceptual model representation, suitable validation methods and the need for participatory modelling.

3.1 Agent-based model representation

In simulation modelling, we select a certain portion of the real world system to be simulated for specific objectives. The process of capturing the essential elements of the system is referred to as conceptual modelling and the resulting model is referred to as a simulation conceptual model (Pidd 2004). In a typical simulation-modelling process, the conceptual-modelling stage is followed by computer implementation, validation, experimentation and implementation on a real system (Pidd 2004). There is no consensus in the literature on the definition of a conceptual model in simulation. Robinson (2008) discusses and compares a number of different definitions. At some stage in a simulation project, the conceptual model needs to be communicated to project stakeholders. Hence, the role of conceptual-model representation is crucial because it facilitates effective communication between stakeholders. Nance (1994) refers to conceptual-model representation for this purpose as a communicative model. Stakeholders may have different expertise in their own field but not in another one. Hence, communicating computer code for a model may not be a good idea if some stakeholders have limited or no computer-programming knowledge. For example, someone who has limited programming knowledge may find it difficult to understand the code presented in Section 2. Hence, a good conceptual-model representation for ABS is needed to help the communication between different stakeholders in a simulation project. Effective communication is essential for the success of a simulation project (Robinson and Pidd 1998).

The three commonly used simulation-modelling paradigms in operational research and management applications are DES, ABS and SD. While the conceptual-model representations in DES and SD have been
dominated by process-flow and stock-and-flow diagrams, respectively, research into conceptual-model representation in ABS is relatively new. Onggo (2013) discusses and compares five conceptual model-representation methods for ABS (i.e. pseudocode/flowchart, Petri Nets, DEVS, UML and BPMN). Although each representation method has its strengths and weaknesses, he argues that BPMN is more friendly for stakeholders who may not be familiar with computer programming or software-engineering concepts. Examples of how BPMN have been used as a communicative model for ABS can be found in Onggo and Karpat (2011), Onggo (2012) and Onggo et al. (2013). However, there remain questions that have not been addressed as to how effectively to represent an ABS model using BPMN. For example, the ability of BPMN to represent the complex attributes of an agent (such as memory, beliefs and perceptions), environment (spatial and relational) and complex learning behaviour (which may alter the rules/behaviour) is limited. Furthermore, work on the specification language needed to run an ABS model represented using BPMN is also limited. There is some early work by Onggo and Karpat (2011) in which an ABS model can be written using BPMN and a detailed specification can be added to the model using an agent-based simulation language (ABSL). The combination of BPMN and ABSL facilitates the simulation of an ABS model.

3.2 Validation
Stakeholders will need assurance that the ABS model that is used as a decision-support tool is valid before they make a decision based on the model. DES and ABS are typically used to represent stochastic dynamic systems and there is a need to track individuals’ states in the simulation. Given this similarity, the validation techniques commonly used in DES, such as face validity, operational validity, white-box validation and black-box validation, are also suitable for ABS. However, model validation in ABS is especially demanding. First, there is a need to validate ABS models at various levels (agent level, system level and, possibly, some intermediate levels). A second challenge arises due to the fact that we often need to represent behaviour in ABS using rules or algorithms. Hence, we need to check if the rules used in our ABS model represent the rules used by real-world agents. The heterogeneity of the agents makes this issue even more challenging. Finally, an ABS model often requires high-fidelity data, which are not always available. Hence, validation using empirical data may not always be possible. Given these challenges, research that creates validation techniques and tools for ABS (e.g. Duong 2010, Niazzì et al. 2009, Collier and Ozik 2013, Gurcan et al. 2013, Onggo et al. 2014, Onggo and Karatas 2015) is needed. These tools can help to increase the confidence of stakeholders when using ABS for decision-making. More research is needed in this area.

3.3 Facilitated ABS Modelling
A simulation project that involves many stakeholders with differing views and objectives is challenging. A good model representation can help to improve the communication between stakeholders. However, in the case of differences in stakeholders’ perspectives, the modelling process also needs to be designed to deal with the conflicting interests of multiple stakeholders. One possible approach is the group model-building that has been widely used in SD (Vennix 1999). The objective of group model-building is to build a socially-constructed description of an organization based on an understanding of multiple stakeholders (Andersen et al. 2007). There has been similar work in DES recently, e.g. facilitated modelling using DES (Tako and Kotiadis 2015) and SimLean (Robinson et al. 2012). As far as I know, a similar approach for ABS is still in the very early stages, e.g. in Newig et al. (2008). Facilitated modelling can help modellers to engage stakeholders during a simulation project and manage potentially conflicting interests from multiple stakeholders. Hence, research in this area will provide us with tools that can improve the involvement of stakeholders in the ABS model-building process and improve the credibility of the ABS model.
REFERENCES


**AUTHOR BIOGRAPHY**

**BHAKTI STEPHAN ONGGO** is a Lecturer (Assistant Professor) in the Department of Management Science at the Lancaster University Management School (LUMS), Lancaster, United Kingdom. He completed his PhD in Computer Science at the National University of Singapore and his MSc in Management Science is from Lancaster University. His research interests are in the areas of simulation methodology (conceptual modelling, discrete-event simulation, system dynamics and agent-based simulation) and simulation applications in business and management. His email address is s.onggo@lancaster.ac.uk.
CONCEPTUAL MODELLING FOR SIMULATION: A TUTORIAL

Prof. Stewart Robinson

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

ABSTRACT

Conceptual modelling is the abstraction of a simulation model from the part of the real world it is representing; in other words, choosing what to model, and what not to model. This is generally agreed to be the most difficult, least understood and most important task to be carried out in a simulation study. In this tutorial the problem of conceptual modelling is first illustrated through an example of modelling a hospital clinic. We define the term ‘conceptual model’ and go on to identify the artefacts of conceptual modelling and hence the role of conceptual modelling in the simulation project life-cycle. The discussion then focuses on the requirements of a conceptual model, the benefits and approaches for documenting a conceptual model, and frameworks for guiding the conceptual modelling activity. One specific framework is described and illustrated in more detail. The tutorial concludes with a discussion on the level of abstraction.

Keywords: Conceptual Modelling, Simulation

1 INTRODUCTION

One of the most difficult issues in simulation modelling is determining the content of the simulation model. The job of the modeller is to understand the real system that is the subject of the simulation study and to turn this into an appropriate simulation model. The chosen model could range from a very simple single server and queue, through to a model that tries to encapsulate every aspect of the system. In effect, there are an infinite number of models that could be selected within this range, each with a slightly, or even very, different content. The question is: which model should we choose? We explore the answer to this question in this tutorial.

On the surface we might suggest the answer is to build the model that contains as much detail as possible. After all, this model will be the closest to the real system and so surely the most accurate. This might be true if we had complete knowledge of the real system and a very large amount of time available to develop and run the model. But what if we only have limited knowledge of the real system and limited time? Indeed, we rarely have the luxury of vast quantities of either knowledge or time, not least because the real system rarely exists at the time of modelling (it is a proposed world) and a decision needs to be made according to a tight time schedule. Further, a simpler model is often sufficient to address the problem at hand and so there is no need for a more complex model.

So, if we need to develop a simpler model, we need to determine the level of abstraction at which to work. This process of abstracting a model from the real world is known as conceptual modelling. We shall define conceptual modelling, its requirements and the process of conceptual modelling in more detail in a while, but first it is useful to illustrate the issues involved in conceptual modelling with a practical example.
CONCEPTUAL MODELLING EXAMPLE: SIMULATING AN OUTPATIENTS CLINIC

Our simulation and modelling team was asked to develop a simulation model for a hospital in Birmingham, UK. The hospital was investing in a new outpatients building, a multi-million dollar project, and their key question was how many consultation rooms are required? They had performed some calculations based on expected patient flows and on observations of the current outpatients system. However, there was obviously some concern with making major investment decisions based on these limited data.

We were quick to point out the problems of making calculations based on static data which do not take into account the effects of variability in patient flows and consultation times. This is something for which discrete-event simulation is very well suited.

When asked to build a model such as this, the typical approach would be to start collecting data and to develop a detailed model of the system. However, the more we investigated how an outpatients system works the more we realized just how complex the system is. There are many specialties using the facility, each with its own clinical team. Patients can progress through a series of tests and consultations. For some specialties, such as ophthalmology, specialist equipment and dedicated rooms are required. Scheduling patient appointments is a significant task and then there is the matter of late arrivals and non-attendances. Staff shifts, working practices and skills all impact upon the functioning of the system.

Given appropriate data, it would be quite possible to build a simulation model that took account of all these details. There were, however, two issues that made such a model infeasible:

- **Lack of data**: much of the necessary data had not previously been collected and even if we were to try, issues of patient confidentiality (e.g. you cannot sit in a consultation room timing consultation times) would make it impossible to collect all the data we needed.
- **Lack of time**: the hospital required an answer within a few weeks and we had very limited time and resource to devote to the modelling work given the number of parallel activities in which we were engaged.

So what did we do? We focused on the critical issue of how many rooms were required and designed a simple model that would give at least an indication upon which the hospital managers could base a decision. Our world view was that the additional information a basic simulation could offer would be more beneficial than no simulation at all.

The simple model we constructed took a couple of days to build and experiment with. It provided a lower bound on the rooms required. In doing so it provided information that would give a greater level of confidence in making the decision that the hospital faced. This was all that was possible given the data and resource available, but it was still valuable.

The model we designed is outlined in Figure 1. Patient arrivals were based on the busiest period of the week – a Monday morning. All patients scheduled to arrive for each clinic, on a typical Monday, arrived into the model at the start of the simulation run, that is, 9.00am. For this model we were not concerned with waiting time, so it was not necessary to model when exactly a patient arrived, only the number that arrived.

A proportion of patients do not attend their allotted clinic. Typical proportions of patients that do not attend were sampled at the start of the simulation run and these were removed before entering the waiting line.

![Figure 1 Process flow diagram of the simple outpatients building model.](image-url)
Data on the time in a consultation room were limited, since they had not specifically been timed, but there were norms to which the clinical staff aimed to work. These data were available by clinic type and we used these as the mean of an Erlang-3 distribution to give an approximation for the variability in consultation time.

The input variable for the simulation experiments was the number of consultation rooms, which were varied from 20 to 60 in steps of 10. The main output variable was the time it took until the last patient left the system. A key simplification, which all involved recognized, was that there were no limitations on staff or equipment availability. Albeit extremely unlikely that this would be the case, the model was predicting a lower bound on the rooms required. In other words, shortages of staff and equipment would only increase the need for consultation rooms with patients waiting in the rooms while the resource became available.

For each room scenario the model was replicated 1000 times and a frequency chart was generated showing the probability that the system would be cleared in under 3 hours – the hospital’s target. Figure 2 shows an example of these results.

![Figure 2](image_url)

**Figure 2** Example of results from the outpatients building model: frequency distributions for time until last patient leaves.

This example illustrates the very essence of conceptual modelling; abstracting a model from the real system. In this case, the real system was not in existence, but it was a proposed system. The model involved simplifications such as modelling only Monday morning’s clinic and not modelling staff and equipment. It also involved assumptions about, among others, the consultation times. Because of the constraints on data and time, the conceptual model involved a great deal of simplification; as such, it might be described as a ‘far’ abstraction.

Whether we got the conceptual model right is in large measure a matter of opinion and one we will leave the reader to judge. It is certain that readers will form quite different judgments on the credibility of the model and so whether it was a good model or not.

### 3 WHAT IS CONCEPTUAL MODELLING?

Conceptual modelling is the abstraction of a simulation model from the part of the real world it is representing (‘the real system’). The real system may, or may not, currently exist. Abstraction implies the need for a simplified representation of the real system in the simulation model. The secret to good conceptual modelling is to get the level of simplification correct, that is, to abstract at the right level.
Because all models are simplifications of the real world, all simulation modelling involves conceptual modelling. Even the most complex and detailed simulation still makes various assumptions about the real world and chooses to ignore certain details.

### 3.1 Definition of a Conceptual Model

More formally we define a conceptual model as follows: ‘… 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 2008a)

Let us explore this definition in some more detail. First, this definition highlights the separation of the conceptual model from the computer model. The latter is software specific, that is, it represents the conceptual model in a specific computer code. The conceptual model is not specific to the software in which it is developed. It forms the foundation for developing the computer code.

Second, the conceptual model describes the computer simulation model. It describes how we conceive the model; it does not describe the real system. In other words, the conceptual model describes how we have abstracted the model away from our understanding of the real world. This distinction is important because of the need for model abstraction in simulation. Consider the model described in section 2. Our conception of this model (conceptual model) is very distinct (and ‘far’) from our description of the real world.

Third, it is stated that the description is of a computer simulation model that ‘that will be, is or has been developed.’ This serves to highlight the persistent nature of the conceptual model. It is not an artefact that gets created and is then dispensed with once the computer code has been written. It describes the concept of the computer model prior to development, during development and after development. Indeed, the conceptual model persists long beyond the end of the simulation study, since we cannot dispose of the model concept. Of course, because the modelling process is iterative in nature (Balci 1994; Willemain 1995; Robinson 2014), the conceptual model is continually subject to change throughout the life-cycle of a simulation study.

Finally, the definition is completed by a list of what a conceptual model describes. It is vital that the objectives of the model are known in forming the conceptual model. The model is designed for a specific purpose and without knowing this purpose it is impossible to create an appropriate simplification. Consider what would have happened if the purpose of the outpatients building model had not been properly understood. We would almost certainly have been driven to a more general purpose, and by nature much more complex, model. Poorly understood modelling objectives can lead to an overly complex model. Instead, because the purpose of the model was clear we were able to create a very simple model.

It is useful to know the model inputs and outputs prior to thinking about the content of the model. The inputs are the experimental factors that are altered in order to try and achieve the modelling objectives. In the example above, this was the number of consultation rooms in the outpatients building. The outputs are the reports that inform us as to whether the modelling objectives are being achieved (e.g. the time to clear all patients from the outpatient system) and if not, why they are not being achieved (e.g. the utilization of the consulting rooms).

Knowing the objectives, inputs and outputs of the model help inform the content of the model. In particular, the model must be able to receive the inputs (e.g. it must model the consultation rooms) and it must provide the outputs (e.g. it must model the flow of patients until all have exited the system). The model content can be thought of in terms of two dimensions:

- **The scope of the model**: the model boundary or the breadth of the real system that is to be included in the model.
- **The level of detail**: the detail to be included for each component in the model’s scope.

The final two items in the list of what a conceptual model describes are the assumptions and simplifications of the model. These are quite distinct concepts (Robinson 2008a):

- **Assumptions** are made either when there are uncertainties or beliefs about the real world being modelled.
Robinson

- **Simplifications** are incorporated in the model to enable more rapid model development and use, and to improve the transparency of the model.

So, assumptions are a facet of limited knowledge or presumptions, while simplifications are a facet of the desire to create simple models.

### 3.2 Artefacts of Conceptual Modelling

To understand conceptual modelling further it is useful to set it within the wider context of the modelling process for simulation. Figure 3 shows the key artefacts of conceptual modelling. The ‘cloud’ represents the real world (current or future) within which the problem situation resides; this is the problem that is the basis for the simulation study. The four rectangles represent specific artefacts of the (conceptual) modelling process. These are as follows:

- **System description**: a description of the problem situation and those elements of the real world that relate to the problem.
- **Conceptual model**: as defined in section 3.1
- **Model design**: the design of the constructs for the computer model (data, components, model execution, etc.) (Fishwick 1995).
- **Computer model**: a software specific representation of the conceptual model.

![Figure 3 The artefacts of conceptual modelling (Robinson 2011).](image)

These artefacts are quite separate. This is not to say that they are always explicitly expressed, with the exception of the computer model. For instance, the system description, conceptual model and model design may not be (fully) documented and can remain within the minds of the modeller and the problem owners. It is, of course, good modelling practice to document each of these artefacts and to use this as a means of communicating their content with the simulation project clients.

The model design and computer model are not strictly part of conceptual modelling, but they do embody the conceptual model within the design and code of the model. These artefacts are included in Figure 3 for completeness. Our main interest here is in the system description and conceptual model which make up the process of conceptual modelling; as represented by the shape with a dashed outline in Figure 3. Unlike the model design and computer model, these two artefacts are independent of the software that will ultimately be used for developing the simulation model.
It is important to recognize the distinction between the system description and the conceptual model. The system description relates to the problem domain, that is, it describes the problem and those elements of the real world that relate to the problem. The conceptual model belongs to the model domain in that it describes those parts of the system description that are included in the simulation model and at what level of detail. The author’s experience is that these two artefacts are often confused and seen as indistinct. Indeed, a major failure in any simulation project is to try and model the system description (i.e. everything that is known about the real system) and to not attempt any form of model abstraction; this leads to overly complex models.

The arrows in Figure 3 represent the flow of information, for instance, information about the real world feeds into the system description. The processes that drive the flow of information are described as knowledge acquisition, model abstraction, design and coding. The arrows are not specifically representative of the ordering of the steps within the modelling process, which we know are highly iterative (Balci 1994; Willemain 1995; Robinson 2014). In other words, a modeller may return to any of the four processes at any point in a simulation study, although there is some sense of ordering in that information from one artefact is required to feed the next artefact.

The specific and different roles of assumptions and simplifications are highlighted in Figure 3. Assumptions relate to knowledge acquisition, that is, they fill in the gaps in the knowledge that can be acquired about the real world. Meanwhile, simplifications relate to model abstraction, since they are deliberate choices to model the world more simply.

The dashed arrow in Figure 3 shows that there is a correspondence between the computer model and the real world. The degree of correspondence depends on the degree to which the model contains assumptions that are correct, the simplifications maintain the accuracy of the model, and the model design and computer code are free of errors. Because the model is developed for a specific purpose, the correspondence with the real world only relates to that specific purpose. In other words, the model is not a general model of the real world, but a simplified representation developed for a specific purpose. The issue of whether the level of correspondence between the model and the real world is sufficient is an issue of validation (Landry, Malouin, and Oral 1983; Balci 1994; Robinson 1999; Sargent 2013). Both conceptual modelling and validation are concerned with developing a simulation of sufficient accuracy for the purpose of the problem being addressed. As a result, there is a strong relationship between the two topics, conceptual modelling being concerned with developing an appropriate model and validation being concerned with whether the developed model is appropriate.

The artefacts described in this section are similar to Zeigler’s concepts of the real system, the experimental frame, the base model, the lumped model, and the computer. The interested reader is referred to Zeigler (1976).

4 REQUIREMENTS OF A CONCEPTUAL MODEL

Before discussing how to perform conceptual modelling, let us consider what makes for a good conceptual model. The key requirements are that the model should be valid, credible, feasible and have utility (Robinson 2008a). By these we mean the model should:

- Produce sufficiently accurate results for the purpose: understanding the number of rooms required in the building (validity);
- Be believed by the clients (credibility);
- Be feasible to build within the constraints of the available data and time;
- Have utility, that is, sufficiently easy to use, flexible, visual and quick to run.

Overarching all of this is the requirement to build the simplest model possible to meet the objectives of the simulation study. According to Innis and Rexstad (1983), Ward, (1989), Salt (1993), Chwif, Barreto, and Paul (2000), Lucas and McGunnigle (2003), and Thomas and Charpentier (2005), simpler models are preferred because:

- Simple models can be developed faster;
Simple models are more flexible;
- Simple models require less data;
- Simple models run faster;
- The results are easier to interpret since the structure of the model is better understood.

As such, the need to abstract a conceptual model from the system description becomes even more pertinent. This does not, of course, mean that we should never develop more complex models, but that we should only develop them if they are required to meet the modelling objectives. For further discussion on the topic of model complexity, with respect to how a model is used, see Pidd (2010).

Figure 4 illustrates the relationship between model accuracy and model complexity (scope and level of detail). It shows that with increasing levels of complexity we obtaining diminishing returns in terms of accuracy, never reaching 100% accuracy. Eventually we may even find that the accuracy of the model reduces. This is because we do not have the knowledge or data to support the complexity that is being included in the model and we start to make assumptions that are incorrect.

So which conceptual model should we choose? We might argue that the model at point x in Figure 4 is the best. At this point we have gained a high level of accuracy for a low level of complexity. Moving beyond x will only marginally increase accuracy and adding further complexity generally requires ever increasing effort. Of course, if we have a specific need for an accuracy level greater than that provided by x, we will need to increase the complexity of the model. Indeed, Sargent (2013) suggests that a model’s ‘acceptable range of accuracy’ should be determined early in the modelling process so it can guide both model development and model validation.

The difficulty is in finding point x. Conceptual modelling frameworks, such as the ones listed below, aim to help us in that quest, but conceptual modelling is more of an art than a science (we might prefer to use the word ‘craft’). As a result, we can only really hope to get close to x. In other words, there may be a ‘best’ model, but we are extremely unlikely to find it among an infinite set of models. What we should hope to do is identify the best model we can. As such, our quest is for better models, not necessarily the best.

5 DOCUMENTING THE CONCEPTUAL MODEL

As stated above, the conceptual model is not always explicitly expressed, but can remain within the mind of the modeller. That said, it is good practice to document the conceptual model and in so doing to provide a means of communication between all parties in a simulation study (e.g. the modeller, code developers, domain experts, end users and clients). In so doing it helps to build a consensus, or least an accommodation, about the nature of the model and its use. A documented conceptual model:

- Minimizes the likelihood of incomplete, unclear, inconsistent and wrong requirements
Robinson

- Helps build the credibility of the model
- Guides the development of the computer model
- Forms the basis for model verification and guides model validation
- Guides experimentation by expressing the modelling objectives, and model inputs and outputs
- Provides the basis of the model documentation
- Can act as an aid to independent verification and validation when it is required
- Helps determine the appropriateness of the model or its parts for model reuse and distributed simulation

There are no set standards for documenting discrete-event simulation conceptual models, but a range of approaches have been proposed, including (references to examples are provided in brackets):

- Component list
- Process flow diagram (Robinson 2014)
- Activity cycle diagram (Robinson 2014)
- Logic flow diagram (Robinson 2014)
- List of assumptions and simplifications
- Unified modelling language (UML) (Richter and März 2000)
- Petri nets (Torn 1981)
- Condition specification (Overstreet and Nance 1985)

The documentation for a conceptual model should be kept simple and focus on identifying what is to be modelled and what is not to be modelled. There is no need for elaborate documentation because detailed decisions about how to model something are not being taken during conceptual modelling. These decisions are taken while creating the model design from the conceptual model. For example, in conceptual modelling a decision is taken to model a flight schedule; in model design the way in which that flight schedule is to be modelled is determined. Hence, the conceptual model documentation only needs to state that the flight schedule is to be modelled, while the model design documentation needs to provide the detail of how to model the flight schedule.

6 FRAMEWORKS FOR CONCEPTUAL MODELLING

A framework for conceptual modelling provides a set of steps and tools that guide a modeller through the development of a conceptual model. It is also useful for teaching conceptual modelling, especially to novice modellers. The simulation literature, however, provides very few such frameworks. Some examples, that the reader may wish to explore further are:

- Conceptual modelling framework for manufacturing (van der Zee 2007)
- The ABCmod conceptual modelling framework (Arbez and Birta 2011)
- Karagöz and Demirörs (2011) present a number of conceptual modelling frameworks: Conceptual Model Development Tool (KAMA), Federation Development and Execution Process (FEDEP), Conceptual Models of the Mission Space (CMMS), Defense Conceptual Modelling Framework (DCMF), and Base Object Model (BOM)
- Conceptual modelling with Onto-UML (Guizzardi and Wagner 2012)
- The PartiSim framework (Tako and Kotiadis 2015)

For a more detailed discussion on conceptual modelling frameworks see Robinson et al. (2011). Here a very brief outline and illustration of Robinson’s framework for conceptual modelling is given. For a more detailed account, and an illustration of the framework in use, see Robinson (2008b).
Figure 5 outlines Robinson’s conceptual modelling framework. In this framework, conceptual modelling involves five activities that are performed roughly in this order:

- Understanding the problem situation
- Determining the modelling and general project objectives
- Identifying the model outputs (responses)
- Identifying the model inputs (experimental factors)
- Determining the model content (scope and level of detail), identifying any assumptions and simplifications

Starting with an understanding of the problem situation, a set of modelling and general project objectives are determined. These objectives then drive the derivation of the conceptual model, first by defining the outputs (responses) of the model, then the inputs (experimental factors), and finally the model content in terms of its scope and level of detail. Assumptions and simplifications are identified throughout this process.

The ordering of the activities described above is not strict. Indeed, we would expect much iteration between these activities and with the other activities involved in a simulation study: data collection and analysis, model coding, verification and validation, experimentation and implementation.

The framework is supported by a conceptual model template which provides a set of tables that describe each element of the conceptual model. These tables describe:

- Modelling and general project objectives (organisational aim, modelling objectives, general project objectives)
- Model outputs/responses (outputs to determine achievement of objectives, outputs to determine reasons for failure to meet objectives)
- Experimental factors
- Model scope
- Model level of detail
- Modelling assumptions
- Model simplifications

Beyond completing these tables, it is also useful to provide a diagram of the model. For instance, process flow diagrams, similar to that presented in Figure 1, are useful for communicating the conceptual model.
The modeller works through these tables with the support of the stakeholders and domain experts, iteratively improving them to the point that the modeller and stakeholders are satisfied that the conceptual model meets the requirements for validity, credibility, feasibility and utility. This provides a structured framework for making the conceptual modelling decisions explicit (documentation) and for debating ways of improving the conceptual model.

7 LEVELS OF ABSTRACTION

The conceptual modelling example in section 2 is described as a ‘far’ abstraction. By this we mean that the conceptual model involves many simplifications and so it is removed a long way from the system description. The implication of this is that the computer model is a highly simplified representation of the real world. At the extreme, a far abstraction can lead to a (conceptual) model that bears little resemblance to the real world. A good illustration of this is Schelling’s model of segregation, which contains no real world data and only the simplest possible representation of the phenomena under study (Schelling 1971). Although a far abstraction, Schelling’s model has certainly attracted a lot of attention.

However, we would not want to leave the impression that conceptual models have to be so far abstracted. Indeed it is not always desirable to abstract to this degree and for some simulation studies it is appropriate to model much of the scope and detail in the problem domain. We refer to this as ‘near’ abstraction. For an example, see the Ford engine plant model described in Robinson (2008a, 2008b). These papers describe a simulation that was designed to determine the throughput of a new engine assembly plant. The model contained much detail about the real system and took a considerable time to develop.

The level of abstraction should be determined by the requirement for the model to be valid, credible, feasible and have utility. One danger with far abstraction is that whilst the model may be valid, it may lack credibility. Hence, we may need to reduce the level of abstraction, making the model nearer to the system description, to increase the credibility of the model.

8 CONCLUSION

Conceptual modelling is the abstraction of a simulation model from the part of the real world it is representing. It is probably the most important aspect of any simulation study. Get the conceptual model right and the rest of the simulation work will be more straightforward, providing the right information in the right time-scale.

This tutorial provides an illustration of how appropriate conceptual modelling, through far abstraction, made a simulation study feasible within the constraints of data and time available. The discussion that follows defines conceptual modelling, its artefacts and its requirements; it also discusses the benefits and approaches to documenting the conceptual model. From this base, some frameworks for conceptual modelling are listed and Robinson’s framework is outlined in more detail. The framework aims to guide a modeller through the process of creating and documenting a conceptual model. We also discuss levels of abstraction, from near to far.

Conceptual modelling is not a science, but a craft or even an art. As with any craft, it can be learned and it can be improved upon with experience. Frameworks provide a good way of learning about conceptual modelling and for helping to do it better. At present, however, there are very few examples of conceptual modelling frameworks and this is an area where more research needs to be undertaken.

ACKNOWLEDGEMENTS

REFERENCES


AUTHOR BIOGRAPHY

STEWART ROBINSON is Professor of Management Science and Acting Dean 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 five 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 Simulation Workshop conference series. He was President of the Operational Research Society (2014-15). Home page: www.stewartrobinson.co.uk.
INCLUDING PARAMETER UNCERTAINTY IN SIMULATION EXPERIMENTS

Russell C.H. Cheng
University of Southampton
Highfield
cheng@btinternet.com

ABSTRACT

A simulation model usually depends on parameters whose values will affect its performance. For example in a single server queue, the queue length and customer waiting times depend on the customer arrival and server service rate parameters. Simulation experiments involving such a model often assume fixed, given values of these parameters. The only uncertainty is then the simulation uncertainty which is just the random variation that is built into the simulation model itself and which occurs when the model is run. Where there is uncertainty concerning parameter values, then this parameter uncertainty also has to be taken into account in analysing simulation results. This tutorial show how to allow for both forms of uncertainty in simulation experimentation.

Keywords: Sample Document, Guidelines

1 INTRODUCTION

We consider a simulation experiment made up of identical but statistically independent runs of a model where the output of interest, denoted by $Y$, is an average. There are two sources of variability in $Y$. One is due to simulation uncertainty which comes from the random variation that is built into the simulation model itself. The other is due to parameter uncertainty when parameters on which system behaviour depends are not known and have to be estimated. We examine how to take account of both sources of uncertainty. This problem has been considered by Cheng and Holland (1997, 1998, 2004) and more recently by Lin, Song and Nelson (2015) who considered how the asymptotic method discussed in Cheng and Holland (1997) might be refined using a single-experiment approach.

In this tutorial we outline the basic asymptotic and bootstrap (BS) methods given in Cheng and Holland (1997), focussing on the latter with a detailed numerical example. The reader is left to examine the original papers cited above for details of extensions to these two basic methods.

The above approaches are based on a frequentist viewpoint. A natural alternative is to use a Bayesian formulation of input uncertainty. This has the advantage of enabling expert opinion to be incorporated into the formulation and this will also be discussed briefly.

2 A METAMODEL FOR PARAMETER AND SIMULATION UNCERTAINTY

A convenient way of analysing the output from a simulation model is to use a statistical metamodel. We therefore need to be clear what is meant by a metamodel and this is discussed first.

Figure 1 illustrates the situation where we have data, $Y$ (here and throughout this article, a quantity is written in bold to indicate that it is a vector quantity), available concerning the behaviour of a system under study. The system itself, represented by the box in the middle, will typically be complicated or even not fully known. We call $Y$ the output and this is what we wish to analyse.
We also have *input* quantities, whose values are expected to influence the output. The inputs are divided into three types. The input $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a vector of $k$ explanatory variables. These are known quantities and indeed may possibly be under the control of the investigator. The input $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ is a vector of parameters that influences the output but whose values are not controllable. Often they will be unknown. Their values would therefore have to be *estimated*. We denote the estimate of $\theta$ by $\hat{\theta}$. Both possibilities are indicated in the Figure.

In addition, the output $Y$ will be stochastic, due to randomness arising in the system itself. We shall think of this randomness as being generated by a sequence of uniform $U(0,1)$ random variables, in the figure denoted by $U$. We regard the $U$ as being internally generated by the system so that it is actually part of the system. However we have depicted it separately in the figure to emphasize that it can be treated as an input, with the effect of making the output $Y$ stochastic.

Our objective is to analyse $Y$ and in particular to identify how the inputs $\lambda$ and $\theta$ and the random input $U$ influence $Y$. We use a *statistical model* for doing this. This does not have to model the specific workings of the real system, but represents the statistical behaviour of the entire system including all the features just described with particular emphasis on $Y$. This statistical model is shown in Figure 1 by a dashed box. We have also placed the entire system being studied in a dashed box to emphasize that the statistical model will represent the statistical aspects of all of it.

Note that the statistical model can represent *either* the output $Y$ of the real system (coloured red in the figure) or the output of the simulation model (coloured blue). In the second case the statistical model is still a statistical model, but is then often called a statistical *metamodel* to emphasise that it is a (statistical) model of output from a (simulation) model. It will be clear that whatever form the statistical model takes, it is determined purely by the form of the output $Y$ and not by its origin. Thus the *same* statistical model can be applied whether to output came from a real system or a simulation model.
Note that the statistical model can range from being simple to being very complex. Indeed if it represents the system in sufficient detail it can be the simulation model. However this would defeat the purpose of having a statistical model, the idea behind which is that it should be capable of representing the stochastic behaviour of \( Y \) of interest, but be sufficiently simple to analyse using standard statistical techniques.

3 THE STATISTICAL MODEL

We discuss the second scenario depicted in Figure 1, where we have a computer simulation model of a real system and a statistical metamodel of the simulation model. We consider in particular the long run average scenario discussed by Cheng and Holland (1997) who assume the quantity of interest is an average obtained over the run length which moreover is strongly dependent on quantities that also are averages over the run length. We follow the notation used in that paper.

The simulation experiment is assumed to comprise making a number, \( r \) say, of identical but statistically independent runs of the simulation model with each run carried out at the same setting of the input variables \( \theta, \lambda \). Then

\[
Y_j = y(\lambda, \theta, U_j) \quad j = 1, 2, ..., r, \tag{1}
\]

with

\[
E(Y_j) = \eta(\lambda, \theta), \tag{2}
\]

and then, under the long term average scenario, assuming \( Y \) is an average over the run length we can assume

\[
\text{Var}(Y_j) = \tau^2(\lambda, \theta)/l. \tag{3}
\]

The overall average

\[
\bar{Y} = \sum_{j=1}^{r} Y_j
\]

has mean and variance

\[
E(\bar{Y}) = \eta(\lambda, \theta), \quad \text{Var}(\bar{Y}) = \tau^2(\lambda, \theta)/rl. \tag{4}
\]

Under the long term average scenario, when quantities of interest are asymptotically normally distributed by the central limit theorem, a not unreasonable first approximation for the distribution of \( Y \) is

\[
Y_i \sim N(\eta(\lambda, \theta), \tau^2(\lambda, \theta)/l). \tag{5}
\]

This is useful for obtaining asymptotic results, but as will become clear is not so important using bootstrap analysis.

In the situation where \( \theta \) is known then little more is needed. We have

\[
\text{Var}(\bar{Y}) = S^2 / r \quad \text{where} \quad S^2 = \sum_{j=1}^{r} (Y_j - \bar{Y})^2 / (r-1). \tag{6}
\]

From this the usual \((1-\alpha)\)100% confidence interval can be calculated

\[
(\bar{Y} - z_{\alpha/2}S/\sqrt{r}, \quad \bar{Y} + z_{\alpha/2}S/\sqrt{r})
\]

where \( z_{\alpha/2} \) is the upper \( \alpha/2 \) quantile of the standard normal distribution.

For \( r \) sufficiently large it is probably more robust to take the empirical distribution function of the \( Y_i \) values and calculate the percentile confidence interval, discussed for example by Hjorth (1994), from this:
(Y(i), \bar{Y}(n)), \quad [l = (\alpha / 2)r], \quad [h = (1 - \alpha / 2)r].

(7)

where the \( Y(i), i = 1, 2, \ldots, r \) are the rank ordered \( Y_i \).

4 PARAMETER UNCERTAINTY

We now consider the situation where \( \theta \) is not known and has to be estimated. It may be that some components are known and others not. In this case what follows applies just to the components that are unknown, but to avoid unnecessary complication in the notation we simply assume that all components are unknown.

The question of interest is how the uncertainty arising from having to estimate \( \theta \) will affect the variability of \( \bar{Y} \). We use asymptotic theory to answer this. Assume that there is a true value \( \theta^0 \) but that this is unknown. Now assume that we have \( m \) data samples

\[ x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \quad i = 1, 2, \ldots, m \]

where each \( x_i \) is a random sample from a probability distribution which for simplicity we assume is continuous with probability density function (PDF) \( f_i(x, \theta) \), on which \( \theta \) depends. Let

\[ X = (x_1, x_2, \ldots, x_m). \]

(8)

The loglikelihood is defined as

\[ L(\theta) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \ln f_i(x_j, \theta), \]

and the maximum likelihood (ML) estimate is defined as

\[ \hat{\theta} = \arg\max_\theta L(\theta). \]

To make use of standard asymptotic theory, see for example Young and Smith (2005), we assume that \( n_i = \alpha_i n \) with \( \alpha_i > 0 \) for all \( i \). This allows asymptotic results to be given in the form \( n \to \infty \) instead of \( n_i \) individually.

The following result given by Cheng and Holland (1997) shows that the variability in \( \bar{Y} \) is composed of two distinct parts, one involving just the simulation uncertainty the other involving just the parameter uncertainty.

**Theorem 1:** Suppose \( \eta(\lambda, \theta) \) and \( \tau^2(\lambda, \theta) \) are both twice continuously differentiable in \( \theta \) in an open neighborhood of \( \theta^0 \). Then

\[ Var(\bar{Y}) = g^T(\lambda, \theta^0)V(\lambda, \theta^0)g(\lambda, \theta^0) + \tau^2(\lambda, \theta^0)/(rl) + R \]

\[ = \sigma^2(\lambda, \theta^0)/n + \tau^2(\lambda, \theta^0)/(rl) + R, \quad \text{say}, \]

\[ = O(n^{-1}) + O((rl)^{-1}) + R, \quad (9) \]

where \( g(\lambda, \theta) = \partial \eta(\lambda, \theta)/\partial \theta \) and \( V(\lambda, \theta) \) is the variance-covariance matrix of \( \hat{\theta} \), and \( R = O_p(n^{-3/2}) + O_p[(nrl)^{-1}] \). Thus \( R \) becomes negligible as \( n \) and \( rl \to \infty \), with \( \sigma^2(\lambda, \theta^0)/n \) the variance due to parameter
uncertainty and $\tau^2(\lambda, \theta^0)/(rl)$ the variance due to simulation uncertainty. The theorem shows that there is no point estimating $\theta^0$ too accurately on its own or in making a large number of simulation runs on its own, as the other uncertainty would then dominate.

Here and in what follows we write derivatives as $g(\lambda, \theta^0)$ simply to mean the derivative is evaluated at $\theta = \theta^0$. Asymptotically as $n \to \infty$ we have

$$V(\lambda, \theta^0) = -[\partial^2 L(\lambda, \theta^0)/\partial \theta^2]^{-1} \equiv -[\partial^2 L(\lambda, \hat{\theta})/\partial \theta^2]^{-1} = I^{-1}(\lambda, \hat{\theta}),$$  \hspace{1cm} (10)

so that use of $\hat{\theta}$ for $\theta^0$ is an acceptable substitution.

The formula (9) can be used to estimate $Var(\bar{Y})$ but this is not straightforward as the derivative $g(\lambda, \theta) = \partial \eta(\lambda, \theta)/\partial \theta$ is needed, so that if $\eta(\lambda, \theta)$ is not known it has to be estimated. But note that an alternative has recently been suggested by Lin et al. (2015) that simplifies estimation of $g(\lambda, \theta)$. Additionally Cheng and Holland (1997) point out two difficulties when $p$ is large, or more precisely when the unknown number of parameters is large. These are: (i) the calculation of $g(\lambda, \theta)$ is expensive as it has to be done numerically using simulation runs to estimate $\eta(\lambda, \theta)$ at different $\theta$ so that finite difference formulas can be used for the derivatives; (ii) if many of the components of $g(\lambda, \theta^0)$ are zero or nearly so, sampling error may inflate their estimated values resulting in the estimate of $\sigma^2(\lambda, \theta)$ being spuriously overlarge.

Bootstrapping provides an alternative way of calculating $Var(\bar{Y})$ that avoids having to use (9). We describe this next.

5 Bootstrap Method

Parameter uncertainty is due to the stochastic variability of $\hat{\theta}$. Imagine therefore that we had $B$ samples $X^{(b)} b = 1, 2, ..., B$, with each sample distributed exactly as $X$ in (8) but with the samples mutually independent. We could then carry out the entire simulation experiment $B$ times, with $\hat{\theta} = \hat{\theta}^{(b)}$ calculated from the sample $X^{(b)}$ in the $b$th experiment. By the Glivenko-Cantelli lemma the empirical distribution function (EDF) of the $\hat{\theta}^{(b)} b = 1, 2, ..., B$ converges to the population distribution of $\hat{\theta}$ with probability one, and it follows that the EDF of the corresponding $\bar{Y} = \bar{Y}^{(b)} b = 1, 2, ..., B$ converges to the distribution function of $\bar{Y}$.

We do not have $B$ such samples $X^{(b)}$ of course. However we can use the parametric bootstrap, see Hjorth (1994) to obtain $B$ bootstrap (BS) samples. Thus the $b$th BS sample is

$$X^{*(b)} = (x^{*(b)}_1, x^{*(b)}_2, ..., x^{*(b)}_k)$$

with

$$x^{*(b)}_i = (x^{*(b)}_{i1}, x^{*(b)}_{i2}, ..., x^{*(b)}_{im}) \quad \text{where} \quad x^{*(b)}_{ij} \sim F_i(x, \hat{\theta}) \quad j = 1, 2, ..., m. \quad$$

The BS estimate $\hat{\theta}^{*(b)}$ used in the $b$th simulation experiment is then obtained by applying the ML method to $X^{*(b)}$. We denote the output from the $b$th experiment by $\bar{Y}^{*(b)}$. We have the following

**Theorem 2** (Cheng and Holland, 1997):

$$Var(\bar{Y}^{*(b)}) = \sigma^2(\lambda, \theta^0)/n + \tau^2(\lambda, \theta^0)/(rl) + R$$  \hspace{1cm} (11)

where
Thus bootstrapping gives essentially the same result (9) as asymptotic theory.
A simple approximation for \( n \) sufficiently large is to sample the \( \hat{\theta}^{* (b)} \) directly from its asymptotic normal distribution using (10) so that:

\[
\hat{\theta}^{* (b)} \sim N(\hat{\theta}, I^{-1}(\lambda, \hat{\lambda})).
\]

However usually the simulation runs are computationally the most expensive part of the whole experiment, so the saving in obtaining \( \hat{\theta}^{* (b)} \) in this way is not great.

### 6 TOLL BOOTH EXAMPLE

We give a numerical example of the bootstrap methodology discussed.

In a study made of the operation of toll booths of the old Severn River bridge, see Griffiths and Williams (1984), each toll booth was modelled as a single server queue and data was collected of the service time of vehicles, that is the time taken for a vehicle to pay at the toll booth before crossing the bridge. Suppose each run of a simulation model of the queue simulates the service of \( l \) vehicles. Suppose the output of interest is the average vehicle waiting time in the queue in a run. We denote this by \( W \) rather than \( Y \). This depends on the arrival pattern and the service time distribution. Usually arrivals follow a Poisson process so that interarrival times can be treated as exponentially distributed with rate \( \lambda \). In the simulation we can use uniform random numbers \( U_{ij} \) \( j = 1,2, ..., l \) to generate the interarrival times in the \( i \)th simulation run using the inverse distribution function transform (IDFT) method. Thus, for exponentially distributed interarrival times, we would use

\[
A_{ij} = -\lambda^{-1} \ln(U_{ij}) \quad j = 1,2,\ldots
\]

We shall assume that the service time times follow the gamma distribution with PDF

\[
g(x, \alpha, \beta) = \frac{x^{\alpha-1} \exp[-(x/\beta)]}{\Gamma(\alpha)\beta^\alpha} \quad x > 0, \quad \alpha, \beta > 0.
\]

Using the GB acceptance/rejection method suggest by Cheng (1977) in our Excel implementation of the simulation experiment being discussed. Excel provides an IDFT function subroutine for generating gamma variates which we did try but it is very slow.

Suppose we are interested in how \( W \) varies with \( \lambda \). We therefore treat \( \lambda \) as being a design variable taking values in a given range \( \overline{\lambda} \leq \lambda \leq \overline{\lambda} \). We treat the parameters \( \alpha \) and \( \beta \) of the gamma distribution as making up the parameter \( \theta \).

In this example \( W \) is likely to depend strongly on the sample average values of interarrival times and services times so that the normality assumption (5) seems a reasonable initial approximation.

Griffiths and Williams (1984) discuss the collection of the service time data, that is the time taken for a vehicle to pay at the toll booth before crossing the bridge. We take typical sample, and as it is for illustration only, of small size \( N = 47 \). The observations are in seconds and given in Table 1.

| Table 1: 47 Toll Booth Service Times | 30 |
We denote this original sample by $x$. The ML estimators obtained from this sample are $\hat{\alpha} = 9.33$ and $\hat{\beta} = 0.644$ so that the mean service rate is $\hat{\alpha}\hat{\beta} = 6.01$ seconds. For this M/G/1 queue we actually know from the Pollacyek-Khinchin formula that the average waiting time of a vehicle in the queue is

$$W(\lambda; \alpha, \beta) = \frac{\lambda(1+\alpha)\alpha\beta^2}{2(1-\alpha\beta\lambda)}, \quad \underline{\lambda} \leq \lambda \leq \bar{\lambda}$$

(14)

where, as indicated in (13), we are interested in calculating $W$ over a range of $\lambda$ values.

We consider three scenarios for calculating a confidence band which, with given confidence say 90%, will entirely contain the graph of $W(\lambda; \alpha^0, \beta^0)$, where $\alpha^0, \beta^0$ are unknown, over the whole interval $\underline{\lambda} \leq \lambda \leq \bar{\lambda}$. Cheng (2015) shows how bootstrapping provides a simple way of doing this. Restrictions on paper length prevents details being presented here, but these are given in Cheng (2015).

Case 1. Confidence Band for $W$ based on Parameter Uncertainty Only. Simulation runs are not used in this case. The calculation is based on asymptotic theory only. We assume the mathematical form for $W(\lambda; \alpha, \beta)$ is (14) using this with $\hat{\alpha} = 9.33$ and $\hat{\beta} = 0.644$ to calculate $W(\lambda, \hat{\alpha}, \hat{\beta})$ for a range of values of arrival rates of interest. We took $\lambda = \lambda_m$, $m=1,2,...,M$ covering the range $\underline{\lambda} \leq \lambda \leq \bar{\lambda}$. To calculate the variance of $W(\lambda, \hat{\alpha}, \hat{\beta})$ at each $\lambda_m$, we use the expression $\sigma^2(\lambda, \hat{\theta})$ as given in (9) but with $\hat{\theta}$ instead of $\theta^0$. The resulting confidence band is therefore based on asymptotic theory therefore. Figure 2 shows the band for this case.

Case 2. Confidence Band for $W$ based on Simulation Uncertainty Only. The form of $W(\lambda; \alpha, \beta)$ is assumed to be unknown, so that (14) is not used. Instead we carry out a set of $B$ independent simulation experiments. In each experiment we make a set of $M = 10$ runs of a simulation model of the toll booth queue. The runs are identical, simulating the processing of $l = 10,000$ vehicles, except that we vary the arrival rate between runs using equally spaced values $\lambda_m$, $m=1,2,...,10$, in the range $\underline{\lambda} = 0 < \lambda < 0.1$ with the corresponding outputs $\{W^{(b)}(\lambda_1), W^{(b)}(\lambda_2),..., W^{(b)}(\lambda_{10})\} = W^{(b)}$ forming the results of the $b$th experiment. In all the runs of all the experiments we generate service times that are gamma distributed with PDF (12), using $\alpha = \hat{\alpha} = 9.33$ and $\beta = \hat{\beta} = 0.644$ for the parameter values throughout. A point to note is that it is best to generate the interarrival and service times using two different sequences of random numbers, one for the interarrival stream and one for the service time stream, but to use these same two streams in generating the interarrival and service times in each of the runs of a given experiment. This positively correlates the outputs $W^{(b)}(\lambda_i)$ of the runs in the same experiment reducing the statistical variability between the outputs of the different runs.

If we treat the points $\lambda_i, W^{(b)}(\lambda_i)$, $(\lambda_2, W^{(b)}(\lambda_2))$, ..., $(\lambda_M, W^{(b)}(\lambda_M))$ as being estimates of points on the graph of $W$ then the variation is due only to simulation uncertainty. We can then use the bootstrap
percentile method to calculate the confidence band in this case with the variation in the $W^{(b)}$ as $b$ varies due only to the simulation uncertainty. This is depicted in Figure 3.

Figure 2 90% Confidence Band for the Long-term Waiting Time in the Toll Booth Example Assuming Parameter Uncertainty through Using ML Estimates for Unknown True Parameter Values in the Gamma Service Time Distribution.

Figure 3 90% Confidence Band for the Long-term Waiting Time in the Toll Booth Example. The Dashed Line Band is that Assuming Simulation Uncertainty Only. The Solid Line Band Includes both Simulation Uncertainty and the Parameter Uncertainty through Using ML Estimates for Unknown True Parameter Values in the Gamma Service Time Distribution.
Case 3. Confidence Band for $W$ based on both Parameter and Simulation Uncertainty. We conduct $B$ experiments exactly as in Case 2 except that we generate service times that are gamma distributed with PDF (13), using estimates of the parameters $\alpha, \beta$ that change from experiment to experiment.

Thus in the $b$th experiment we use $\hat{\alpha}^{*(b)}$ and $\hat{\beta}^{*(b)}$ for the parameters of the gamma distributed service times, however this is for all the runs of the experiment. These parameter values are the ML estimates obtained from a different parametric bootstrap sample $x^{*(b)}$ for each experiment. These BS samples are identical in form to the original sample $x$, but are mutually independent. The only difference is that each observation making up $x^{*(b)}$ is drawn from the gamma distribution with PDF $g(x, \hat{\alpha}, \hat{\beta})$ rather than $g(x, \alpha^0, \beta^0)$. The estimate the parameters by ML from this sample are the parameter values $\hat{\alpha}^{*(b)}, \hat{\beta}^{*(b)}$.

Each of the $B$ simulation experiments has the same form as in Case 2, comprising $M (= 10)$ simulation runs at different arrival rates $\lambda = \lambda_m$, $m = 1, 2, ..., M$. However in the runs of the $b$th experiment we generate service times with gamma distribution where the parameter values are $\hat{\alpha}^{*(b)}, \hat{\beta}^{*(b)}$, yielding a set of bootstrap outputs $\{W^{*(b)}(\lambda_1), W^{*(b)}(\lambda_2), ..., W^{*(b)}(\lambda_M)\} = W^{*(b)}$. As in Case 2 it is best to generate the interarrival time and service time sequences using the same uniform random sequences in all the runs of a given experiment to minimize the random variation between the outputs of the experiment.

This last Case is where Theorem 2 applies. We can therefore calculate a BS confidence band exactly as in Case 2, only now the $W^{*(b)}$ include the parameter uncertainty arising from estimating $\alpha, \beta$, as well as simulation uncertainty. Figure 3 also shows the effect of this.

It will be seen that the effect of parameter uncertainty is very marked, this probably being due to the small size of the sample used in estimating $\alpha, \beta$.

7 DISCUSSION

An Excel workbook is available from the author’s university personal website. This carries all the analyses discussed in this paper using VBA subroutines. The entire analysis is carried out by just clicking on appropriate buttons. The routines are easily modifiable to handle similar examples.

It has to be said that the bootstrap method is computationally intensive. The Case 3 toll booth experiment comprised $M \times B = 10 \times 500$ simulation runs each of $l = 10,000$ so that $5 \times 10^7$ vehicles are processed in each overall experiment. The most expensive aspect is actually generation of the gamma variates. The VBA function WorksheetFunction.GammaInv() appears to be an inverse distribution function transform routine, but it is very slow, so that the bootstrap analysis of the toll booth example took some hours to carry out. Using the GB method of Cheng (1997) to generate the gamma variates speeds up the entire calculation nearly 30-fold, reducing the calculation time to just a few minutes.

A Bayesian analysis can be carried out using the Case 3 scenario of Section 5. Thus in the toll booth example, one would have to decide on a prior distribution for each of the gamma distribution parameters. A posterior distribution can then be calculated using the service time data of Table 1. Then the $B$ ‘bootstrap’ experiments are carried out but with the gamma distribution parameters $\alpha^{*(b)}$ and $\beta^{*(b)}$ sampled from the posterior distribution. If no data is available, one simply samples from the prior distribution instead.

REFERENCES


**AUTHOR BIOGRAPHIES**

**RUSSELL C. H. CHENG** is Emeritus Professor of Operational Research at the University of Southampton. He has an M.A. and the Diploma in Mathematical Statistics from Cambridge University, England. He obtained his Ph.D. from Bath University. He is a former Chairman of the U.K. Simulation Society, a Fellow of the Royal Statistical Society and Fellow of the Institute of Mathematics and Its Applications. His research interests include: design and analysis of simulation experiments and parametric estimation methods. He was a Joint Editor of the *IMA Journal of Management Mathematics*. His email and web addresses are <R.C.H.Cheng@soton.ac.uk> and www.personal.soton.ac.uk/rchc
AN AGENT-BASED MODEL OF SUPPLY CHAIN COLLABORATION: INVESTIGATING MANUFACTURER LOYALTY

Niniet I. Arvitrida
Loughborough University
Loughborough, LE11 3TU, UK
N.I.Arvitrida@lboro.ac.uk

Prof. Stewart Robinson
Loughborough University
Loughborough, LE11 3TU, UK
S.L.Robinson@lboro.ac.uk

Dr. Antuela A. Tako
Loughborough University
Loughborough, LE11 3TU, UK
A.Takou@lboro.ac.uk

Dr. Duncan A. Robertson
Loughborough University
Loughborough, LE11 3TU, UK
D.A.Robertson@lboro.ac.uk

ABSTRACT
Collaboration is believed to be a key driver to supply chain success. However, the ideal collaboration practice is difficult to achieve. Firms’ behaviour in supply chain collaborations is identified as the main reason for supply chain failure. To study how collaboration behaviour affects supply chain, an agent-based model is proposed. It represents two-stage supply chains, consisting of customers, manufacturers, and suppliers. The firms exist in a two-dimensional supply chain strategic space defined by dimensions of efficiency and responsiveness. In this paper, we examine the effect of manufacturer loyalty on supply chains’ performance in an innovative products market from a system perspective. Measuring the supply chains fill rate and number of surviving supply chains in the market, the results indicate that manufacturer loyalty at intermediate levels does not guarantee supply chain success in a market of innovative products, unless it is extremely high or does not exist at all.

Keywords: agent-based modelling, collaboration, supply chain, loyalty

1 INTRODUCTION
Supply chain management (SCM) practices are an essential component of business success. It shifts the conventional perspective from firm competition to supply chain competition. This point of view makes collaboration between firms in supply chains crucial in achieving supply chain success (Chopra and Meindl, 2007; Christopher, 2000; Lee, 2004). However, many firms have failed to establish successful collaborations leading to ineffective and inefficient SCM practices (Barratt, 2004; Cao and Zhang, 2011; Holweg et al 2005; Lambert and Cooper, 2000).

According to the supply chain literature, such as in Simchi-Levi et al (2000) and Chopra & Meindl (2007), collaboration between firms is achieved by maintaining long-term partnerships and a single supplier (single-sourcing). Collaboration strategies in SCM are shown to optimise the supply chain’s competitive advantage. Long term collaboration is considered to promote better communication between firms and accelerate the innovation process (Boddy et al 1998). Meanwhile, a single sourcing strategy for long term collaboration can minimise uncertainties in the supply side (Kraljic, 1983) and lead time to market (Christopher, 2000). Single-sourcing can also reduce the degree of competition in the market(Lee, 2004).
These claims are supported by examples of best practices in SCM, which are practiced by successful companies, such as Toyota and Benetton (Chopra & Meindl, 2007).

However, these partnership strategies do not suit all supply chains. Several findings argue that this approach does not always lead to a better supply chain performance, such as Anderson and Jap (2005), Burke et al (2007), Leeuw and Fransoo (2009), Squire et al (2009), and Sun and Debo (2014). Furthermore, the strategic management view recommends that single sourcing should be avoided as it does not encourage suppliers to improve their performance (Porter, 1997).

Based on these contradicting views, this study aims to obtain a better understanding of how collaboration approach suggested in SCM affects supply chains in the market. Instead of observing an individual supply chain, a system perspective is employed to assess the effect on market performance. As investigating the collaboration failure from system perspective is hard to perform by using empirical approaches, an agent-based modelling approach (ABM) is used in this study.

The use of ABM to study collaboration in supply chains is still limited to date. Studies that use ABM to study SCM focus primarily on software architecture rather than supply chain analysis, such as Barbuceanu et al (1997), Parunak et al (1998), Barbuceanu (1999), García-Flores et al (2000), Jiao et al (2006), Kwon et al (2007, 2011), and Siebers and Onggo (2014). In addition, ABM research that has addressed collaboration issues in SCM, such as Zhu (2008) and Chen et al (2013), only focuses on a single supply chain. They do not examine the supply chains at market level, taking a system perspective.

The modelling approach of this work is presented in a previous paper (Arvitrida et al 2015). The paper provides the conceptual model and several preliminary results for face validation. The discussion of the effect of particular collaboration behaviour on the supply chain is presented in this paper. It focuses on examining how firms’ behaviour in implementing supply chain collaboration strategies affects supply chains at a market level. The main hypothesis examined in the model is that collaboration between firms in the supply chain improves the supply chain performance considered from a market level perspective. The model proposed here represents two-echelon supply chains consisting manufacturer and supplier agents. The supply chains serve a market represented by customer agents. The model assumes a simplified strategic landscape where the manufacturer and supplier agents attempt to reach the best strategic fit on two dimensions (criteria) - responsiveness and efficiency. These dimensions are adopted from the definition of the basic supply chain capabilities required to satisfy customer priorities (Chopra and Meindl, 2007). A theory-driven approach is used to develop the model and observe the emerging outcome as a result of intrinsic behaviour of agents (manufacturer and supplier). Instead of observing a single supply chain, we examine the supply chain performance at market level, taking a system perspective. This is a novel approach that has not been implemented by previous studies in the SCM literature.

The remainder of this paper is organised as follows. Section 2 describes the agent-based model, including the conceptual and computer model. Section 3 explains the experiments and shows some preliminary results of the effect of a manufacturer’s collaboration behaviour, which is manufacturer loyalty to supplier. Section 4 discusses the results, and section 5 concludes the paper.

2 THE AGENT-BASED MODEL OF COLLABORATION IN SUPPLY CHAINS

The conceptual modelling framework introduced by Robinson (2014) is used to define the model. It specifies the following elements of a conceptual model: the model contents, inputs or experimental factors, outputs, assumptions and simplifications of the model. The model content described in this study incorporates the main features of the ABM described by North and Macal (2007), and Robertson and Caldart (2009). The ABM characteristics describe the model’s scope and level of detail.

The contents of the model consist of the agent, the environment, the interaction, and the behavioural rules (Figure 1). Three types of agents are simulated in this model: suppliers, manufacturers, and customers. They are located in a two-dimensional environment of the supply chain’s strategic position to represent the level of responsiveness and efficiency. Responsiveness that is described as y-axis is represented as the level of innovation from a customer’s view and manufacturing flexibility from the firm’s perspective.
Meanwhile, efficiency (x-axis) reflects the price and product value from the customer’s point of view and operational expenses from the firms’ viewpoint. Within the landscape, two infeasible areas are set to reflect the limits to the competitive space. These represent strategic spaces where a firm or a product with a relatively high level of innovation, is impossible to have a very low price (or cost) and product value, and vice versa.

The interaction and collaboration behavioural rules are described as follows. Customers create links with a manufacturer, which represents the decision to purchase the manufacturer’s product. At the same time, manufacturers create links with suppliers, which reflect the partnership decision. Manufacturers have a loyalty to represent the probability of selecting the same supplier for the next collaboration. Suppliers also have loyalty to the manufacturer. The supplier loyalty represents the probability that suppliers will imitate the manufacturer’s strategic movement to maintain their current partnership. As with manufacturer agents, suppliers have a maximum number of relationships with manufacturers. It represents a situation where suppliers can supply more than one manufacturer. The platform used in this study is NetLogo and the computer representation of the agent-based model is presented in Figure 1.

![Figure 1 The representation of the model (revised from Arvitrida et al 2015)]

The key inputs or experimental factors of the model are the collaboration strategy rules. They are the duration of collaboration between a manufacturer and a supplier, and the manufacturer maximum number of sourcings. The duration of collaboration represents the length of partnership between manufacturers and suppliers, and the maximum number of sourcing reflects the manufacturer’s maximum number of suppliers allowed to collaborate with.

To view the supply chain performance from system perspective, the outputs of this model are defined at market level. The performance is assessed by supply chains fill rate and the number of supply chains in the market. The supply chains fill rate is measured by dividing the number of customers served by the total number of customers in the system. Meanwhile, the number of supply chains in the market represents the number of supply chains that can survive in the market. For instance, if an experiment ends with only one surviving supply chain, it indicates that the collaboration approach pursued does not promote competition
in the market. The reader is referred to Arvitrida et al (2015) for the further detail. The summary of the conceptual model of this study is presented in Table 1.

<table>
<thead>
<tr>
<th>Table 1 The conceptual model (revised from Arvitrida et al 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model contents</strong></td>
</tr>
<tr>
<td><strong>(Scope and Level of Detail)</strong></td>
</tr>
<tr>
<td><strong>The agents:</strong></td>
</tr>
<tr>
<td>Customers, manufacturers, and suppliers.</td>
</tr>
<tr>
<td><strong>The environment:</strong></td>
</tr>
<tr>
<td>Two-dimensional strategic space defined as the degree of efficiency and responsiveness.</td>
</tr>
<tr>
<td><strong>The interaction:</strong></td>
</tr>
<tr>
<td>Each customer creates a link with a manufacturer, and each manufacturer makes connection/s (collaborations) with one or several suppliers.</td>
</tr>
<tr>
<td><strong>The collaboration behavioural rules (autonomy):</strong></td>
</tr>
<tr>
<td><strong>CUSTOMERS</strong></td>
</tr>
<tr>
<td>Each customer selects a manufacturer in accordance with its preference presented by its position.</td>
</tr>
<tr>
<td><strong>MANUFACTURERS</strong></td>
</tr>
<tr>
<td>Each manufacturer selects a supplier based on its preference presented by its position, within a probability of manufacturer loyalty to the supplier/s.</td>
</tr>
<tr>
<td><strong>SUPPLIERS</strong></td>
</tr>
<tr>
<td>Suppliers have a probability of loyalty to the manufacturers and maximum number of relationship with manufacturers.</td>
</tr>
<tr>
<td><strong>Inputs / Experimental Factors</strong></td>
</tr>
<tr>
<td><strong>Collaboration strategy:</strong></td>
</tr>
<tr>
<td>1. The duration of collaboration between supplier and manufacturer, and</td>
</tr>
<tr>
<td>2. Manufacturers' number of sourcing.</td>
</tr>
<tr>
<td><strong>Outputs</strong></td>
</tr>
<tr>
<td>1. Supply chains fill rate, and</td>
</tr>
<tr>
<td>2. Number of supply chains in the market.</td>
</tr>
</tbody>
</table>

3 EXPERIMENT AND PRELIMINARY RESULTS

We simulate supply chains that operate in a market of innovative products, such as automobile and high technology devices. A thousand customer agents are randomly located within the environment of the model. Their positions in the space represent their unique preference to buy a product from a manufacturer. They are set to have a very limited degree to compromise with their preference to decide which manufacturer is most suitable to supplying their taste or requirement preferences. They will only create a link with a manufacturer if the manufacturer stays within their radius of compromise, measured from their position in the two-dimensional space. The radius is determined by using a percentage of the diagonal length of NetLogo world (the agent’s space or environment, where the percentage used is 10%). They are also situated to have no loyalty to the manufacturer, so they will select another manufacturer if it is closer to their preference than the previous one.

Ten manufacturer and ten supplier agents are set to collaborate and compete with each other. The duration of collaboration is homogenously set to be short-term partnership, which is defined as four ticks. A tick reflects the time unit in which a firm can change its strategic position by one grid unit. We assume that one tick represents three months, so that the length of collaboration in this experiment is one year. Manufacturer and supplier who cannot manage to find a firm for collaboration will die after they have exceeded a limit to loss, which is defined as four ticks or a year.

A single-sourcing supply chain is applied to the relationship between manufacturer and supplier. It means that a manufacturer is only allowed to collaborate with one supplier. This rule also applies to the suppliers, which are limited to have only one link with manufacturer. In this situation, the supplier agents have no loyalty to the manufacturer, so they will move to approach another manufacturer instead of following the strategic movement of the currently linked manufacturer.
This paper presents the analysis of the results on the effect of manufacturer loyalty on the market performance. The performance is presented in two measures: supply chains fill rate and number of supply chains in the market.

Manufacturer loyalty represents the manufacturer trust to work with the supplier. The value is defined at five levels (or scenarios) of loyalty as a probability: 0%, 25%, 50%, 75%, and 100%. These values are selected to represent the following degrees: extremely disloyal (0%), disloyal (25%), moderately loyal (50%), loyal (75%), and extremely loyal (100%). These scales are chosen empirically to observe the effect of different firms’ level of loyalty on the model outputs. The variation of the variables for the scenarios is summarised in Table 2.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>0%</th>
<th>Extremely disloyal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 2</td>
<td>25%</td>
<td>Disloyal</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>50%</td>
<td>Moderately loyal</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>75%</td>
<td>Loyal</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>100%</td>
<td>Extremely loyal</td>
</tr>
</tbody>
</table>

Each scenario is run for 50 replications with 1000 ticks for each replication. As we assuming a tick represents three months, 1000 ticks are equal to 3000 months or 250 years. This time period is considered to be a sensible extreme duration, as many old firms can survive for more than 200 years, such as Royal Delft - a porcelain company from Netherlands that has been active for more than 350 years, Ekelund – a Swedish clothing company that has survived since 1692, and Lambertz – a German chocolate company that has been operated since 1688. Common random numbers are used in the simulations in order to reduce the variations of the simulation outputs (Robinson, 2014). It means that the similar streams of random numbers are applied for all scenarios. This approach can also be useful to reproduce the simulation results for allowing further investigation.

We use descriptive and inferential statistical analysis to interpret the simulation results. Boxplots are employed to obtain the visual descriptive patterns of the results, and confidence intervals are used to infer the significant difference between scenarios.

### 3.1 Boxplot Analysis

Boxplot analysis allows us to describe the data characteristics of the outputs. As shown by Figure 2, the boxplots show that there is a pattern for the effect of manufacturer loyalty to both supply chains fill rate and number of supply chains in the market.

Figure 2a shows that the median and mean of supply chains fill rate follow a slight U shaped pattern as the probability of manufacturer loyalty increases. In scenario 1 (manufacturer loyalty is 0%), the median of supply chains fill rate is 9.3% and the mean is 10.47%. Then, both values gradually decrease as the probability of manufacturer loyalty increases up to 75% (scenario 4). The values rise to 14.3% for the median and 14.37% for the mean when the manufacturer loyalty is 100% (scenario 5). In the last scenario, the model provides the highest outputs comparing to other scenarios.

Meanwhile, the median and the mean of the number of supply chains in the market also depict a slight U shaped (Figure 2b). Both the median and the mean of the output decrease from three supply chains to two supply chains, from scenario 1 to scenario 2 (manufacturer loyalty is 25%). The values do not change until the manufacturer loyalty is 75% in scenario 4. In the last scenario, both values increase to four supply chains.

The variation of the simulation outputs can also be observed by looking at the tall of both boxplots. The charts show that the highest variations of both outputs are in scenario 5, where 100% of manufacturer loyalty is applied. The first scenario provides the second highest variation, while the others (scenario 2, 3, and 4) reveal the same values of the outputs.
In general, the boxplots indicate that 100% of manufacturer loyalty has the greatest effect to both supply chains fill rate and “the number of supply chains in the market”, and 0% of manufacturer loyalty provides the second highest outputs. It also implies that scenario 2, 3, and 4 provides no different effect to the simulation outputs.

Figure 2 Boxplots of the model outputs with a line of mean values for all scenarios of manufacturer loyalty. a) Supply chains fill rate. b) Number of supply chains in the market

3.2 Confidence Intervals for Multiple Comparisons

Confidence intervals (10% overall significance level) are constructed in order to compare the scenarios. If the interval includes zero, it can be concluded that there is no significant difference between two scenarios. The Bonferroni inequality is used to determine the overall level of significance (Robinson, 2014). In order to compare the 5 scenarios amongst themselves, ten confidence intervals are required. Hence in order to maintain an overall 10% significance level, the individual level of significance for each confidence interval is 1% (overall significance level/number of confidence intervals = 10/10=1).

As shown in Table 3, scenario 5 (100%) is the most different compared to the other four scenarios, in terms of supply chains fill rate. It provides the highest value of supply chains fill rate. Scenario 1 (0%) is also significantly different, but it provides similar results to scenario 2 (25%). The supply chains fill rate is higher compared to scenario 3 (50%) and 4 (75%), but it is lower compared to scenario 5. Scenario 2 (25%) is significantly different from scenario 4 and 5, but it is insignificant with scenario 1 and 3. It provides higher supply chains fill rate than scenario 4, but it is lower than scenario 5. Scenario 3 (50%) is only significantly different from scenario 1 and 5. It has lower supply chains fill rate than scenarios 1 and 5. Lastly, the supply chains fill rate in scenario 4 (75%) is significantly lower comparing to scenarios 1, 2, and 5.

A similar approach is applied to infer the difference between scenarios for the number of supply chains in the market (Table 4). In general, scenario 5 (100%) provides a significantly higher number of supply chains in the market than other scenarios, particularly when it is compared to scenario 2, 3, and 4. It is
Arvitrida, Robinson, Tako, and Robertson

considered to be only insignificantly different when it is compared to scenario 1. Scenario 1 (0%) is only significantly different from scenario 3. It generates more number of supply chains in the market than scenario 3. Finally, insignificant difference between scenarios is concluded for the remaining comparisons.

Table 3  Confidence interval comparison of supply chains fill rate between all scenarios of manufacturer loyalty (presented in percentage, with overall confidence interval ≥ 90%)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2 (25%)</th>
<th>3 (50%)</th>
<th>4 (75%)</th>
<th>5 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (0%)</td>
<td>(-1.44, 2.81)</td>
<td>(0.29, 3.69)</td>
<td>(0.89, 4.29)</td>
<td>(-6.65, -1.15)</td>
</tr>
<tr>
<td></td>
<td>No significant difference</td>
<td>Scen.1 (0%) &gt; Scen.3 (50%)</td>
<td>Scen.1 (0%) &gt; Scen.4 (75%)</td>
<td>Scen.1 (0%) &lt; Scen.5 (100%)</td>
</tr>
<tr>
<td>2 (25%)</td>
<td>(-0.66, 3.28)</td>
<td>(0.14, 3.67)</td>
<td>(-7.37, -1.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No significant difference</td>
<td>Scen.2 (25%) &gt; Scen.3 (75%)</td>
<td>Scen.2 (25%) &lt; Scen.5 (100%)</td>
<td></td>
</tr>
<tr>
<td>3 (50%)</td>
<td>(-0.97, 2.16)</td>
<td>No significant difference</td>
<td>Scen.3 (50%) &lt; Scen.5 (100%)</td>
<td></td>
</tr>
<tr>
<td>4 (75%)</td>
<td>Scen.4 (75%) &lt; Scen.5 (100%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4  Confidence interval comparison of number of supply chains in the market between all scenarios of manufacturer loyalty (with overall confidence interval ≥ 90%)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2 (25%)</th>
<th>3 (50%)</th>
<th>4 (75%)</th>
<th>5 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (0%)</td>
<td>(0, 2)</td>
<td>(0, 2)</td>
<td>(1, 2)</td>
<td>(-2, 0)</td>
</tr>
<tr>
<td></td>
<td>No significant difference</td>
<td>No significant difference</td>
<td>Scen.1 (0%) &gt; Scen.4 (75%)</td>
<td>No significant difference</td>
</tr>
<tr>
<td>2 (25%)</td>
<td>(0, 1)</td>
<td>No significant difference</td>
<td>Scen.2 (25%) &lt; Scen.5 (100%)</td>
<td></td>
</tr>
<tr>
<td>3 (50%)</td>
<td>(0, 1)</td>
<td>No significant difference</td>
<td>Scen.3 (50%) &lt; Scen.5 (100%)</td>
<td></td>
</tr>
<tr>
<td>4 (75%)</td>
<td>No significant difference</td>
<td>Scen.4 (75%) &lt; Scen.5 (100%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4  DISCUSSION

Both boxplots and confidence intervals indicate that very loyal manufacturers (when manufacturer loyalty is 100%) can lead to better supply chain performance in the market. The supply chains fill rate is significantly higher compared to collaboration with less manufacturer loyalty. It also helps more supply chains to survive longer in the market compared to the in between degrees of manufacturer loyalty. Even though the outputs are lower than the results of 100% manufacturer loyalty, it is significantly higher than having a moderate degree of loyalty.

However, disloyal manufacturers (when manufacturer loyalty is 0%) also provide a better supply chains fill rate and number of supply chains in the market as opposed to the in between degrees of manufacturer loyalty. Even though the outputs are lower than the results of 100% manufacturer loyalty, it is significantly higher than having a moderate degree of loyalty.

This result is counter to the current SCM concept of achieving supply chain success through collaboration achievement. A possible explanation for this could be that when all manufacturers in the market are disloyal with the suppliers, it can support a perfect competition environment that benefits the
supply chain as a system. In other words, extreme loyalty (0% and 100%) can be beneficial to support the performance of all supply chains to serve the market.

In general, these results suggest that the firms’ collaborative behaviour affects the supply chains in the market. It contributes to the overall service level of all supply chains in the market, represented by the supply chains fill rate, and the number of supply chains who can survive in the market. We consider that a higher supply chains fill rate is better for the market because more demand is fulfilled. Meanwhile, the more supply chains that survive in the market do not always represent a better performance for the market. Having more supply chains survived in the market may be undesirable for the firms who compete, even though it represents lesser supply chain failures in the operating market. However, having only one supply chain survives in the market is also not preferred. If this situation is emerged in a significant frequency, it suggests that the behavioural rules implemented potentially leads to a monopoly situation. It indicates that market intervention may be required to prevent this from happening.

5 CONCLUSIONS

An agent-based model of collaboration in supply chains has been developed in this study. An experiment on supply chains of innovative products is performed to examine the effect of manufacturer loyalty to the supply chains fill rate and number of supply chains in the market.

The preliminary runs of the model show that the manufacturer loyalty affects market performance, in terms of the supply chains fill rate and number of supply chains in the market. If manufacturers are perfectly loyal to the supplier, by choosing the same supplier every period, it would lead to a higher supply chains fill rate and a higher number of supply chains in the market. On the other hand, when a manufacturer has no loyalty to the supplier, the supply chains fill rate and number of supply chains in the market are notably higher compared to when a manufacturer has a moderate degree of loyalty. These results are contradictory with the current concept of SCM even though very loyal manufacturers generate better supply chains fill rate and more number of supply chains in the market rather than very disloyal manufacturers.

The experiments performed in this paper consider only one aspect of supply chain collaboration in the market, supply chain loyalty. As a next step we plan to consider further aspects of collaboration and competition behaviour, including the distance of manufacturer strategic movement, supplier loyalty, and the supplier’s maximum number of relationships. The results are also analysed by disregarding the interaction between collaboration factors, such as duration of collaboration and number of sourcing. A similar analysis is also under way considering supply chain competition. Varying the model content, such as supply and demand characteristics, will also be examined in the future in order to gain a better understanding of collaboration in the supply chain. Moreover, several case studies will be used to test the model.

The model presented in this study represents the strategic space and firm behaviour in a very simplified fashion. However, it can be extended to address more supply chain issues, such as collaboration in functional products market, to study the market at a system level. Moreover, incorporating a learning capability for each agent and providing alternative measures of performance could enrich the model.

REFERENCES


AUTHOR BIOGRAPHIES

NINIET INDAH ARVITRIDA is a PhD student in School of Business and Economics, Loughborough University. She has a Master of Engineering (cum laude/with honours) in Supply Chain Management and Bachelor of Engineering (cum laude/with honours) in Industrial Engineering, both obtained in Sepuluh Nopember Institute of Technology, Indonesia. She has seven years academic experience, which is five years in Indonesia and two years in the UK. She managed a full studentship for her PhD from Directorate General of Higher Education, Ministry of Education and Culture, Indonesia.

STEWART ROBINSON is Professor of Management Science at Loughborough University, School of Business and Economics. He is also Acting Dean of School of Business and Economics at Loughborough University, President of the Operational Research Society, and Chair of the Management Group for the Institute for Sports Research; a collaboration between Loughborough University and Nanyang. He holds an honour degree in Management Science (Operational Research) (1985) and a PhD in Management Science (1998), both from Lancaster University. Key areas of interest are conceptual modelling, model validation, output analysis and alternative simulation methods (discrete-event, system dynamics and agent based). Home page: www.stewartrobinson.co.uk.

ANTUELA A. TAKO is a Lecturer in Operations 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 and conceptual modelling. She is an Associate Fellow of the Operational Research Society (AFORS). Home page: www.lboro.ac.uk/departments/sbe/staff/profiles/takoantuela/tako-antuelaanthi.html.

DUNCAN ROBERTSON is a member of the management sciences and operations management group at Loughborough University, previously having worked at Warwick Business School, Manchester Business School, Sun Yat-sen University, and researched at The Wharton School at the University of Pennsylvania. He is a Fellow of St Catherine’s College, University of Oxford. Before undertaking his DPhil at Saïd Business School, he qualified as a Chartered Accountant with KPMG. He originally trained as a physicist at Imperial College London. Home page: http://www.lboro.ac.uk/departments/sbe/staff/profiles/robertsonduncan/robertson-duncan.html.
AGENT-BASED SIMULATION OF HERDING IN FINANCIAL MARKETS

Xin Liu
Department of Management Science
Lancaster University Management School
Lancaster, LA1 4YX, U.K.
x.liu1@lancaster.ac.uk

Dr. Roger J. Brooks
Department of Management Science
Lancaster University Management School
Lancaster, LA1 4YX, U.K.
roger.brooks@lancaster.ac.uk

ABSTRACT

There are several models of financial markets which look at the herding effect. This is a situation where many market traders act as a herd in that they all behave in a similar way with their trading. This type of behaviour may explain certain observed characteristics (or ‘stylised facts’) in real markets. However, the various models have different herding mechanisms and market settings. This paper sets out the rationale of our approach and our initial work in trying to get a better understanding of herding in financial markets. Our research, though, is at an early stage. The basic methodology is to reproduce and compare some of the existing models, hopefully leading to a more general understanding and measure of herding and the relationship with market behaviour. One model has been investigated so far and this is described. A more general issue is the research importance of reproducing previous studies.

Keywords: Agent based simulation, Financial markets, Herding, Gurus, Reproducibility

1 INTRODUCTION

This introduction begins with a definition of agent-based simulation of financial markets, moving on to herding. Then the objective of the research and the chosen methodology are discussed. The initial herding model is described in section 2 with the outputs set out in section 3.

1.1 Agent-based simulation and financial markets

Agent-based simulation is a simulation technique based on the agent. There is no agreed definition of an agent. However, one possible definition is an entity which models a cognitive process such as an individual’s intention and belief (Edmonds and Mohring, 2005). For example, in the context of the economy, consumers, companies, regulators and governments could be agents.

Financial markets are markets where funds get exchanged. Stock and bond markets are examples of financial markets. Several empirical studies of financial markets have found common statistical features across financial instruments like cash or bonds, and across different time scales (Cont, 2001). These common statistical features are called stylized facts which financial market models are expected to produce. There are three well known stylized facts: real-returns follow a non-Gaussian distribution; absolute value or square value of real returns for each period are correlated; real returns for each period are not correlated (Taylor, 2005). Real returns are non-Gaussian in that compared to the normal, the distribution is more sharp and narrow in the middle with fat tails. This means that it has a positive kurtosis (Cont, 2005). Correlations of absolute value or square value of real returns is often described as volatility clustering. Big returns tend to be followed by big returns and small returns by small returns and so there are periods where the market is very volatile and periods where it is fairly stable (Cont, 2005). On the other hand there is little or no...
correlation in whether successive periods give a positive or negative return, hence real returns are not correlated.

Agent-based simulation in financial markets models the behaviour of the different traders with their interactions generating the market and determining the market price. The behaviour may include learning and adaptation. Agent based simulation in financial markets is mainly divided into two types: N-type models and autonomous agent models (Chen et al 2012). The N-type models (Lux, 1998; Brock and Hommes, 1998) divide the agents into N different types of beliefs although their individual characteristics with each group may vary. The autonomous agent models (Arifovic, 1996; LeBaron, 2001) divide individuals into different rules.

Unfortunately, there is no agent based modelling that generates all stylised facts. Herding has been used in agent based simulation of financial markets for explaining some of stylised facts and this is discussed in the next section.

1.2 Herding

Herding describes a behaviour where groups of people keep making similar decisions. This may be due to some type of interaction between them or just because they are using similar decision making rules. From a review of 20 models in the literature which focus on herding behaviour, there are neither general model mechanisms nor herding mechanisms. The agents in the market differ from N-type (Carro et al 2015; Lux 1995) to autonomous type (Lebaron and Yamamoto, 2007; Mauri and Tettamanzi, 2012). In addition, there are different rules of price setting like excess demand (Lux and Marchesi, 1999; Chowdhury and Stauffer, 1999; Alfarano et al 2005), and order-driven market (Lebaron and Yamamoto, 2007; Tedeschi et al 2012). In the Yamamoto (2011) and Kaizoji et al (2011) models, two stocks are considered in the market while in others only one stock is considered.

The most important part which decides the herding factor is also different in the literature. Some papers (Alfarano et al, 2005; Alfarano and Milaković, 2009; Carro et al., 2015) are based on Kirman’s idea of ant behaviour (1993) when facing two identical foods. The herding factor in these models is based on a discrete choice using a probability factor which usually affects the transition probability of two groups such as pessimistic and optimistic traders. Some papers (Lux, 1995; Lux, 1998; Lux and Marchesi, 1999) describe a herding factor with a continuous time discrete model which is also similar to Kirman’s idea but with different probability formulae. Unlike Kirman’s switching type model, there are several models mainly based on social imitation. The type of the agents in these models does not change, and the herd effect is modelled through impacting the agent’s decisions. The imitation rules in these models vary: the agents in the spin models (Chowdhury and Stauffer, 1999; Bornholdt, 2001; Kaizoji et al., 2002) imitate their nearest neighbours; the agents in the models (Markose et al., 2007; Kaizoji et al., 2011; Tedeschi et al., 2012; Mauri and Tettamanzi, 2012; Yang et al., 2012) are influenced by other agents such as the opinions of majorities or of successful agents; some autonomous models (Labaron and Yamamoto, 2007; Yamamoto, 2011; Chen and Yeh, 1999) set the herd mechanism through a genetic algorithm or genetic programming learning; Some models divide the agents into different groups with different size to replicate the herd by clustering (Chen et al, 2013; Lee and Lee, 2013; Manahov and Hudson, 2013).

1.3 Objective and Method of Research

The overall objective of this research is to improve the understanding of the nature and effects of herding in financial markets. Herding, as an important behaviour in financial behaviour studies, is a likely partial explanation of some of the observed stylized facts like volatility clustering and fat tails.

As discussed in section 1.2, there are several papers modelling herding but with different mechanisms and also applied to different situations. It would be very useful to find a more general measurement of herding, irrespective of the specific herding mechanism, where the same herding level produces the same general market behaviour. The initial research approach therefore is to make a comparison of some of the previous models from the literature with a variety of herding mechanisms.
A ‘reproducing’, ‘repeating’ or ‘replicating’ method is taken by which we mean attempting to build a model from a previous study, as described in the paper of that study, and hopefully obtain similar results. Although our overall aim is an improved general understanding of herding we believe that reproduction of studies is a valuable research undertaking in itself and one that is not done enough in simulation and Operational Research (OR), and perhaps even across science in general.

Repeating experiments is a key aspect of the scientific method. Karl Popper’s (Popper, 2002) view was: “We do not take even our own observations seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated ‘coincidence’, but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable. Every experimental physicist knows those surprising and inexplicable apparent ‘effects’ which in his laboratory can perhaps even be reproduced for some time, but which finally disappear without trace.” In a traditional natural science experiment, then, replication increases the confidence in the generalisability of the results. A single experiment may produce misleading results due to various factors such as errors, the specific conditions, or chance. A recent large scale project reproducing 100 psychology studies from top journals found that “only 39% of effects were subjectively rated to have replicated the original result” with the mean effect size being just half that of that of the original results (Open Science Collaboration, 2015). Very low levels of reproducibility have also been found in oncology studies in medicine (Begley and Ellis, 2012). One issue is a publication bias in that papers are more likely both to be submitted and published if they contain novel and unexpected (i.e., ‘exciting’) results. This can lead to bias in the method such as the selective reporting of results, and under certain assumptions the chance of a study being correct may be very low (Ioannidis, 2005). Of course, the peer review system does not involve checking the details of the study and its effectiveness is arguable (Smith, 2015).

Similar issues can potentially apply to a simulation or OR study. Results could be misleading due to errors, specific (perhaps implicit) assumptions, the specific conditions (e.g., parameter values or the scenario), and the experimentation and analysis process. Hence, one benefit and reason for reproducing a simulation study is a verification check of the model. Even high profile studies from famous scientists can contain basic errors. Reinhart and Rogoff (2010) published an economics paper on the relationship between debt and GDP for countries. The results have been quoted by key policymakers in the debates on austerity and economic policy. However, student Thomas Herndon found when attempting to reproduce the results that simple mean and median formulae in the original Excel spreadsheet referred to the incorrect cell range and so accidentally excluded five countries (Herndon et al., 2014). A second benefit is that reproducing a study should show up the assumptions being made. This also applied in the Herndon et al. (2014) reproducing study where they found that in the original study some data had been excluded and also that in the calculations for different debt / GDP categories a single average value was used for each country irrespective of the number of years of data. They were critical of both of these assumptions.

In addition to these benefits a simulation reproducing study can go further than simply repeating the original study. New results can be generated through different experiments, new analysis, additional outputs etc. This may also lead to new insights and to a greater depth of understanding of the reasons for the model behaviour.

The Wilensky and Rand (2007) paper has a good discussion of the nature and benefits of reproducing simulation models in the context of agent-based simulation in the social sciences. They list six dimensions which may differ between the original and reproducing study: “time, hardware, languages, toolkits, algorithms, authors.” Of course one problem is that if the study cannot be reproduced then it may not be clear whether there is an error or whether it is due to one of these dimensions. Another practical problem is that the original paper may not contain all the details of the model. What is required is the conceptual model – i.e., a complete software independent description of the model (Brooks and Robinson, 2001). In forecasting, Boylan et al. (2015) were unable to reproduce a previous study and one issue was insufficient information on the methods and data used.

Based on the above discussion, the five main steps planned for our research are as follows:
1. Selecting: several herding papers will be selected based on the level of model description and the model mechanism.
2. Checking: the models will be built based on the papers to check if the results can be reproduced or not.
3. Extending: then an extended model will be produced to obtain a more detailed understanding of the models through new experiments and analysis.
4. Comparing: after reproducing several papers, the results will be compared to see under what conditions similar results are obtained.
5. Generalising: attempt to generate a general herding measure or rule.

To date, one paper has been selected and the next section describes our current progress in reproducing it.

2 THE MODEL

This model is based on the Tedeschi et al. (2012) paper called ‘Herding effects in order driven markets: The rise and fall of gurus’. The model uses zero intelligent agents which stems from Gode and Sunder (1993) where agents trade according to their random behaviour. However, each agent is influenced by one other agent who is most likely to be a successful trader. The guru is the trader who has the most imitators. Each trader can view the current bid (buy) and ask (sell) price and submit a market order or limit order. A market order is an order which can be traded immediately fully or partially. A limit order is an order which cannot be traded immediately and is added to the order book. All parameter settings follow the original paper unless there is no information in the paper. The traders in the model have the same situation at the start. They all have 100 stock at price 1000 and 100,000 cash.

2.1 Reproducibility issues

To date, we have not managed to reproduce the results from the original paper. One issue is that although the paper contains a good level of detail in the description of the model, two of the parameter values are not specified. Another issue may be that the trader’s behaviour is based on a utility function which may make the model quite sensitive to the specific parameter values. In the results in the Tedeschi paper we also observe that the market price starts at 1000 but very quickly drops to about 500 and then fluctuates around that level. This initial drop is not particularly realistic if we take the initial conditions to be a stable situation. In view of these difficulties we decided to change to trading behaviour from the utility function to alternative rules that we considered are reasonably realistic of typical trading behaviour. This leaves the essential herding part of the model unchanged.

The model description below describes our current model. Sections 2.2 (network) and 2.3 (expectation) therefore follow Tedeschi et al. (2012), with section 2.4 (market mechanism) being our revised market trading rules.

2.2 The Network

The network is the starting point of this whole model to build a communication picture among the agents who are traders in this model. All agents are nodes in this network and the edges are the communication links. For each node, there is just one out-going link to keep it simple. This constrains that one agent can just get advice from one other agent.

In all the formulae below, the superscript \( i \) is the particular agent \( i \) from 0 to 149 (total of 150 agents) and the subscript \( t \) is the time period \( t \) from 0 to 1000. The wealth is equal to the current value of the stock plus the cash holding in equation (1):

\[
W^i_t = S^i_t p_t + C^i_t \quad (1)
\]

In equation (1): \( W, S, C \) and \( p \) with just subscript \( i \) are the symbols of wealth, stock, cash and price. Then, \( W^i_t \) is the wealth for agent \( i \) at time period \( t \).
The fitness function measures the level of wealth for each agent in equation (2):

$$f_t^i = \frac{W_t^i}{W_t^{max}}$$  \hspace{1cm} (2)

In equation (2), $f$ represents the fitness. $W_t^{max}$ is the wealth of the agent who has the maximum wealth among all agents at time $t$.

Then the probability function based on the fitness forms the whole communication network. For each agent, there is one random assigned neighbour that is a potential newly formed link at the beginning of each period. Each agent is faced with a choice of keeping the existing neighbour or linking to the newly formed neighbour. The probability function (3) decides the probability of switching to the new link.

$$p_t^i = \frac{1}{1 + e^{-\beta_t^i (f_t^i - f_t^k)}}$$  \hspace{1cm} (3)

The $p_t^i$ with superscript in equation (3) is the symbol of agent $i$’s probability of switching to the newly formed link. $\beta_t^i$ is a random number that follows the uniform distribution $(5, 45)$. This random number protects from locking to imitate the same guru. The superscript $k$ is the existing neighbour of agent $i$ and $j$ is the newly random formed neighbour of agent $i$.

### 2.3 The Expectation

The agent’s expectation of future returns is based on the agent’s own idiosyncratic expectation and the neighbour’s expectation. The returns in these formulae are spot returns with time interval from time $t$ to time $t + \tau$. The agent’s idiosyncratic expectation is based on one volatility factor and one other normal noise. The volatility factor is:

$$\sigma_t^i = \sigma_0^i (1 + l_t^i \% (1 - w))$$  \hspace{1cm} (4)

The $\sigma$ is the volatility, $w$ is the herding factor explained below and $l_t^i \%$ is the percentage of incoming links for agent $i$ at time period $t$. $\sigma_0^i$ is a uniform distributed factor for agent $i$ from 0 to $\sigma_0$ where $\sigma_0$ is a uniform distribution from 0 to 0.07. Then the idiosyncratic expectation is based on the result obtained from formula (4) multiplied by the normal noise:

$$\hat{r}_{t,t+\tau}^i = \sigma_t^i \epsilon_t^i$$  \hspace{1cm} (5)

The $\hat{r}$ is the symbol of agent’s idiosyncratic expected return and $\epsilon$ is the normal noise with normal distribution N (0,1) with mean 0 and standard deviation 1.

Then the return of each agent follows:

$$r_{t,t+\tau}^i = w \hat{r}_{t,t+\tau}^i + (1 - w) r_{t,t+\tau}^j$$  \hspace{1cm} (6)

In this formula, $r$ is the symbol of return and $\hat{r}_{t,t+\tau}^j$ is the neighbor $j$’s idiosyncratic return when agent $j$ is the neighbour of agent $i$. The herding factor $w$ in equation (4) and (6) takes values between 0 and 1. Lower $w$ means more herding.

### 2.4 The Market

The future price for each agent is also formed based on their expectation. The expected price at time $t + \tau$ when the agents are currently at time $t$ is:

$$\hat{p}_{t,t+\tau}^i = p_t^i (1 + r_{t,t+\tau}^i)$$  \hspace{1cm} (7)

The $\hat{p}$ here in the equation (7) is the future price. At each time, each agent always has just one order in the market. The discount factor in this model is ignored. They compare the current price with the expected price from equation (7) to submit a buy or sell order. Obviously, if the current price is greater than the expected price, traders sell now and hope to buy in the future to make a gain; otherwise, the traders buy now at a lower current price and hope to sell at the higher future price to make profits. When the difference between the current price and expected price is greater than or equal to 50, the buy order price for certain agent is the current price plus 50 and the sell order price is the current price minus 50 to ensure some gains. Otherwise, traders submit limit orders at buy order prices or sell order prices which are expected price minus 50 or expected price plus 50 to ensure at least 50 profit if the expectation is right. The amount of the
order for a certain agent to submit is based on their cash, the probable gain they make and also the risk attitude. The probable gain is equal to:

\[ g_i^t = \frac{|p_{oi}^t - \hat{p}_{oi}^{t+1}|}{p_{oi}^t} \quad (8) \]

The \( g_i^t \) is the probable gain for trader \( i \) at time \( t \) and \( Po \) is the price submitted into the order driven market. The equation for amount the agent to submit is:

\[ s_i^t = \frac{s_i^t g_i^t}{Po_i^t Max} \quad \text{(when } g_i^t \leq Max) \]
\[ s_i^t = \frac{c_i^t}{Po_i^t} \quad \text{(when } g_i^t > Max) \quad (9) \]

The \( s \) and \( Max \) stands for the stock amount to submit in the market and the acceptable percentage of gain to use the full cash. \( Max \) is a uniform distribution from 0.5 to 1.

The agent’s order is submitted on a rolling basis in a random sequence. Trading follow the rules in order driven market derived from Chiarella et al (2009). During time period \( t \), when agents are entering the market, the current price is equal to the average of best bid and best ask. If the price of the buy order is greater than the best ask or the price of the sell order is smaller than the best bid, there is a match. The new order is the market order and trades at best ask or bid. Once there is no match, and the new order still has some untraded amount of order, the remaining amount of new order goes into the order book and the best bid or best ask is updated accordingly.

3 RESULTS

The current model is not producing a strong herding effect. Changing the herding parameter does not produce a noticeable change in the pattern of the results or the output values. Also gurus have a short lifespan (i.e., the guru keeps changing from one agent to another) and only a little more wealth than the other agents. Therefore the model needs further development. Altering the parameters and model details such as the fitness function in appropriate ways should give a stronger herding effect. For example, by increasing the probability of an agent switching to a successful agent as their neighbour.

Since the model is not yet producing suitable results this section will focus on listing the main model outputs. The results reported in Tedeschi et al. (2012) include for a single illustrative run: diagram of the final network; time series charts of which agent is the guru, percentage of links to the guru, guru fitness, wealth of guru and imitators and others, market price. For 100 runs the results include guru average life, average wealth of guru and imitators and others, decumulative distribution functions for wealth and stocks, average values for various market values such as the volume of orders. Tedeschi et al. (2012) ran the model using herding parameter \( w \) values of 0.1, 0.5 and 1. We have added other outputs to measure the stylised facts of price distribution (fat tails) and price correlations (volatility clustering).

Figure 1 shows an illustrative time series chart of market price for one run of our model with \( w = 0.1 \).

4 CONCLUSION AND RECOMMENDATIONS

The paper has set out the research objective of improving the understanding of herding in financial markets. The approach is based on reproducing and comparing previous herding models. One difficulty, however, is in getting enough detail of the original study and we have encountered such problems. Nevertheless, reproducing is important for increasing confidence in scientific results and assessing their generalisability. This approach should be given more emphasis and priority in simulation research.
**Figure 1** Time series of market price for one run with w = 0.1

**REFERENCES**


Liu and Brooks


Lee S and Lee K (2013). Heterogeneous expectations leading to bubbles and crashes in asset markets: Tipping point, herding behavior and group effect in agent-based model. At press.


Liu and Brooks


AUTHOR BIOGRAPHIES

XIN LIU received a BSc (Hons) Accounting and Finance from Lancaster University in 2013 and an MRes Management Science from the same university in 2014. She is now doing her PhD at the same university.

ROGER J. BROOKS is a lecturer in the Management Science department at Lancaster University. He has a B.A. (Hons) degree in mathematics from Oxford University and an MSc (Eng) and PhD in Operational Research from Birmingham University. He is co-author with Stewart Robinson of a textbook on simulation. His research areas include conceptual modelling, Boolean networks, agent-based simulation, and sports statistics.
EFFECTIVE QUALITY ASSURANCE OF GOVERNMENT MODELS USED TO SUPPORT BUSINESS CRITICAL DECISIONS

Mr Alan Robinson
Defence Science and Technology Laboratory
Portsdown Hill Road, Fareham
aprobinson@dstl.gov.uk

Mr Paul Glover
Defence Science and Technology Laboratory
Portsdown Hill Road, Fareham
peglover@dstl.gov.uk

ABSTRACT

Modelling underpins much of the work of government and provides evidence in support of decisions which significantly affect people’s lives and have major financial implications. It is vital, therefore, that this modelling is fit-for-purpose. Much work has been undertaken across Government in recent years to enhance the quality – and the quality assurance - of models and modelling, largely prompted by experience with the Intercity West Coast franchise competition in 2012. A key element of this was a cross-government review of the quality assurance of government analytical models, completed in 2013 (the Macpherson review) which led to consolidated guidance published in Spring 2015 (the Aqua book).

This paper briefly reviews the key points from the MacPherson review and then discusses the main elements of the Aqua book. Finally, it highlights ongoing activity to further develop and share best practice guidance across Government and more widely.

Keywords: best-practice, guidance, modelling, analysis, Aqua book.

1 INTRODUCTION AND BACKGROUND

The issues associated with the award process for the InterCity West Coast mainline franchise by the Department of Transport (DfT) in 2012 were well documented at the time including via two cross-government reviews. The Laidlaw Review (2012) addressed the specific issue of “what went wrong” in this particular case. Given that modelling was a major – but not sole - input to the failed decision making process a broader review was also commissioned to make sure that general lessons were learned across Government. This second review - the Macpherson Review (2013) - considered broader issues associated with the quality and the quality assurance of models and modelling undertaken in support of Governmental decision making.

Robinson and Glover (2014) reported on the themes and ideas that were emerging from the multidisciplinary Macpherson review at that stage, in particular related to best1 practice, its documentation and promulgation. This paper brings the story up to date, by:

• Briefly reprising the key points from the Macpherson review and summarising progress in implementing its findings via a “year-on” review (HM Treasury 2015b).

1 It is, of course, arguable whether a single set of best practice exists or whether it is better to think of this as a body of good practice that constantly evolves as well as needing to be tailored to individual situations. Although the latter is undoubtedly closer to ground-truth, the term “best” practice is the one being used within Government for this and is thus the term used in this document.
Robinson and Glover

• Introducing the now published best practice guidance2 (HM Treasury 2015a) – the Analysis QUality Assurance (Aqua) book - and describing its key elements. This best practice guidance is designed for both analysts and for those commissioning analysis and, as such, is a document of key top-level principles, checklists and pointers to potential issues along with how they can be recognised and mitigated.

• Providing an overview of the key ongoing developments and research areas that are being pursued so as to ensure that the guidance remains relevant and up-to-date. A key element of this is the establishment of a repository of more detailed guidance that builds on the Aqua principles and provides more specific guidance tailored to specific eventualities and/or modelling paradigms. Documentation at this level is essentially similar in nature to more tactical guidance present in various standards and documents such as under the Simulation Interoperability Standards Organisation (SISO), and the Institute of Electrical and Electronic Engineers (IEEE).

It should be noted that, although both authors are from the Defence Science and Technology Laboratory (Dstl), which is an Agency of the UK Ministry of Defence (MoD), they have been involved in implementation at the cross-government level, within MoD and, more parochially, within Dstl itself.

2 MACPHERSON REVIEW AND SUBSEQUENT PROGRESS

The original review under Sir Nick Macpherson was supported by a senior level steering group, a cross-Government Working Group and a review implementation team. Work was undertaken across all Government Departments to: identify business critical models (see further below) and associated approaches to quality assurance (QA); calibrate best practice against public and private sector organisations and professional bodies; and to produce overall recommendations.

In broad terms, the review concluded that most components of best practice in modelling QA fall under two headings: modelling environment (including culture), and modelling process. In particular:

• **Environment**: creating the conditions in which QA processes to deliver *quality of the outcome* (Robinson, 2002) can operate effectively, including through a culture that values QA and welcomes challenge. And, providing a well understood chain of responsibility and sufficient time for QA. It also requires adequate capacity, including specialist skills and sufficient time to conduct QA effectively.

• **Process**: establishing a clear process for every stage of the model life-cycle. This includes addressing *quality of process* (Robinson, 2002) issues through working alongside the customer to ensure there is a shared understanding about the purpose and any limitations of the model or models used. It is also about a systematic approach to make QA accessible, easy and comprehensive. This requires clear guidance on QA and clear documentation for every model.

Additionally:

• Models have a life-cycle – scope, specify and design; build and populate; test; deliver and use; and, finally, decommission - and therefore QA activities need to be tailored and progressive. Equally, models in this sense include all elements of the overall modelling approach, including not just the software itself, but also the data with which it is populated and the skills and experience of its users.

• The emphasis of the review was on those models that were deemed to be *business critical*, for example due to the extent to which the model drives key financial and funding decisions and thus where error could lead to serious financial, legal or reputational damage. However, all models and modelling

---

2 The intention to publish consolidated best practice guidance was discussed in Robinson and Glover (2014); however, the original working title – the Rainbow book – changed as the work matured.
activities need an appropriate level of QA; this needs to be proportionate to the decision being supported and any resultant risk associated with error or uncertainty.

Subsequent to the formal review, a cross Government working group remains in place to assist with coherence and the dissemination of best practice; in addition to leading on the assimilation and publication of best practice guidance as discussed in the next section, this has for example led to a number of modelling school initiatives to build on and socialize the published guidance.

Many Departments, in particular those with a significant level of modelling effort such as MoD, set up their own working groups to support the review and subsequently to implement its findings; indeed the resultant MoD action plan was independently assessed by its Defence Science Advisory Council (DSAC) (Rawle and Johnson, 2014) to provide assurance that it would deliver effectively. Additionally, cross-Government follow-up activity – a “year-on” review - was commissioned to monitor implementation progress (HM Treasury, 2015b), which demonstrated that the necessary changes are being embedded within Departments, although full implementation is not yet complete in all cases.

3 THE AQUA BOOK

The Aqua Book (HM Treasury, 2015a) has been developed to outline a sensible, achievable set of top-level principles to help ensure that the analysis work conducted on behalf of government can be trusted to inform good decision making. This remit was chosen so as to both embrace the findings of the Macpherson review and provide guidance concerning the context within which analytical modelling is conducted, so as to promote the achievement of the outcomes that Government seeks.

3.1 Key Principles

The key principles set forth in the Aqua Book are:

- Proportionality, in order to appropriately support management of the risk associated with the intended usage of analysis work;
- Continuous assurance, delivered throughout the analysis process and supported by the recommended development of a number of key assurance products (Glover, 2014);
- Verification and Validation (Glover, 2014);
- Analysis with RIGOUR\(^3\), so as to position the analysis and its context holistically and thereby understand how much we can rely upon analysis for a given purpose.

Arguably, these top-level principles (see Figure 1) – albeit possibly couched in slightly different nomenclature – would be recognised by most, if not all, good modellers and analysts; however, codifying the principles remains worthwhile and supports broad and consistent application of modelling across Government. From a modelling and simulation perspective, the main thrust of the Aqua Book and its supporting material can be thought of as facilitating being “fascinated equally on the one hand by possible meanings, theories and tentative models … and on the other with the factual implications deducible from tentative theories, models and hypotheses” (Box, 1976, p792). The Aqua book also supports delivering this in a managed way, with clear demarcation of ownership and responsibility, so as to build upon the concepts set forth by Robinson (2002) as outlined in Robinson and Glover (2014).

---

\(^3\) The mnemonic RIGOUR was originally coined by Graham Mathieson, but has since been reworked for the Aqua book so as to better take account of the Post-Modernist critique (Jackson, 2003, pp 257 – 261) in order to achieve an appropriately holistic point of view. RIGOUR stands for Repeatable, Independent analysis, that is Grounded in reality, so as to be Objective, Uncertainty managed and Robust. In other words, something well worth delivering!
3.2 Clear Cut Roles

To achieve the intended quality of delivery (see Figure 2), the Aqua Book recognises three clear cut roles:

- **Commissioner** of the analysis, responsible for *quality of outcome*, whose role is to lead on the framing of the analysis and ensure appropriate stakeholder engagement so as to achieve acceptability of the work;
- **Analytical Assurer**, responsible for the *quality of the process* by which delivery is achieved in order to ensure the credibility of the product of the work;
- **Analyst**, responsible for *quality of content* through detailed conceptualisation of and engagement with the work.

**Figure 1 AQUA Book Principles**

**Figure 2 Quality of Delivery**
The Aqua Book provides material to support these three roles in its three constituent parts:

- **Part A** which has been written for commissioners of analysis and those that have accountability for programmes where analysis is important to successful delivery. This is written largely for the non-expert commissioner who nevertheless needs to understand what good, fit-for-purpose modelling looks and feels like and the types of questions and challenges that he/she needs to be aware of.
- **Part B** which has been written for the analytical assurer and the analyst. This provides the top-level principles that underpin successful, fit-for-purpose modelling and analysis work.
- **Part C** which provides an overview of the accompanying resources that are available. This provides further supporting detail, including via a web-based repository which continues to be developed and enhanced (see 4.1).

### 3.3 Theoretical Underpinnings

The following summarises some of the key theoretical considerations which guided the development of the Aqua book. At its heart the delivery of good quality simulation work, much like any other analysis, entails a journey from current belief to a better understanding of the issues that are explored. As shown in Figure 1, key to this journey is to: make sure that the right problem is being investigated and has been framed appropriately; confirm that the selected approach has the requisite discriminatory and explanatory power; and to make sure that the results and insights generated are free from any misperceptions or biases. For more complex parts of the simulation problem space this journey can have a sense of passing “through the looking glass” (Carroll, 1871). It is this realisation, encapsulated in the observations on “possible meanings, theories and tentative models” (Box, 1976, p792) which led to the reformulation of the RIGOUR mnemonic\(^3\). It is this process of phenomenological engagement, at a stage in the study process where by definition there are no certainties, simply a number of different claims and possibilities which proceeds model building for simulation (as for example outlined in Robinson and Glover, 2006).

To deliver good quality is to successfully engage with a diversity of perspectives concerning the simulation problem space, drawing together as a team so as to address the right problem. At its simplest this involves reconciliation of the views of the broader project team\(^4\), remembering that “essentially, all models are wrong but some are useful” (Box and Draper, 1987, p424). The following paragraphs outline the key considerations to be addressed in order to achieve such reconciliation while maintaining an appropriately holistic point of view. To do otherwise would be to risk a Type 3 error, literally to address the wrong problem, but in practice to address a complex issue from an overly narrow perspective.

The challenge of achieving holism, breaking through misconceptions, is often exacerbated by limited time series about which to reason (Foucault, 1969) which for a number of key shaping reasons has no simple resolution through ‘big data analytics’. These shaping reasons include the considerations raised by the Grue Paradox (DiFate, 2007), such that at the limit all extant information series prior to those derived directly from the current theoretical claims are rejected because the world is believed to be about to change (e.g. Moffat and Gardner (2006)). Particular care is needed here to shape simulation experiments such that they are more than a simple mirror of belief unless the aim is to simply explore what-if? Instead simulation work generally requires a deeper claim to being founded, if it is to be relied upon to support decision making. This in turn raises broader perspectival questions concerning intentionality (Fuenmayor, 1991, pp 473-474). Specifically, the way in which experts unfold their understanding is based on what they believe to be relevant when asked\(^5\) (Husserl, 1982; Polanyi, 1998). This phenomenon has been both empirically described (Lotka, 1926; Bradford, 1934; de Solla Price, 1986) and theoretically explained (Zipf, 1949). This means that experts need to be kept engaged at all stages of the simulation project since what they are

---

\(^3\) Here the broader project team is understood to include those in the analyst, analytical assurer and commissioner roles along with the key stakeholders.

\(^5\) This consideration particularly applies when seeking to work ‘at pace’ within a situation demanding a rapid but none the less considered response.
exposed to can lead to further clarifying revelation, which can in turn significantly change the understanding of those engaged in the work, enabling appropriate characterisation of the pluralistic perspective required to support evidence based decision making.

At the limit, even the simplest validity questions concerning phenomenological engagement are complex to resolve due to the complexity of the challenges faced by government, where understanding of process is often under resolved and deep understanding incomplete. Thus the analyst seeking to engage with detailed questions of criterion validity is faced by dilemmas and uncertainties (Derrida, 1992, pp 26-28) including those raised by the Ravens Paradox (DiFate, 2007) such that at the limit all available information could be interpreted as evidence of pre-held belief, making it challenging to both identify and break misconceptions down. Here the concept of RIGOUR, used to provide challenge from a post-modernist perspective, helps to promote both self-awareness and understanding of others’ stance, so as to achieve a properly holistic and appropriately impartial view (Jackson, 2003, pp 257-261). Thus allowing the broader project team to reflect upon and understand:

- The risk arising from power / knowledge connections which demands that the simulation work takes the broader project team upon a journey of discovery in order to achieve validity;
- The shaping power of ‘Grand narratives’ found in the accepted literature, which at their heart usually have unrecognised assumptions which once engaged with enable new understanding;
- The limitations of technical discourse used by the stakeholder community, which while enabling rapid and precise communication within their area of expertise also conceals blind spots in their current understanding (including the systemic assumptions made in their phenomenological engagement (Kleindorfer and Ganeshan, 1993) and the need to manage the consequences which these raise);
- Awareness of sources of indeterminacy, seeking to mitigate these risks as far as is practical within project timescales and achieve ownership of the risks which remain by the broader project team.

The first inputs that the Analyst role receives in a complex situation concern current belief, where the belief shapes the perception of what is evidence (Kleindorfer and Ganeshan, 1993) which then re-enforces the belief (Jackson, 2003, pp 259-260) and this is often happening against a background where there is time pressure to act (appropriately). Here, stakeholder situation assessment pattern matching (Robinson and Glover, 2006) only considers a situation in terms of the lenses that they already use. This produces presentational difficulty with expressing unexpected findings from the work which can be resolved through finding an accepted lens which uncovers such alternative narrative, that the stakeholders have thus far forgotten to apply. In other words to connect with an autopietic narrative (Mingers, 2014) that is extant, in a way that properly satisfies concerns of RIGOUR and validity (Glover, 2014).

---

6 A council of perfection could be to engage in research concerning the understanding of everything that is or yet shall be. However, as framed, this is both unaffordable and since socio-technical systems evolve as their detailed contexts change, such ambition could never be more than a questing beast.

7 The concept of criterion validity examines the extent to which the work conducted actually engages with the phenomena that it claims to (Glover, 2014, p11).

8 See the previous paragraph as to why the stakeholders, as experts in their perspectives, may not necessarily realise the full implications of what they know or understand, without assistance from a simulation study process.

9 At its simplest an autopietic narrative is one on which a stakeholder relies in order to situate their understanding of their place within the world.

10 In extremis it may be necessary to simply map who will benefit, who will be non-plused and who will suffer given the choices that are perceived, allowing the decision maker to exercise their judgment call.
3.4 Implementation Practicalities

While appropriately responding to the considerations identified above and the advice so far issued through the Aqua book will significantly reduce the risks of failing to produce good quality analysis it is important to understand the key components of quality assurance so that we know that this is so. A number of approaches to information capture are therefore recommended in support of the RIGOUR mnemonic so as to achieve an analytical transparency from a simulation project (Glover, 2014):

- Repeatable: Analytical Estimate\(^\text{11}\); Master Data and Assumptions List; Technical Reporting. The key issue being resolved by this documentation is to appropriately map the assumptions necessary to deliver the required pace of the work.
- Independent analysis: Concept of Analysis. Setting out the purposes of the analytical means being used to conduct the analysis and the role of the simulation aspects within this scheme.
- Grounded in reality: Verification and Validation Logbook. Setting out the case by which it is claimed that the simulation work has been verified, demonstrating that it has been done in the right way and validated, setting out the claim that the right work has been done.
- Objective: Reflective Journal. Mapping the evolving understanding of those in the analyst role such that emergent evidence from the work can be noted, apparent ambiguities engaged with and as far as possible understanding resolved.
- Uncertainty managed: Presentations, Key Summaries for the Busy Reader. Drawing out the key things that the decision maker needs to know, both in terms of findings and also remaining parts of the associated risk space (typically expressed as other contributory factors) which they need to understand and own.
- Robust: Customer Reporting. Formally capturing both the boundaries of the simulation work and what has been determined in order to assist the decision making.

3.5 Uncertainty

In addition to consideration of uncertainty as a general part of attention to RIGOUR, the Aqua book also contains explicit guidance on uncertainty and its management. This is because decision-makers need information on the uncertainty of decision outcomes, that is the range of outcomes that may occur and their relative likelihoods, in order to act appropriately and with confidence in using the modelling to support decision-making. In particular, this requires early identification of key causes of uncertainty whether that be in the form of uncertainties associated with data and assumptions used in the analysis; changes in the wider environment; or unexpected events or risks that may influence the outcome. In the context of simulation it is particularly important to understand which aspects of the problem are intrinsically uncertain (some call this variability) and which could in principle be tackled for example through additional targeted data capture, via greater run sizes or through better designed computer experiments. Of course, whether such reductions in uncertainty are practical will depend inter alia on the time, cost and effort involved in any remedial action. Equally, even if it is impractical to address some issues in the specific study – for example due to time or budget constraints – the requirement to tackle those issues can usefully inform a longer-term research agenda so as to better position modelling capability in support of future customer questions. Ultimately, however, it is incumbent on all analysts (and assurers and commissioners) to recognise that absolute certainty is rarely achievable in most real-world situations and thus management and communication of uncertainty is a key component of good modelling and analysis.

\(^{11}\) The purpose of an analytical estimate is to establish a common understanding between commissioner and analyst concerning the customer question to be addressed and the type of product jointly agreed to be appropriate.
4 CURRENT DEVELOPMENTS

Although the Aqua book is now available, there is an ongoing need to ensure that best practice is maintained and developed; and that it is shared and used! Efforts both across Government and within the MoD are, therefore, concentrating on both the enduring cross-cutting communities required and on further developing guidance in key areas as briefly discussed below.

4.1 Continuing Communities

Across Government, the community that produced the Aqua book remains in place and meets regularly. Three key areas are worthy of specific comment:

- First, the community has a key role in training and education at all levels. This includes senior policy staff, in particular commissioners, who need to understand how and where modelling can be used effectively, and has included development of a modelling school for such staff. Similarly, analytical assurers and analysts also need access to suitable training resources to supplement the Aqua book and these are also being put in place.
- Second, has been the adoption of some cross-government approaches to sharing, currently via an Analytical Professions Network (APN) hosted on a cross-Government platform which enables promulgation and socialization of best practice and, for example, enables staff in smaller Departments to seek advice and input from across the broader community. These networks were used, for example, to capture some of the best practice that is now contained in the Aqua book as well as to road-test some of the advice it contains before it was formally launched. Ongoing work is looking at opportunities to host the APN on more widely accessible networks so that key peers and external suppliers can also contribute to – and benefit from – the available material.
- Third, best practice is rarely static, although the top-level principles are well established. Therefore a growing raft of more detailed information is being developed and shared – building on the original Aqua principles and shared through an open website.

4.2 Current Issues

Key areas in which best practice advice continues to be developed include:

- Third Party Assurance. In many cases, Government modelling is delivered through external suppliers and thus there is an ongoing issue of how to ensure appropriate adherence to Aqua principles in such cases without having to resort to “man-marking” in review (that is where external suppliers assure their own work and Government also conducts exactly the same reviews). Cascading Aqua guidance effectively, whilst still tailoring it to the specific situation in hand, is thus important and can be exacerbated in cases where a chain of supply is involved such as where an external supplier acting as a prime-contractor itself commissions work from other contributors as part of its overall response.
- Through-Life Tracking of Quality. Another key development area is to take some of the advice in the Aqua book and to develop key supporting resources and templates that make the advice easy and practical to follow in a consistent way. Thus, for example, prototype work is in hand on a through-life quality tracker as a means of capturing and recording agreements and progress against a small set of key headings – has commissioner agreement been gained to the proposed scope and bounds of the analysis; have key data sources been identified; what validation and verification status do the proposed model(s) have, and so on. Monitoring such aspects through the course of a project can ensure that any key challenges can be identified early and monitored so that any remedial action can be taken.
Robinson and Glover

- Work is also in hand on a more explicit evidence framework based on Cynefin principles (French, 2012); this is being developed to assist both analysts and commissioners in scoping study questions and also in shaping an appropriate modelling response.

5 CLOSING COMMENTS

Supporting the full range of diverse and high-impact decisions across Government requires fit for purpose models and modelling supported by sufficient and appropriate assurance activity to generate confidence in the results and insights used in support of significant decisions.

The key recommendations from the Macpherson review are being successfully implemented across Government in particular by means of the Aqua book and through ongoing activity to ensure that the guidance and best practice it contains remains relevant, up-to-date and used.

REFERENCES


**AUTHOR BIOGRAPHIES**

**ALAN ROBINSON** has been working in Defence Operational Analysis for over 30 years. He is currently a Dstl Fellow of the Defence and Security Analysis (DSA) Division of Dstl, which is an Agency of the Ministry of Defence, advising the DSA Capability Lead on the technical quality of the work delivered by DSA Division.

**PAUL GLOVER** received a BSc (Hons) in Archaeological Sciences from the University of Bradford in 1985. He then completed an MSc in Information Processing at the University of York in 1986 and has been working in Defence Operational Analysis ever since. Paul is currently a principal analyst in DSA Division at Dstl, with responsibility for the quality assurance, technical development and delivery of modelling and simulation methods and techniques.
AN AGENT-BASED MODEL OF KNOWLEDGE TRANSFER: EXPLORING THE NEED FOR CLOSURE AND COGNITION

Dr. Duncan A. Robertson
Loughborough University
Loughborough, LE11 3TU, UK
d.a.robertson@lboro.ac.uk

Prof. L. Alberto Franco
Loughborough University
Loughborough, LE11 3TU, UK
l.a.franco@lboro.ac.uk

ABSTRACT

We set out a model of inter-team knowledge evolution through inter-group interaction. We investigate individuals’ Need for Closure and Need for Cognition and the effect on the quality of the group’s decision. We set out a description of the model and representative results.

Keywords: agent-based model, knowledge transfer, decision making, social networks

1 INTRODUCTION

Agent-based modelling and complexity science have been cited as possible solutions for investigating team interactions. At its heart lies the interaction between human agents within and comprising a complex system of interactions.

Agent-based models have been thought of as dividing into two camps: as a method for studying the dynamics of social systems (the ‘microworld approach’), or as a type of boundary object (Elsawah et al 2015). The use of agent-based models within operational research has been limited, with debate existing as to the use of agent-based models rather than actually building agent-based models. While the microworld and boundary object perspectives are not necessarily mutually exclusive, this paper takes the former approach: modelling the actors within the social system.

2 MOTIVATION

A recent stream of research has studied the interactions between individuals within workshop settings (Tavella and Franco 2014), groups (Arrow and Henry 2010), and dialogue (Tsoukas 2009). We combine this with agent-based approaches of firm interaction (Robertson and Caldart 2009) and the dissemination of culture (Axelrod 1997) to create a model of inter-personal interaction within workshop settings.

We instead use an agent-based simulation approach to explore the Need for Cognition and the Need for Closure that play a part in group decision making.

Need for Closure (NFClos) (Kruglanski and Webster 1996) is a desire for information, where individuals are goal oriented. High NFClos individuals are typically inclined to (Kruglanski and Fishman 2009): attain closure as quickly as possible, and maintain it for as long as possible; they achieve this by relying on past knowledge and avoiding new information.

The Need for Cognition (NFCog) however shows a tendency to engage in and enjoy cognitive efforts (Cacioppo and Petty 1982). High NFCog individuals typically: care about how and why something works, not only that it works; are more likely to engage in information seeking, seek more information, and evaluate information more thoroughly.
Previous work on the modelling of group process modelling have included Larson’s (2007) $N$ dimensional problem space model where heterogeneous groups produce better solutions. This work is restricted to a Boolean search space with little interaction between members. Rousseau and van der Veen’s (2006) work shows the emergence of a shared identity but with a limited number of possible outcomes and with the restriction of being a cellular automata model.

3 AGENT-BASED MODEL

We first generate a fitness landscape (Wright 1932, Levinthal 1997) over which agents can move. This is created by adding $M$ Gaussians at random positions. Figure 1 shows a fitness landscape for $M=1$ – one central peak.

![Figure 1: Fitness Landscape](image)

We then add $N$ participants – agents – who are positioned randomly on the fitness landscape. In our model, the fitness of the solution is represented by the height of the landscape.

Each participant has a NFClos boundary – if another participant is within this boundary, the participants will ‘seize and freeze’ (Kruglanski and Webster 1996). Participants who have not frozen will continue to move around the fitness landscape. This movement is confined to the NFCog radius. Figure 2 shows (in purple) the NFClos radius, and (in red) the NFCog radius. The grey dots show areas that are searched as part of the search process, and the brightness of the green shows the height of the landscape (this is another view of Figure 1, except viewed from above).

Another parameter of the model is the ability to accept a participant’s viewpoint. We model this by setting a minimum threshold for the height of the participant with which the interaction is taking place. If their height is below the threshold, they will continue to search.
If another participant is not found when the locations within the NFCog radius are searched, the participant will move to the highest location that they have searched within the NFCog radius. This repeats until all participants are ‘closed’ or after a certain time period.

### 3.1 Parameters of the Model

Model parameters include:

- **Number of participants**: the number of participant agents that are initialized. This number remains fixed over each simulation run;
- **Need for closure radius**: at each tick of the simulation run, a participant will search for a partner within this radius;
- **Need for cognition radius**: if a participant is not closed, they will search a radius around their initial location;
- **Height threshold**: if a participant detects a partner, they will have to meet this height requirement in order for the agent to be closed.
3.2 Model Logic

The model logic is shown in Figure 3 below.

![Model Logic Diagram]

**Figure 3 Model and Agent Behaviour**

4 RESULTS

By controlling the number of participants in each run of the model, we are able to determine the effect on the quality of the group’s solution (measured by the average height of participants on the fitness landscape).

4.1 Effect of Need for Closure

Figure 4 shows the effect of varying the participants’ Need for Closure. Increasing Need for Closure decreases the best solution, which is more pronounced with more participants.

4.2 Effect of Partner Height Threshold

Experiments were conducted to determine the effect of varying the height threshold. As can be seen from Figure 5a (for two participants) and 5b (for four participants) below, the results for a zero threshold revert to the results shown in Figure 4.

For two participants (Figure 5a), when the threshold is set at 100 (the maximum height of the fitness landscape), the results are largely independent of the Need for Closure. This is due to the participants being trapped in a loop of dissatisfaction – where they will continue to search. This leads to local movement around the landscape peak.

For four participants (Figure 5b), the efficacy of the threshold diminishes. This is due to the participants requiring only their partner to be above the threshold. When there are two participants, A requires B to be above the threshold and B requires A to be over the threshold; this mutuality is destroyed with more than two participants.
Figure 4 The Effect of varying Need for Closure

Figure 5a The Effect of varying Height Threshold with two participants
4.3 **Effect of Need for Cognition**

As the Need for Cognition radius increases, there is a decrease of height of the participants for height thresholds of zero and 100. This is due to the participants being unclosed at the end of the simulation, meaning that they are searching for a solution. As the Need for Cognition variable increases, the search area scales approximately as its square, meaning that participants are likely to be further away from the peak of the landscape, and as such the quality of their solution decreases. This is shown in Figure 6, below.

5 **DISCUSSION**

For the first time, we have produced a model that combines Need for Closure and Need for Cognition. The use of agent-based modelling allows us to extend these models to intra-group heterogeneity, and opens up a wide range of extensions to the model.

The results are consistent with existing understanding of the Need for Closure – that high Need for Closure individuals can get trapped into sub-optimal equilibria. The introduction of a threshold for partner solution height produces interesting results, particularly at very high or very low levels. This is particularly important in two-person situations, where the existence of mutual thresholds produce better outcomes.

Furthermore, the results show a counter-intuitive results where increased Need for Cognition (which extant literature sees as an individual asset) can result in lower results in a group situation.
Robertson and Franco

**Figure 6** The Effect of varying Height Threshold with four participants

REFERENCES


Robertson and Franco


AUTHOR BIOGRAPHIES

**DUNCAN ROBERTSON** joined the MSOM Group at Loughborough in 2014. Prior to that, he held posts at Warwick Business School (as Assistant Professor of Strategic Management), Manchester Business School, UEA London, Sun Yat-sen University in China, as well as holding a visiting post at The Wharton School of the University of Pennsylvania and teaching at London Business School. He is a Fellow and Trustee of St Catherine’s College in the University of Oxford. Before undertaking his DPhil at Saïd Business School at the University of Oxford, he qualified as a Chartered Accountant with KPMG working in the City of London, Australia, and New Zealand, where he specialized in advisory and corporate finance roles within the banking sector. He originally trained as a physicist at Imperial College London.

**ALBERTO FRANCO** is Professor of Management Sciences at the School of Business and Economics, Loughborough University. With a background in civil engineering, he worked as a consultant in the area of soil mechanics before moving to academia. During 1997-2000, he held research positions at the London School of Economics and Strathclyde Business School, followed by a three-year lectureship post at Kingston Business School. He then spent 8 years at Warwick Business School (2003-2011), first as assistant professor and, from 2006, as associate professor. During his time at Warwick he was Course Director of the MSc in Operational Research and Management Science (2006-2010), and led the design and launch of the MSc in Business Analytics and Consulting, also as Course Director (2008-2010). He was also a Visiting Professor for the 2007-2008 academic year at IE Business School (Spain). Prior to joining Loughborough, he was Professor of Problem Structuring Methods at the Centre for Systems Studies, University of Hull (2011-2013). Alberto joined Loughborough in 2013 as Head of the Management Science and Operations Management Discipline Group.
USING SIMULATION GAMING TO VALIDATE A MATHEMATICAL MODELING PLATFORM FOR RESOURCE ALLOCATION IN DISASTERS

Dr Sara Hasani
University of Bedfordshire
University Square, Luton, Bedfordshire
LU1 3JU
sara.hasani@beds.ac.uk

Dr Emel Aktas
Cranfield University
College Rd, Cranfield
MK43 0AL
emel.aktas@cranfield.ac.uk

ABSTRACT

The extraordinary conditions of a disaster require the mobilisation of all available resources, inducing the rush of humanitarian partners into the affected area. This phenomenon called the proliferation of actors, causes serious problems during the disaster response phase including the oversupply, duplicated efforts, lack of planning. In an attempt to reduce the partner proliferation problem, a framework called PREDIS (PREDictive model for DISaster response partner selection) is put forward to configure the humanitarian network within early hours after disaster strike when the information is scarce. To verify this model, a simulation game is designed using two sets of real decision makers (experts and non-experts) in the disaster Haiyan scenario. The result shows that using the PREDIS framework, 100% of the experts could make the same decisions less than six hours comparing to 72 hours. Also, between 71% and 86% of the times experts and non-experts decide similarly using the PREDIS framework.

Keywords: Decision-making, Disaster response, MCDM, Simulation game, Proliferation of partners

1 INTRODUCTION

The Multi-Cluster/Sector Initial Rapid Assessment (MIRA) report is released 72 hours to three weeks after a disaster strikes, and contains valuable information about various impacts of the disaster. The problem is that in the early hours after a disaster strike this report is yet to be released whilst most of the decisions about partners and emergency aids need to be taken. The PREDIS model is an attempt to bridge the gap in the early hours of the disaster aftermath where no official report regarding the human impact and the needs assessment required for disaster response decisions exist. Most of these decisions such as allocation of resources and selecting the partners need to be made before the MIRA report is released in order to prevent fatalities and the spread of disease and so on. This is the period where the PREDIS model is the most useful. The purpose of the PREDIS model is to provide a predictive evaluation of the disaster impact during the first 72 hours after any natural onset disasters. This provides the decision makers during this period with information, which can then be used to allocate the limited resources and recruit suitable partners. In another words this is a decision making platform for making more efficient allocation of the resources, sourced from the most suitable partners chosen for that type of disaster and that level of impact. In this paper the PREDIS model is used in a simulation process where the impact of using model in making decisions is compared to the decisions made without the model. In this hypothetical simulation game, there is an intention to investigate how well the model can assist the decision maker in making faster decisions. Also it is investigated that how well the model can assist the non-expert decision makers, make decisions similar to the experts. Simulation games are normally used in the situations when the re-creation of real situation is dangerous or immoral, namely military field or medical field. Disaster situation is also part of this real situations where creating a disaster for the purpose of validating the model is unethical if not impossible.
Hasani and Aktas

Instead the simulation game re-creates the real world in an observable situation where the decision made by each group of experts and non-experts can be analysed and compared in a safe environment. The research contributes to various areas of disaster management, decision-making and simulation game as well as serious games by being part of the new use of simulation game for validation instead of sole purpose of education. The paper begins with outlining the existing experience in simulation games to paint a picture of how other scholars used this tools to simplify their complex real worlds. Then the principles of the PREDIS model are briefly outlined The process of simulation game then is put forward before the result is analysed. The paper concludes with the summary and the limitations of the study.

2 THE EXISTING EXPERIENCE IN SIMULATION GAMES

Simulation as a validation tool itself is defined as a representation of a real-world environment, when the real system may not be observed directly because of the inaccessibility, cost and danger (Barton, 1994). Participatory simulation or simulation games bring the simulation into the experiential world of the players (Colella, 2000). In that sense the simulation games are the most dynamic forms of simulated decision-making models (including discrete choice, stated preference, and agent based modelling) because multiple agents are interacting with each other for decision-making and the effects of the choices are reciprocated by decisions of all agents (Anand, 2013). Although the simulation games are performed in various disciplines, scholars (Elgood, 1997; Lewis and Maylor, 2007) identify 572 simulation games in which almost half (222 games) are related to operations management. Simulation games, which have been used for validating decision-making models in various studies as, provided in table 1.

Table 1 Example of simulation designs

<table>
<thead>
<tr>
<th>Subject</th>
<th>Objective</th>
<th>Related work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Servitisation</td>
<td>Understanding the customer needs and define the scope of servitisation</td>
<td>Laine et al (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision-making</td>
<td>Strategic policy formulation</td>
<td>Oderanti and Wilde (2010)</td>
</tr>
<tr>
<td>under uncertainty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td>Literature review</td>
<td>Lopes et al (2013)</td>
</tr>
<tr>
<td>development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City logistics</td>
<td>Validating the proposed framework</td>
<td>Anand et al (2013)</td>
</tr>
<tr>
<td>Medical decision</td>
<td>Understanding the decision-making process</td>
<td>Mohan et al (2014)</td>
</tr>
<tr>
<td>Making</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agricultural policy</td>
<td>Analysis of the impact of a policy</td>
<td>Musshoffa and Hirschauerb (2014)</td>
</tr>
<tr>
<td>Economic behaviour</td>
<td>Analysing the Bounded Rationality Behaviour</td>
<td>Musshof et al (2011)</td>
</tr>
<tr>
<td>Mergers and Acquisition</td>
<td>Understanding in what economy the mergers and acquisitions take place</td>
<td>Thavikulwat et al (2013)</td>
</tr>
<tr>
<td>Strategic decision</td>
<td>Decision about evacuation modelling</td>
<td>Thompson and Merchant (1995)</td>
</tr>
<tr>
<td>Decision support system</td>
<td>Efficacy of simulation game</td>
<td>Ben Zvi (2010)</td>
</tr>
<tr>
<td>Supply chain</td>
<td>Effect of coordination risk on bullwhip phenomenon</td>
<td>Croson et al (2014)</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the use of simulation games in the variety of studies; for example in medical treatment (Reichlin et al, 2011) to investigate the effect of increase in the knowledge of the patient on the cancer treatment decision. Another example is the use of agent-based model for designing a serious game to
develop an entrepreneurial mind in the user (Gentile et al, 2014). Also simulation is used in cross-cultural decision-making (Madni, 2013) to promote the ad-hoc decision-making in supply chain. These decisions are basically made for a special and immediate purpose, without previous planning; which makes the result difficult to generalise.

It has also been used to validate a model about deciding the profit margin and supply partners in city logistics (Anand et al, 2013). Another example is related to the validation of the decision-making in land use (Villamore, 2013) to examine if rewarding the decision-maker for eco-friendly behaviour instead of profit-oriented behaviour affects the decision. In the present research a simulation game is used to highlight the strengths and weaknesses of the model and to justify its usage in the real-world situations. To that end the simulation game is embedded in the treatment phase of a quasi-experiment design as is explained in the next section.

3 THE PREDIS MODEL

The PREDIS model formulates an MCDM platform in combination to linear programing to optimise resource allocation. The input data are drawn from a panel analysis of historical records (4,252 natural onset disasters between 1980 to 2013). This data then is examined using a pattern recognition technique to predict the human impact of the disaster (fatality, injured, homeless) with up to 3% prediction error. This enables the decision makers to estimate the required needs for each disaster and prioritise them based on the disaster type and socio-economic situation of the affected country. It also allows to rank and optimise the desired partners based on the decision maker’s preferences. The logic behind this attempt is that if a predictive technique could be developed to approximately estimate the human impact, and the needs of the affected population immediately after the disaster strike, then it is possible to combine the decision makers’ expertise and experiences with the data about the available resources to enable decisions about which partner should meet which requirements. The focus of this paper is on the verification of above model using a simulation game. The PREDIS model in the simulation game is used in the treatment phase of the quasi-experiment design.

4 SIMULATION GAME

To validate the PREDIS model two hypotheses were considered:

Hypothesis 1: ‘The PREDIS model assists the decision makers in making the similar/comparable decisions faster’

Hypothesis 2: ‘The PREDIS model assists the non-experts in making decisions as well as experts’

If these hypotheses are confirmed, it can be said that the PREDIS model not only helps the experts to decide faster but also helps the non-experts in making decisions as well as the experts. These hypotheses were examined in a quasi-experimental design (Figure 1) where the simulation game is observed as the treatment to a quota sample selected equally from a mixed population of experts and non-experts. The use of the experimental design in simulation games is popular due to its resemblance to the laboratory conditions (Simon, 1961; Norris and Snyder, 1982; Ben-Zvi, 2010; Musshoff, 2011; Croson et al, 2014). Experimental approaches are used in various studies where participants have to make entrepreneurial decisions within the systematically controlled rules of the game (Longworth, 1969, 2008; Tanner, 1975; Keys and Wolfe, 1990; Musshoff and Hirschauer, 2014).

Figure 1 shows that this empirical study (simulation game) contains three phases of the quasi-experiment design. It starts with the pre-test where the participants are asked to decide about a disaster scenario (here disaster Haiyan). At this point they are offered the data in the MIRA report and are asked to make decisions about the requirements and the humanitarian partners using their experience and judgement. The second step is the treatment where the participants are exposed to the principles of PREDIS model. In this model the human impact of the disasters is predicted using a rule based technique and the required aids are matched with the available humanitarian partners. The last phase or post-test is where both groups are
Hasani and Aktas

asked to make decisions using the principles of the PREDIS model. In this phase no information is provided other than the date, the type and the country of the disaster.

**Figure 1** Mapping the process of quasi-experiment simulation game

The simulation game process took three weeks to complete for 44 expert and non-expert participants. The expert participants were selected from volunteers who responded to a series of announcements from different humanitarian networks on the related conferences, organisations, LinkedIn special interest groups and humanitarian summons. They were gradually filtered after preliminary information regarding the process, their availability and their experience. The numbers in non-expert group then was matched to the remaining experts in order to make the two groups comparable.

5 PROTOCOL FOR RUNNING THE SIMULATION GAME

The protocol for the simulation game is built upon Sterman’s (1987) example of a simulated game such as a beer game for designing the instructions, strategies and debriefing questionnaires as well as the Van Sickle (1977) version of dangerous parallel for designing the implementation process.

The purpose of the game is twofold: To introduce and evaluate the model and to identify the weakness of the model in a real-world-like situation.

The goal of the first questionnaire is to define the characteristics of the desirable partners in the view of each decision maker and give each criterion the numerical preferences. The data about the characteristics of the desired partner in terms of the following is gathered: Type of the partner (governmental, NGO, International, Military or Volunteer organization), Size of the partners based on ANLAP’s (2012) categories for humanitarian organisations), Experience of the partners, Partner’s surge capacity (the ability to rapidly expand beyond normal capacity to meet the increased demand), Partner’s international expansion and ability to address the needs for humanitarian cluster being WASH, Nutrition, Health, and Shelter. These preferences calculated by AHP are combined with utility function for each resource as well as the utility of that resource for that partner to form an optimisation problem with two restrictions. First restriction shows that the total units acquired from all partners should not exceed the 100% of the total resources required. The second restriction shows that the number of the units obtained from the partner should not exceed the resources available to that partner. The partners then can be ranked based on their utility. For example, for
these particular participants, the utility of partners can be calculated and be used to rank the partners as exhibited in table 2.

**Table 1 Example of partners ranked / participant's preferences**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Partner</th>
<th>Total Utility</th>
<th>Type</th>
<th>Size</th>
<th>Expansion</th>
<th>Experience</th>
<th>Surge capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>15205</td>
<td>Government</td>
<td>Small</td>
<td>Yes</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>13716</td>
<td>Government</td>
<td>Small</td>
<td>No</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>13549</td>
<td>Government</td>
<td>Small</td>
<td>Yes</td>
<td>Low</td>
<td>Very high</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>13078</td>
<td>Government</td>
<td>Small</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>11643</td>
<td>Government</td>
<td>Medium</td>
<td>Yes</td>
<td>Low</td>
<td>Very high</td>
</tr>
</tbody>
</table>

Table 2 shows an example of the rankings of the partners based on this participant’s preferences. For example, Partner 5 is the most desirable partner with a utility of 1520. This also shows that the most desirable partners for these participants are small governmental entities. In addition, it seems that this participant does not value the experience or the surge capacity of the partners as critical requirements for a disaster response.

6 THE RESULT

To compare the result between the two groups of experts and non-experts, a variation of outranking method associated to Borda (Marchant, 1996) or Roy (Bouyssou, 2001) has been employed. The reason is that this is a classic MCDM problem, where a set of alternatives is selected based on preferences expressed by decision maker (Bouyssou, 2001). A common solution is to examine if for example, partner (a) is at least as good as partner (b). The outranking techniques under this rule have been used to support decision-making in voting (Jurij, 2006), supplier selection (De Boer et al, 2008) or project assessment (Nurmi and Salonen, 2008) amongst others. If a selection consists of a set D of Decision makers (here 22 decision makers for each group), each having a preference order for a set of C candidates (here 20 partners), the Borda rule here is calculated where a partner receives \( n \) points each time they are selected as the most desirable, \( n-1 \) points when they are selected second to most desirable, and no points every time they are selected as the least desirable (Russell, 2007). An example of these results for experts is exhibited in Table 3.

**Error! Reference source not found.** shows that the total Borda count for partner 1,2,3,4 is calculated as 144, 456, 326, and 500. This means that in this set, partner 4 is the most desirable in the overall view of the experts. The final results of the Borda counts are calculated for experts and are ranked in table 4.

Table 4 shows that based on the Borda count, for the group of experts, partner 4 who is a small military organisation with a high surge capacity, no international expansion, and low experience is the most desirable (with a 500 Borda count). Partner 9, who is a medium sized government organisation with no expansion, and a high degree of experience and surge capacity is the least desirable (with a 105 Borda count). The same process has been repeated for the non-expert group and the comparison of the result shows that for example the non-experts preferred partner 5 and 12 equally (333 Borda count) mostly because they are both small governmental organisations, with international expansion. It seems that the non-experts care less about the surge capacity and experience. Their least favourite are partners 20 and 17 with (a 135 and 136 Borda count), who are very big organisations with international expansion, and low surge capacity and experience.

As far as the comparison of first and last choices of the experts and non-experts reveals, there is no evidence that by using the PREDIS model these two groups make the same choices. However, the NRMSE has been used to calculate a more precise percentage of error between the choices of the two groups. The NRMSE for difference between the two is calculated as 29% (Error between non-experts and experts) and 14% (Error between experts and non-experts). This means that at least 14% and at most 29% of the times, the non-experts’ choices are different from the experts’ choices.
Hasani and Aktas

the experts. This also means although the first and last choice of the majority of decision makers in the two groups are not the same, between 71% and 86% of the times experts and non-experts decide similarly using the PREDIS framework. The significance of this result is that the non-expert does so with no prior training or data other than the data that are freely available on the Internet through the UN related and World Bank related websites (including HDI, DRI, population, population density, and disaster type). Therefore, it is possible to conclude that although the result shows that the experts and non-experts may have various preferences, the model enables the non-experts to choose partners similarly to experts, if necessary.

Table 3 Example of the Borda count for the group of experts

<table>
<thead>
<tr>
<th>Choice</th>
<th>Partner 1</th>
<th>Partner 2</th>
<th>Partner 3</th>
<th>Partner 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Borda</td>
<td>Frequency</td>
<td>Borda</td>
</tr>
<tr>
<td>1st</td>
<td>n</td>
<td>0</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>2nd</td>
<td>n-1</td>
<td>0</td>
<td>3</td>
<td>57</td>
</tr>
<tr>
<td>3rd</td>
<td>n-2</td>
<td>0</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>4th</td>
<td>n-3</td>
<td>0</td>
<td>9</td>
<td>153</td>
</tr>
<tr>
<td>...20th</td>
<td>n-19</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total Borda count</td>
<td>144</td>
<td>356</td>
<td>326</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 4 Example of the expert borda count ranking

<table>
<thead>
<tr>
<th>Partner</th>
<th>Borda count</th>
<th>Type</th>
<th>Size</th>
<th>Expansion</th>
<th>Experience</th>
<th>Surge capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>500</td>
<td>Military</td>
<td>Small</td>
<td>No</td>
<td>Low</td>
<td>Very high</td>
</tr>
<tr>
<td>5</td>
<td>427</td>
<td>Government</td>
<td>Small</td>
<td>Yes</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>356</td>
<td>Military</td>
<td>Small</td>
<td>Yes</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>... 9</td>
<td>105</td>
<td>Government</td>
<td>Medium</td>
<td>No</td>
<td>High</td>
<td>Very high</td>
</tr>
</tbody>
</table>

7 CONCLUSION

The study was put forward to examine two hypotheses.

The Hypothesis 1 inquired if ‘The PREDIS model assists the decision-makers in making faster decisions’. Firstly, it was initially expected that the majority of the participants have their own decision model. The result shows that even the experts who have frameworks in place (two of them were described earlier) mostly rely on heuristics of trust, previous experiences, self-declared resources, and capabilities. Therefore, the conclusion can be drawn that in practice a specific numerical and measurable guideline, which can clearly compare various partners, is missing. Secondly, it was initially expected that the majority of the participants make their decisions under six hours (the golden hour) in order to be able to perform the initial rescue operations. The result shows that without the PREDIS model, 23% of the experts take less than one-hour to make decisions, 45% take between 1-6 hours to make decisions, and 32% take more than 12 hours to make a decision. However, using the PREDIS model all the participants could make their
decision in less than an hour. This further confirms that the PREDIS model assists decision makers to make faster and comparable quality decisions.

Hypothesis 2 inquired that ‘The PREDIS model assists the non-experts in making decisions as well as experts’. Interpreting the result shows that as far as the comparison of first and last choices of the experts and non-experts reveals, there is no evidence that by using the PREDIS model these two groups make the same choices. However, the Normalised Root Mean Squared Error (NRMSE) has been used to calculate a more precise percentage of error between the choices of the two groups. The NRMSE for difference between the two is calculated as 29% (Error between non-experts and experts) and 14% (Error between experts and non-experts). This means that although the first and last choice of the majority of decision makers in the two groups are not the same, between 71% and 86% of the times experts and non-experts decide similarly using the PREDIS framework. The significance of this result is that the non-expert does so with no prior training or data other than the data available through PREDIS model. Therefore, it is possible to conclude that although the result shows that the experts and non-experts may have various preferences, the model enables the non-experts to choose partners similarly to experts, if necessary.

The aim of this paper was to test the suitability of the PREDIS model for decision-making in the disaster situation. Using two series of expert and non-expert participants, a hypothetical scenario of a previous disaster was re-played. The decisions the two groups made were recorded and compared. The results of the decisions show that the PREDIS model has two major capabilities. It enables the experts and non-experts to predict the disaster results immediately and using the widely available data. It also enables the non-experts to decide almost the same as the experts; either in predicting the human impact of the disaster and estimating the needs or in selecting suitable partners. It is also the only framework of its type, which takes specific numerical values as input, and provides specific numerical values and clear decisions as outputs such as which partners should supply how many units of the requirements.

However, there are some limitations associated with the PREDIS model. First, it is purely theoretical at the moment and has yet to be tested in a real disaster situation. Second, the data used for estimating the needs have been accumulated from various sources and their applicability in the actual disaster scenario might differ for each organization. Hence, further research is required to test the PREDIS model in a real disaster situation.

REFERENCES


Marchant Th (1996) Valued relations aggregation with the borda method Journal of Multiple Criteria Analysis 5:127–132
Norris D and Snyder C (1982) External Validation of Simulation Games Simulation and Gaming 13(1) :73–85
Simon H A 1961 Business games Business Horizon Spring 6–9

AUTHOR BIOGRAPHIES

SARA HASANI is a lecturer in Business Systems in Bedfordshire university and a member of UCL Institute for risk and disaster reduction (IRDR) and Operational Research Socceity. Her research focuses on various aspects of operations research, decision support system, risk management and serious games in the context of disaster management with the focus on segregation between short term and long term collaborative networks.
EMEL AKTAS is a senior lecturer in logistics and supply chain management in Cranfield University, School of Management and a member of the Operational Research Society. Emel’s applied research is informed by mathematical modelling, simulation and optimisation to develop decision support systems for policy and decision makers on decisions at strategic, tactical and operational levels. Her current research focuses on decision making at operational level in the context of manufacturing, supply chains, and humanitarian logistics.
THE IMPACT OF POLICY CHANGE ON MARITIME LOGISTICS
LEVEL OF DISASTER PREPAREDNESS AND RESILIENCE

Dr. John K. Buor
Ghana Institute of Management & Public Administration
P.O. Box AH50, Achimota – Accra, Ghana
kwebuor@yahoo.ca

ABSTRACT
This research employs system dynamics (SD) modelling to analyse the structural behaviour of the interactions between Disaster Preparedness, Environment Instability, and Resilience in maritime logistics industry in response to policy change. Despite the evidential rise in frequency, magnitude, and disruption potentials of catastrophic events in recent times, it appears that industry stakeholders are not able to anticipate the effects of long-term strategic risk management decisions.

Field data and the dynamic models have revealed that there is a strong influence relationship (interdependencies) between Disaster Preparedness, Environment Instability, and Resilience in a logistics/supply chain network. We also found that policy interventions geared towards risk management have the potential to produce unintended consequences basically due to unacknowledged conditions. The research model provides strategic policy makers with real-time decision evaluation tool that can enhance justification for acceptance or rejection of a complex risk management intervention prior to decision implementation.

Keywords: disaster preparedness, environment instability, resilient logistics, system dynamics

1 INTRODUCTION
Several researchers (Craighead et al., 2007; Tang, 2006a; Wagner and Bode, 2006; Zsidisin et al., 2005) assert that supply chain managers do implement initiatives to boost revenues, cut costs, and reduce assets. Obsessed with these strategic management policies, coupled with the perceived environmentally stable conditions enjoyed (in the case study area), stakeholders in risk management may begin to ignore operational safety and sustainability issues, leading to increasing tendency by investors and supply chain managers to reduce the amount of resources and time invested to promoting disaster preparedness.

This paper therefore is set out to find out the effects of frequent policy change on the interdependency between disaster preparedness, environment instability, and the resilience of a maritime logistics network using SD modelling tools. Findings suggest that the level of disaster preparedness and resilience of a logistics system is contingent on the size and physics [environment] of the network.

2 LITERATURE REVIEW
Maritime logistics industry is very complex yet it plays a very significant role in regional growth and economic development (see Polyzos et al., 2008). Such roles are dependent on the supply chain strategies adopted by those who use a particular port facility in a particular location. Heaver et al. (2000) mention the need for continuous role redefinition by ports and port authorities in order to guarantee that the system remains a fully-fledged player in modern day fast evolving markets. However the redefinition of roles may lead to frequent change in policy which can be unintendedly costly due to increasing policy risks.
Under the “theory of structuration”, Giddens (1979) acknowledges that social life is more than random acts that are merely determined by social forces. Giddens contended that human agency and social structure [environment] relate to each other such that the repetitive acts by individual agents [in an environment] can produce structures in the form of traditions, institutions, moral codes, and established ways of doing things. Changes can occur to the social structures when agents start to ignore, replace, or reproduce them in a different way. It is these consequences that we attempt to evaluate using the SD models to analyse the structural behaviour of the interactions between environment instability, disaster preparedness, and resilience, as different policy changes (strategic interventions) are tested on a logistics network.

2.1 Trends in maritime transport and the impact on the environment

The growth in world trade in the past decades and the increasing emphasis it places on maritime transport efficiency requires that port owners and shippers will remain alert and be ready to tackle potential causes of disruptions in their logistics chain. Globalised production, increased reliance on sea transport for import and export (close to 90% of global trade), technological changes and increase in vessel size (‘Gigantism’

1

), port specialisation, and port decongestions are some of the current evolutions in the maritime logistics industry.

Synchronously, the maritime industry has witnessed increases in regulations aimed at port and transport sector decongestions in recent years, due to their environmental impact and port security (Psaraftis, 2005). The post 9/11 era in particularly, has seen an increase in regulations and dominance of environmental sustainability issues at world forums, besides the shift in port ownership and management (Brooks and Cullinane, 2006). Many Governments are withdrawing from port operations and are rather concentrating on monitoring and oversight responsibilities (Baird 1995). See Bailey and Solomon (2004) and Tzannatos (2010) for relationship between the growth of maritime transport and marine environment pollution; for effects of port dredging (see Gupta et al., 2005; Goulielmos and Paldali, 1998); and for effects of ship emission on ecological sites (Geircke, 2003; Isakson et al., 2001; Trozzi et al., 1995). Apparently, the industry-wide revolution can lead to increasing cost of operation, increasing influence of harsh weather, increasing human errors, and many associated risks that may call for regular policy change. Our research therefore intends to investigate the effects of frequent policy change on the interrelationships between environment stability, disaster preparedness, and resilience of a maritime logistics/supply chain network.

2.2 Resilience

“The adaptive capability of a supply chain to prepare for unexpected events, respond to disruptions, recover from disruptions, and maintain continuity of operations at the desired level of connectedness and control over structure and function” (Ponomarov and Holcomb, 2009).

From the above definition, the concept “resilience” seems to be linked with supply chain risk management strategies (interventions) and security. According to Kachali et al. (2012) the resilience of most organisations can be measured on the bases of: leadership, staff engagement, situation awareness, decision making, innovation and creativity, effective partnerships, leveraging knowledge, breaking silos, internal resources, unity of purpose, proactive posture planning strategies, and stress testing of plans.

2.3 Disaster

The number of disaster declared cases in the world is on the rise (Drabek, 1995; Murphy and Bayley, 1989), so also is the value of economic losses, and the number of disaster victims (Blaikie et al., 2004). These phenomena seem to explain the causes for the rise in environmental turbulence and crisis, and therefore need to be constantly evaluated in real-time.

The World Health Organisation (WHO) defines disaster as:

1Gigantism is a hub and spoke relationship between main ports that handle larger sized vessels and other ports in the neighbourhood.
“Disaster is a result of vast ecological breakdown in the relationship between man and his environment, a serious and sudden/slow onset disruption, on such a scale that the stricken entity [individual, community, organisation, society] needs extraordinary efforts to cope with it, often resulting in dependence on outside help or international aid” (WHO).

One may also read McFarlane and Norris (2006, p.16) for classification of disaster into acts of God, large scale industrial accidents and episodes of mass violence. Also, for the definition and categorisation of risk and related terms, readers may refer to Christopher and Peck (2004); Juttner, Peck, and Christopher (2003); as well as Manuj and Mentzer (2008)

3 METHODOLOGY

Due to the complexity of our research objective, we adopted the positivist (quantitative) epistemology (i.e. reality is objective) using Euler’s numerical integration methods to computer simulate [analyse] the structural behaviour of the three state variables considered in the abstract. At the data collection, we employed the interpretivist (qualitative) epistemology (i.e. reality is perceptual) using personal in-depth semi-structured face-to-face interviews with seven (7) individual Chief Risk Officers (CROs) who were drawn from the Port Authority, Port Operators, Liner Agencies, Transporters, and some Academia (who are knowledgeable in marine, coastal, and estuarine studies). We purposely selected this category of elites for interview because we assumed that they hold key information about maritime security, emergency, and risk management. In addition to their daily as risk managers, they could be key players in policy formulation and implementation concerning disaster planning, crisis and risk management. We put forward seven (7) interview questions that covered four thematic areas: risk identification and assessment; current risk/disaster management strategies and procedures; expected changes in the port industry in the next few years (three, five, or ten years) and the potential effects that such changes can have on the maritime logistics industry; and the possible consequences of today’s planning strategies towards future events (scenarios of policy change). Thus quantitative methodology became the primary approach (Morgan, 1998); implying that we employed qualitative methods first in data collection, followed by a qualitative (thematic) data analysis, which we capped with quantitative (simulation, or sensitivity) analysis and interpretations.

We extracted statements such as the ones below from our interview respondents:

“....the more prepared” one is, the “more resilience” one can be; or
“....increased risk preventive compliance measures” can improve “port stability”
“....increased risk preventive compliance measures” can “improve disaster preparedness”; or
“....more policies, sanctions, directives, laws and regulations (risk preventive compliance measures)” can lead to “improvement in port environment sustainability [i.e. reduce port instability]”; or
“....increased disaster awareness and preparedness” consequently “increases resilience in port logistics chain” or
“Resilient logistics operation is depends on environment stability and level of disaster preparedness”
....etc.

These extracts from the raw textual data helped the research to create themes and sub-themes that enhanced the creation of links and structures such as figure (1 and 2) in section (4). Data was processed and analysed applying steps in Grounded Theory’s (Glaser and Strauss, 1967) coding processes (see Corbin and Strauss, 2008; Straus and Corbin; 1990; 1998; Kim and Anderson, 2012; Kopainsky and Luna-Reyes; 2008). SD simulation runs were performed for a100 day period. We select SD modelling because it offers stakeholders a platform to share ideas through scientific debates, to formulate, explain, and effect long-term complex policy changes (Barlas, 2007) with minimum (or no) resistance.
4 SYSTEM DYNAMICS (SD) MODEL/SIMULATION

According to Forrester (1961), SD is “the study of information-feedback characteristics of industrial activity showing how organisational structure, amplification [in policies], and time delays [in decision and actions] interact to influence the success of enterprise”. SD represents real-world processes in terms of stocks (e.g. material, knowledge, money, people, etc.), flows between stocks, and information that determines the value of the flows (Borschchev and Filippov, 2004; Forrester, 1968; Tako and Robinson, 2009). The primary assumption of SD is that the dynamic tendency of a system is determined by the internal causal structure of that system (Meadows and Robinson, 1985). Therefore we propose that the level of disaster preparedness could determine the response and recovery interventions (i.e. sustainable of maritime logistics functions and operations) to employ in the HPC.

In order to form the building block for SD models (Richardson, 1996), and to derive the cause-and-effect relationships between variables that portray the behaviour in maritime industry, we employed the causal loop mapping (CLM) (see Eden, 2004; Williams, Ackermann and Eden, 2003; Morecroft, 1982; Wolstenholme, 1982) to simplify and transform verbal descriptions from our field data (interviewed respondents) into the conceptual model (figure 1) to help us identify and organise principal components and feedback loops of the system being studied (Goodman, 1974) as in figure 2 and 3.

![Figure 1 Basic causal links between variables](image1)

![Figure 2 The causal loop diagram connecting Resilience, Instability, and Preparedness](image2)

For instance, figure (2) suggests that (all things being equal) the more unstable the environment becomes the less prepared will be the logistics industry (-). On the other hand, the more prepared the logistics network is, the more resilient it will be against systemic risks and disruptions (+). Also, the more resilient we perceive a logistics network, the more activities we carry out and hence the less stable the environment becomes (-). The overall feedback loop is reinforcing (+) which requires serious attention. Such diagramming readily reveals the loop structure of the complex model (figure 3) to people who might
be unfamiliar with flow diagrams, or the DYNAMO notation. Since CLM and feedback loops can be employed to illustrate structure and behaviour of system over a time period (Binder et al, 2004; Wolstenholme, 2004; Senge, 1990; Richardson, 1996), we can therefore infer that CLMs (figure 2 and 3) illustrate what the problem owners think, and how they communicate their intentions to others qualitatively (Ossimitz and Lapp, 2006; Ossimitz, 2000). Such feedback loops have further allowed us to explain some non-linear relationships between key variables and also enhance the description of the behaviour of the SD graphs in section (5).

![Diagram](image.png)

**Figure 3** The integrated CLM indicating interdependencies in research variables extracted field data

### 5 MODEL TESTING

The time horizon for the simulation models is 100 days (i.e. approximately 3 months). We selected this time horizon based on the responses we got from the CROs concerning the expected duration of some disruption incidents they illustrated. The responses suggested various time horizons that are event dependent. However, by the nature of their activities, busy ports cannot endure any major disruptions in their logistics network even for one hour. We bear in mind that several modelling gurus (Meadow et al., 1974; Backus, 1996; Bunn and Larsen, 1997; Sterman and Richardson, 1985; Sterman, Richardson and Davidsen, 1990) have advised that time horizon should be selected on the basis of the concept that one is about to model and also based on the resources that need to be measured. The rule of thumb is that ‘time horizon should be set several times as long as the longest time delays in a system’ (Sterman, 2000). Sterman adds that time horizon should extend far enough back into history, and far ahead into the future, to enable one to show how a problem emerged and to describe the symptoms of the effects of strategic policy interventions. Therefore a long time horizon may be a critical antidote to the event-oriented worldview so that patterns of behaviour and the feedback structures that generate them can be identified. Thus it seems that 100 days is long enough for the system we are modelling to be able to learn whether the planned mitigation activity (policy intervention) has failed or otherwise. We presume that any time less or more than the 100 days may not yield a desirable result (see Sterman, 2000, pp.90-94). Below (figure 4 – 12), we analysed a few scenarios about the impact of change in technology capacity on structural behaviour of the maritime logistics industry for a 100 day time horizon and all other variables held constant.
All curves above appear to be non-linear. The trajectories of the curves (figures 4, 5, 7 - 10) show that DP and RPL declined sharply, reached their minimum points, and remained constant (at approximately 20
days). Contemporaneously, the inflow curve (Rate of increase in DP due to resource accumulation) rises very fast to reach a peak whiles the outflow curve (Rate of decrease in DP due to resource usage) declined sharply, reached minimum (approximately in 5 days) then rises fast to reach a peak then remained constant thereafter. We note that the inflow for resilience behaves the same way as outflow from DP since they are one and the same. At the same time that these curves exhibit their respective behaviours, the curve for environment instability (figure 6, 11 and 12) rises steadily towards the end of the period under consideration. Figure (3) to figure (12) describes the scenarios when technology capacity changes and other variables are held constant. The graphs suggest that there is strong influence relationship (or interdependency) between environment instability, risk/disaster preparedness, and the adaptive capacity to bounce back from disruptions (resilience) as the level of technology varies. As the trajectory of the curves of environment instability increase, those for disaster preparedness and resilience decline exponentially, suggesting that increasing environment instability (due to increasing maritime activities) influences levels of preparedness and resilience adversely. We observe some erratic (random) behaviour in the trajectory of graphs of disaster preparedness (figure 8) and resilience (figure 10) when technology capacity declines in the maritime logistics industry. It appears that technology change influences the flows in the variables under consideration. Furthermore, it appears that when disaster preparedness is unstably low due to irregular inflow of resource, it can potentially influence systemic resilience since unstable environment can make it highly difficult to match resource demand against supply.

Therefore we argue that resilience seem to be influenced by changes in the level of the disaster preparedness which is also influenced by the level of environment stability. The more unstable the environment becomes, the more unable it will become for the supply chain to respond to disruptions, and even the more difficult it will become for the system to bounce back from disruptions if they occur. The graphs seem to mimic real world situations and have apparently revealed that ‘the level of disaster preparedness and resilience of a logistics chain is contingent on the size and physics of the logistics network’. For example, by the location effect (i.e. port/vessel size or physics), industry stakeholders may recommend high redundancy (Sheffi, 2005), or take some disaster preparedness (risk prevention) policies as an attempt to overcome the impacts of some anticipated hazards. Such inventory which may include dangerous/obsolete materials may rather increase environmental hazards. Specifically ports rely on hard engineering (e.g. constructing concrete, hard surfaces, floor walls, and all-weather facilities) as flood preventive measures in their quest to dominate the difficult terrain. However, such constructions and extensions can lead to land reclamation/squeezing which may result in floods, and/or disruptions elsewhere in the natural ecological processes at the port’s immediate environs (far and near) thus causing both ‘spatial dimensional effect’ and ‘temporal dimensional effect’ (see Klink and Klink, 1995 in the proposed expansion of the Port of Rotterdam).

The above result arguably suggest that in order to attain resilience, management need to stabilise the environment by adopting good environmental management interventions as well as by ensuring change compliance. We support current policy change (intervention) in the European port industry, particular those directives that deal with River usage, and environment/pollution control regimes (e.g. MARPOL conventions) and especially those being enforced on the Humber Estuary including: environmental impact assessment (EIA); NATURA 2000; Habitat and Species Directive; Wild Birds Directive; the Appropriate Assessment (AP); Integrated Pollution Prevention Control (IPPC); Flood Management Control Directive; to mention just a few.

Because it offers stakeholders with a platform to share ideas through scientific debates, managers in the maritime logistics industry may employ these models to formulate, explain, and effect long-term complex policy changes (Barlas, 2007) if they intend to minimise change effector (stakeholders) resistance.

---

2 The spatial dimension refers to the expected effects of a project at one end of a region/country on the other end of another region/country and possibly that part of the continent.

3 The temporal dimension speaks about the expected life-span of a project. Most projects currently target 50-80 year time scale.
to policy change. Both policy-makers and policy effectors can readily understand the essence of change such that through this learning laboratory, they can anticipate plausible possible outcomes of policy interventions prior to decision implementation. Such dynamic complexity (Newsome, 2008; Wankhade and Dabade, 2006; Santo, Belton and Howick, 2004; Forrester, 1961) approach has been engaged by Kopainsky and Luna-Reyes (2008) as well as Sterman (2000) to structure the behaviour of complex dynamic systems through modelling, simulation, and feedback mechanisms within the field of systems thinking.

A few major challenges (limitations) of the approach we adopted was our inability to do exhaustive iterative data work as required in the Grounded Theory (GT) philosophy and also by the SD modelling approach. Generally, the SD modelling process requires the researcher to work in collaboration with problem owners, to build the model in stages so that there can be full participation by problem owners. But time constraint did not allow that to happen fully. Therefore, one will argue that the premature truncation of data collection (without going through all the stages) may have produced truncation error which can affect the randomly generated numbers that translated into the structure of graphs under the simulation exercises. However we do not believe that these can affect the interview results and the SD simulation models (graphs) that were based on field data significantly.

Further research may be done to enhance result comparability and theory falsification (such that it can lead to generalizability) about the interrelationship between environment, disaster preparedness, and resilience in the maritime logistics industry. We suggest that the model be tested on other major port complexes and industrial clusters elsewhere including: the industrial port complex of Rotterdam, the port complexes of New York and New Jersey, the Los Angeles port complex, the port complex of Guangzhou, the Busan ports, or the Ports of Kembla using the same data collection and analytic tool as the ones employed in this research.

### BIBLIOGRAPHY


Livey, P. (2005), "Inland transport by container: why think water?," Sea and Water Conference'.


Oliver-Smith, A. (1999), "What is a disaster? Anthropological perspectives on a persistent question." The angry earth: Disaster in anthropological perspective, No.18-34.


AUTHOR BIOGRAPHIES

JOHN KWESI BUOR received B.Ed in Mathematics and Science from the University of Cape Coast in 1999, and an MBA in Logistics and Supply Chain Management in KNUST in 2007 (both in Ghana). He completed PhD (Logistics and Supply Chain Management) at the Hull University Business School (UK) in 2015. He is currently a full-time lecturer and researcher at Ghana Institute of Management and Public Administration (GIMPA). John aspires to become an authority in logistics and supply chain risk, security, emergency, and disaster management.
MAXIMISING POLICE OFFICER RESPONSE VIA INFORMED DECISION MAKING

Johanna Leigh
Loughborough University
Loughborough, Leicestershire, LE11 3TU
j.m.leigh@lboro.ac.uk

Dr. Lisa M. Jackson
Loughborough University
Loughborough, Leicestershire, LE11 3TU
l.m.jackson@lboro.ac.uk

Dr. Sarah J. Dunnett
Loughborough University
Loughborough, Leicestershire, LE11 3TU
s.j.dunnett@lboro.ac.uk

ABSTRACT

Under the strain of public service funding cuts finding methods of utilising response officers to their highest efficiency, to ensure the public’s welfare, is a high priority. One concern is the method of dispatch, where current methods lack the necessary information to make an informed decision on the most efficient officer to send to an incident. The developed methodology includes the amalgamation of mapping, routing, and decision making criteria. The decision criteria used is based on incident severity, officer availability, predicted possible response time (which includes traffic conditions and time of day), demand coverage, and driving qualification. The methodology has been automated to maximize the chance of selecting the most efficient officer to attend an incident. Testing and validation of the methodology through the use of simulation has been undertaken, where it has been proved that making an informed decision can increase officer selection efficiency significantly over random selection of officers within the proximity of an incident.

Keywords: Decision Analysis, Resource Allocation, Police Officer Response, Vehicle Routing, Simulation

1 INTRODUCTION

Police forces play a key role in ensuring the public’s safety by responding to emergencies, performing neighbourhood patrols, monitoring criminal activity, making arrests, investigating crimes and testifying in court. Police Forces, along with many other public sector services, are being faced with major budget cuts from the government. This leads to an increasing need to operate more efficiently without jeopardizing standards of service. One area of focus is the first point of response from the emergency service, which is the dispatch process to send officers to incidents reported. The incidents reported are typically classified into four grades; Grade 1 incidents are emergencies (e.g. where there is danger to life) where response is required within fifteen minutes; Grade 2 incidents are priority incidents (e.g. involving vulnerable people) which require response within sixty minutes; Grade 3 and 4 can be resolved by scheduled appointments or over the phone. Grade 1 and 2 incidents require a fast response to ensure public safety and these incidents are handled by a dispatchers who allocate chosen officers to incidents. The aim of dispatchers is to reduce
response times and increase officer availability whilst maintaining demand coverage in high demand areas. A common issue for the dispatcher is the lack of information upon which to base decisions on. There is much more information which could be relevant to the decision making processes which is not currently considered. This leads to uncertainty in decisions which often leads to dispatchers asking officers who can attend an incident rather than making an informed decision, which does not tend to lead to the most efficient officer being chosen. This research details an algorithm, where factors have been identified which should be used to base the decision on, to assist police officer dispatchers in the decision process by making the best use of the information available to recommend appropriate officers to send.

Research into police dispatch is limited. Green (1984) uses a multi-server queuing system with multi-priorities to model response car actions. The purpose of this technique is to provide more accurate queue time estimates by the efficient allocation of response cars. In the queuing process each server is considered identical and calls for service arrive according to a Poisson process at rate λ. Two priority classes are considered where the emergency class is served ahead of the priority class but within each class the situations are served on a first come first served basis. In this system there is no reallocation, once a call for service has been allocated by a server it cannot be reassigned. The study segregates precincts into separate geographical areas. This method of policing is not the most efficient and is being replaced by boundary-less policing.

Receiving more research attention is work on ambulance dispatch. Haghani et al (2003) produced an ambulance dispatch tool through developing an optimisation model. The tool included features such as a weighting system to prioritize more severe incidents and the use of real-time traffic information. It deals with these aspects by reassignment and rerouting when a new incident is reported with higher importance or when traffic conditions are significant. Henderson et al (2004) considered dispatch of ambulances and positioning for ambulance stations using techniques such as queuing theory and Markov chain analysis. Simulation was used to simulate incidents for the ambulances to react to, to determine whether ambulances could meet response times. Further research by Bandara et al (2012) determined that the closest ambulance is not always the quickest, though that is what decisions are typically based on. A dispatch tool was produced using a heuristic. The tool dispatched based on severity of incident measured by survival function. Simulation was used to prove that considering severity decreases mortality rate.

Part of the dispatch process for the police shares similarities with that of the ambulance service, however there are some attributes that are distinct. The methodology developed in this research considers these specific factors for police dispatch, such as driving qualifications and demand coverage, which are not currently identified in combination in other research. This research has developed the most comprehensive dispatch algorithm which considers availability, response times (incorporating driving standard, time of incident and traffic factors), severity of incident and demand coverage. Implementation of the decision model also includes two other main areas: generating a map of the area of concern; and a routing application. The map details the paths possible to travel on and constraints such as one way roads. The routing application determines the quickest path from an officer to an incident using the paths detailed on the map along with traffic information.

Kang et al (2005) investigated optimal route planning based on a real traffic network to avoid congested areas. The research produces an appropriate optimal route plan for a driver not using real-time traffic information by using historical traffic data to predict the speed of traffic flow along certain roads. The data used to predict the speed of traffic flow is road type, facilities type and link type, number of road, time and day. A survey of driver preference was also considered, the results of this survey concluded that the maximum driving speed and number of lanes were key factors in drivers choosing a route. The process of selection is as follows; use Modified Vine Building Algorithm to search for optimal route, calculate traffic costs, and finally perform an optimal route search using traffic info. The paper of Kang et al proves that using historical data is an effective means of predicting traffic effects on route times. The process of finding road weightings for the tool developed within this research varies due to the historical traffic information available for the area considered within the algorithm developed differing from that available in Korea where the study by Kang et al was based.
To test the effects of using this enhanced dispatch methodology discrete event simulation is used
Robinson (2004). This helps predict the variations in efficiency from using the tool in a police dispatch
environment compared to current methods of selection. The simulation runs through typical incidents and
finds the appropriate officer to send to each. The results of this are then compared to random selection of
officers to determine the difference in resource efficiency between each process.

2 FACTORS AFFECTING THE DISPATCH PROCESS

2.1 Defining the Decision Criteria Based on Incident Severity
Through investigation and dispatcher shadowing it was determined that there is considerable information
that dispatchers could use to aid them in assigning officers to reported incidents, with the overarching aim
to maximize incident attendance with response time targets. Evidence gained showed that the dispatcher
makes different decisions depending on the incident that requires attention, hence the required response
times (reflecting incident grading) impacts upon the decision making process. In emergency situations
(grade 1 incidents) a timely response is the main priority whereas in priority situations (grade 2) there is
more room to consider officers who are not immediately available given the longer response target time.
Hence a decision framework was developed for the two differing grade situations. Figure 1 shows the two
different selection criteria for response situations. The first path (Figure 1a) is the criteria for emergency
response and the second path (Figure 1b) is the criteria for priority response. The emergency response
decision considers availability and the predicted response times accounting for traffic and driving standard.
The priority situation considers availability, predicted response time accounting for traffic, and demand
coverage.

![Decision Criteria Diagram]

Figure 1 Decision criteria a) emergency b) priority

2.2 Considering Availability
The decision as to which officer to select for incident attendance can be restricted by the officer’s current
activity status. Their status describes what they are doing at that time and can include available, break,
dealing with incident, dealing with arrest and paper work. Available officers are immediately available to
attend incidents, making these officers eligible for emergency incidents. Other statuses mean the officer is
currently not available but it is possible for certain statuses, such as processing prisoners or on a break, to
allow officers to potentially be eligible for priority dispatch as time to availability can be predicted. The
statuses and the time they have been on this status can be used to determine which officers will be available.
Leigh, Jackson, and Dunnett

in time to meet the response time criteria. Utilizing the status culs the potential attendee set and speeds up the response allocation process.

2.3 Decision Making Considering Demand Coverage

Often it is the policy of forces to have police cover distributed across the policing region – this is referred to as demand coverage. A basic level of demand coverage aims to maintain at least one officer present within each town of the policing region. The officers do not have to be available, only present in the reachable area. To aim to maintain this level, demand coverage is a factor incorporated into the decision making process for selecting officers to respond to incidents.

When selecting an officer in a priority situation the coverage of the area which the officer would leave to attend the incident is considered. The decision process will choose officers from where there is a higher coverage to attend an incident rather than removing an officer from an area leaving it uncovered. In emergency situations response time is the main factor and hence demand coverage is not a decision factor.

2.4 Factoring in Predicted Response Time

The potential time taken for an officer to reach an incident is vital to the decision process. The proximity of the officer and the required destination, along with the route between the two is considered during the decision process. Factors affecting the route time include: road type, time of day, grade of incident, traffic congestion and officer driving qualifications. Weightings are used to reflect these different factors when calculating the route between two points.

Road type determines the speed limit which is a constant limiting factor for any particular road and is accounted for in calculating a route by recognising faster and slower roads. For example typically the motorway speed limit is 70mph hence a low road weighting is used and in residential areas the typical speed limit is 30mph which will be given a high road weighting. Traffic congestion is variable and changes during the day and for different days of the week, with peak times given higher weightings. Weightings are also given to officers travelling on foot or bicycle, based on the speed officers are expected to walk or cycle. Walking is only permitted on paths and cycling is not permitted on primary roads and motorways this is done by making certain roads unavailable.

To operate a police car the driver has to undertake various qualifications: basic, standard and advanced. The first level of training is the basic driving qualification which allows an officer to drive a car but does not allow activation of blue lights and sirens. The standard driving qualification allows activation of the blue lights and sirens in emergency situation. Blue lights and sirens can be used when necessary to warn cars of the officer’s presence, clear a path through traffic and allow the speed limit to be exceeded. The advanced driving qualification adds to the standard qualification the ability to pursue vehicles failing to stop. Basic drivers cannot utilise blue lights and sirens hence it will take them longer to travel the same path as a standard or advanced driver. These qualifications are considered in emergency situations as they affect the speed at which an officer can reach an incident. This is factored into the calculation by adjusting weights in the route calculation.

3 ALGORITHM IMPLEMENTATION

3.1 Overview

In order to enable a prompt response by the dispatcher, the process of officer selection needs to be automated. To facilitate the inclusion of the factors affecting dispatch into the decision making process (detailed in section 2), a mechanism of mapping and routing is required. The full algorithm to yield the most appropriate officer for incident attendance is shown in Figure 2. The map generated plots the roads and paths available to travel on within the police force boundaries, which enables the route generation to determine the quickest route from officer to incident, utilising the decision criteria.
3.2 Mapping

To facilitate calculation of a route between two points, or multiple points, the method of using a directed graph $G = (V, E)$ is used. This equation uses a group of vertices $V$ to represent the longitude and latitude points given as points along a road and uses edges $E$ to represent the roads which join the vertices. An edge is comprised of a pair of vertices $\{u, v\}$. Figure 3 shows how roads are represented in graph form, with an area within Leicestershire used for illustration purposes.

The graph is directed because roads may be one-way and hence can only be travelled in one direction. When this is the case $\{u, v\}$ is an ordered pair of vertices, Delling and Wagner (2009). The edges have weightings associated to them to account for the cost of travelling on the edge. To generate the directed graph information is extracted from OpenStreetMaps, http://www.openstreetmap.org/ accessed 1 September 2014. The information gained from OpenStreetMaps includes locations of points along roads, road type, road names and traffic restrictions such as one-way roads. Filtering of the data is required before use to remove unnecessary data such as building coordinates.

![Directed Graph](image)

**Figure 2** The dispatch algorithm

![Directed Graph](image)

**Figure 3** Directed graph

To navigate through the map’s road system a method of finding a path from a current location to a target location is required, this is found through routing.
3.3 Routing

For this Police application the shortest response time is the main objective so the path is found based on shortest time. To generate the shortest path to the incident Dijkstra’s algorithm is used, Joyner et al (2011). To determine this, a route can be calculated from the officer’s current location to the incident location, where the total distance (TD) is a sum of the distance of each road (D$_K$) on the path to the destination. Being close to an incident means that the route to an incident may be the shortest but it doesn’t necessarily mean that you can respond to an incident in the quickest time. There are a range of factors that must be taken into account to determine which officer will get to an incident the quickest. Given the calculation of the route distance to an incident the additional inclusion of other factors, via weightings, can be used to reflect the complexities of reaching the required destination. Hence the cost of the route is determined using equation 1.

$$\min \sum_{K} \sum_{i,j} W_K(t)D_{K_{ij}}$$  \hspace{1cm} (1)

Where $W_K(t)$ is the road weighting of road $K$, depending on the time $t$. $D$ is the distance travelled between nodes $i$ and $j$ on road $K$. In (1) the sum of the cost of travelling on an individual road is calculated and this cost is then added to the cost of the other roads used.

The road weightings are determined on factors that include driver skill set and geographical location. The adapted route distance calculation can then be used to determine the predicted response time. An example to show impedance to travel on roads is given in table 1, where an illustration of different weightings is shown for three factors: road type, incident type and time of day.

<table>
<thead>
<tr>
<th>Road type</th>
<th>Emergency weightings in peak traffic</th>
<th>Emergency weightings in low traffic</th>
<th>Non-emergency weightings in peak traffic</th>
<th>Non-emergency weightings in low traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorway</td>
<td>0.43</td>
<td>0.33</td>
<td>0.60</td>
<td>0.43</td>
</tr>
<tr>
<td>Trunk road</td>
<td>0.43</td>
<td>0.33</td>
<td>0.60</td>
<td>0.43</td>
</tr>
<tr>
<td>Primary</td>
<td>0.43</td>
<td>0.33</td>
<td>0.60</td>
<td>0.43</td>
</tr>
<tr>
<td>Secondary</td>
<td>0.5</td>
<td>0.38</td>
<td>0.75</td>
<td>0.5</td>
</tr>
<tr>
<td>Tertiary</td>
<td>0.75</td>
<td>0.5</td>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td>Unclassified</td>
<td>0.75</td>
<td>0.60</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Residential</td>
<td>0.75</td>
<td>0.60</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

To save computational time the route from every officer in Leicestershire to an incident is not calculated. Of those available the $n$ closest officers to the incident are found using the straight line distance from officer to incident. The value of $n$ is location dependent, within the city it is higher than in county areas. The routes are shown on the map. Figure 4 shows the route from officers to an incident in an emergency incident determined by using Dijkstra’s algorithm with road weightings. In this situation officer D should be selected above other officers as it has the quickest route to the incident with the use of blue lights and sirens. The route for officer B is not calculated as the officer’s status is attending incident and as it’s an emergency situation the officer cannot be selected to attend.
4 SIMULATION

To validate this dispatch decision making process, discrete event simulation has been used. The events are incidents which are generated by the simulation. Figure 5 shows the simulation process. The simulation starts by creating the map of the area of concern, in this case Leicestershire, and then setting a random testing period and generating a list of incidents which are to be used as events in the simulation. Historical data is used to generate distributions to give a realistic view of the demand officers will face at particular times of the day. The grade of the incident determines which decision process the tool will use. The decision tool selects an officer to send to the incident by reading from the resource list to gain information on officers and then updating the resource list once selected. The officers are then made unavailable for a predetermined length of time based on historical time taken dealing with that type of event. The simulation records the response times, distance travelled by each officer, and availability of officers. This information is used to determine the efficiency of the use of the officers. The simulation is repeated for 100 iterations to find the average savings over multiple periods of time. Currently each officer allocation takes on average 91 seconds when \( n = 4 \).

In order to determine the effectiveness of the developed tool the simulation is also run using random officer selection between the \( n \) closest resources to the incident. This resembles the culture of asking officers to attend rather than allocating officers. The results of the two simulations are then compared to show the difference using the selection tool can have on the efficient use of resources.

5 RESULTS

5.1 Using the decision algorithm

The tool picks the most effective route from each officer to incident and then selects the officer it predicts to be the most efficient. \( n \) has been chosen as 4 for this research. Figure 6 shows one emergency situation generated from the simulation, the officers in the proximity, and the most efficient routes from each officer deemed appropriate to the incident. Table 2 shows the cost of routing for each officer, demonstrating C has the lowest cost.

5.2 Comparison of results

The results of running the simulation within an area of Leicestershire city centre for various staffing levels are shown in Figure 7 and 8. The incidents used within the simulation are from historical data and are a mixture of emergency and priority situations. Figure 7 shows the percentage of incidents attended within
Leigh, Jackson, and Dunnett

the recommended response times depending on the number of officers in that region. The response times using the decision tool is shown by the solid line, and random selection close to the incident by the dotted line. When four officers are in the region it is predicted on average only 56% of crimes will be attended within the recommended response times, this is similar to the 55% shown by random selection. The small difference between the two is due to the limited number of officers available for selection. As the number of officers increases the number of officers available to attend the incident increases hence the difference between choosing an optimal officer and randomly assigning an officer increases. As the number of officers reaches 12 there is on average 100% of incidents being responded to within the recommended response times. When the number of officers reaches 14 both there are more officers available than needed and both the decision tool and random selection can ensure response times within the limits. Figure 8 shows the percentage decrease in routing costs resulting from using the tool developed rather than random selection depending on the number of officers available to respond. The difference in routing costs shows that the lower the number of officers considered the higher the route cost savings made by using the tool. This is due to the fact that when there are fewer officers, on average the officers are further away from the incident location, the difference in route times between the optimal officer and a randomly selected office is then more significant. Hence choosing the most efficient officer in this case gives higher route cost savings. However the response times determined show that it is not possible to use the lower number of officers due to the queue time being too high resulting in the total response times being greater than the 15 minute limit.

Figure 5 Simulation Process

Table 3 Routing Costs

<table>
<thead>
<tr>
<th>Response officer</th>
<th>Routing cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.27</td>
</tr>
</tbody>
</table>
All results obtained are dependent on information used in the simulation. They will vary with factors such as average time spent at incidents, area of concern and road weightings. The time taken to deal with incidents is predicted using historical data from Leicestershire police and hence may be different in other police forces. Different police forces assign response officers varying tasks to perform at the scene of an incident. Increasing time taken to deal with incidents decreases the availability of officers. Decreasing the availability of officers will increase response times. Decreasing time to deal with incidents has the opposite effect. The area of concern will change the level of crime and hence the number of officer required to meet response targets. The road weightings are subject to change depending on the area as traffic levels are based on data from Leicestershire. Changing the road weightings gives a slight variation in the predicted response times.

The results of this simulation show that the model can increase the probability that the most efficient officer is selected to attend the incidents. With the limited resources available within the police it is essential that the most efficient officer is assigned to the incident. Decreasing the response time for a fixed level of officers will offer other benefits such as increased patrol time to deter crime.

Future work will include development of road weightings to account for live traffic information. This will require live traffic data from services such as TrafficMaster. Another development required is a user friendly graphical interface to enable dispatchers to be able to make efficient use of the tool.

6 CONCLUSION

A need for a police officer dispatch tool is clear, any means of increasing police resource efficiency is necessary in the current climate. The tool outlined has proved its worthiness in selecting efficient officers through simulation giving it a place in the current police dispatch environment. It uses information not currently used in the officer selection process, such as driving qualification, and quickest route, which proves to result in a more efficient officer selection. When the future work highlighted above has been carried out the implementation of the tool into a police dispatch area will be possible.
ACKNOWLEDGMENT

The cooperation of the Leicestershire Police is gratefully acknowledged as without this support this project would not be possible. This work was supported by the Economic and Social Research Council [ES/K002392/1].
REFERENCES


AUTHOR BIOGRAPHIES

JOHANNA LEIGH received a Meng (Hons) Aeronautical Engineering from Loughborough University in 2013. She is at present studying for her PhD at Loughborough University. Her current research interests include modelling service support systems, simulation applications and reliability modelling with Petri-Nets.

LISA JACKSON is a senior lecturer at Loughborough University. She holds a BSc and MSc degree in Mathematics from Loughborough University. Her current research interests include applications of reliability methods for enhanced reliability performance, fault diagnostics and prognosis, crime reduction through demand modelling and positioning, optimization, vehicle safety and new hazard analysis technologies and risk and reliability generation methods.

SARAH DUNNETT is a senior lecturer at Loughborough University. She holds a BSc and MSc degree in Mathematics from London University, UK. She also received a PhD from Leeds University, UK. Her current research interests include risk and reliability generation methods, modelling service support system reliability, reliability modelling of phased missions, vehicle safety and new hazard analysis technologies, numerical modelling of problems in aerosol science and modelling of aerosol filtration.
ASIMULATION OF A POLICE PATROL SERVICE SYSTEM WITH MULTI-GRADE TIME-VARYING INCIDENT ARRIVALS

Miss. Hanjing Zhang
Loughborough University
Leicestershire, UK
H.Zhang@lboro.ac.uk

Dr. Lisa Jackson
Loughborough University
Leicestershire, UK
L.M.Jackson@lboro.ac.uk

Dr. Antuela Tako
Loughborough University
Leicestershire, UK
A.Takou@lboro.ac.uk

Prof. Jiyin Liu
Loughborough University
Leicestershire, UK
J.Y.Liu@lboro.ac.uk

ABSTRACT

Due to the squeeze on public expenditure, the funding cuts imposed on the police provide a great impetus to find an efficient incident response sequence with limited resources. This is especially the case for police response systems which exhibit the characteristics of time-varying volume of demand. In this paper, we investigate two types of priority queues in the patrol service system. Both the incident arrival rate and the scheduled staff level change with time. For such a system, there is no analytical model available to give close-form performance, so simulation is used for the study. Although dynamic priority queues which enable more flexibility in setting the sequence of service requests are widely applied in many service systems, such as the NHS service system, the simulation model results show that in police patrol service systems static priority queue performs better.

Keywords: Police patrol service system, Priority queue, Time dependent arrival rate

1 INTRODUCTION

Performance measures of police patrol service systems typically focus on response times, especially for incidents which need to be responded to immediately. Unlike manufacturing or transportation systems with well scheduled demand, police emergency service providers always face time-dependent service requests. Patrol officers must be ready to respond to an assigned incident of which the time and place cannot be known in advance. Management of staff allocation is usually based on a reasonable demand prediction using historical records so moderate understaffed periods may occur in practice.

In police patrol service systems, call handlers in front desks divide reported incidents into four grades based on their urgency. Only two grades of incidents, emergency incidents with threat to life and priority incidents with necessary officer attendance, require immediate response. The common target in the UK is to attend over 85% of emergency incidents within 15 minutes and reaching over 80% of priority incidents in 60 minutes (Leicestershire Police, 2013). With a lack of staff, police patrol service systems may fail to meet service targets, so there is a trade-off between satisfying emergency response targets and satisfying priority response targets. Similar to call centre systems, police patrol service systems can be modelled as a queueing system. A good performance in responding to emergency incidents, with an acceptable priority incident response speed, is preferable for selecting a queue rule since emergency incidents have more
serious consequences if they fail to be responded to in time. It cannot be assumed that such targets can be met when incidents are served on a first come first serve basis. Two types of queues can be used to model the police response system. Static priority queue, which serves priority incidents only when there is no emergency incidents presented, may reduce the speed in responding to priority incidents. Dynamic priority queues, including accumulating priority queues and threshold priority queues, that provide extra flexibility in service sequence have been already investigated in some service systems, such as the NHS. The work presented in this paper investigates the service quality in police patrol service systems with alternate moderate under-staffed and over-staffed periods via simulation. Extending previous work on priority queue systems with multi-grade arrivals, both arrival rate and staff level vary over time in this simulation model. According to the simulation results, static priority queues perform better than dynamic priority queues in police patrol service systems when service targets are hard to be achieved due to lack of staff.

The remainder of the paper is arranged as follows. The next section reviews the existing work on priority queues. Section 3 compares two different dynamic priority queues using simulation. First come first serve queues and static priority queues are discussed as special cases of dynamic priority queues. The last section gives conclusions and discusses the future research direction.

2 EXISTING MODELS OF PRIORITY QUEUES

Traditional static priority queues are analysed under the assumption that requests for service have fixed non-pre-emptive priorities. In such situations, no service resource will be allocated to requests for service from lower urgency grades while there are some high urgent grade requests present. This priority queueing sequence does make sense in some situations such as computer systems and telecommunication systems (Choi and Chang, 1999). However, in some situations like medical emergency service systems and police patrol service systems, requests for service from different urgency grades have different service targets and absolute priority for service does not exist (Henderson and Mason, 2004; Metropolitan Police, 2013). High urgency grade requests for service expect to receive a faster service than low urgency grade requests. Several dynamic priority queues have been proposed to provide more flexibility in response sequences. Among all the priority queues, accumulating priority queues and threshold priority queues attract the most interest.

Accumulating priority queue was used for a steady state queue by Kleinrock (1964). Upon arrival, queueing index of service start accumulating linearly over time. The higher the urgency grade, the greater the accumulation rate of queueing index function. Requests for service from low urgency grades, which experiences a long waiting time, will eventually be able to receive services even if some more urgent requests exist. Then Kleinrock and Finkelstein (1967) extended this method in which queueing index increases in proportion to some arbitrary power of waiting time. Another variant has been applied by Hay et al. (2006) in which initial queueing index also depends on urgency grades of requests for service. Recently Stanford et al. (2014) and Sharif et al. (2014) have derived waiting time distributions for linear accumulated priority queues for single server systems and multi-server systems in steady state.

Threshold priority queues have been investigated by setting thresholds to upgrade requests for service from low urgency grades so that they can be served when high urgency grade requests are still waiting in queues. In the model of Ridley et al. (2003), low urgency grade requests have dynamic priorities where they become high urgency if their waiting time exceeds a given value. Knessl et al. (2003) proposed a dynamic priority queue system in which once there are more than a certain number of requests from low urgency grades waiting for service, the head of low urgency grade request queues will be either upgraded or abandoned. Gurvich and Whitt (2010) proposed a queue length ratio threshold where the next service will be allocated to service requests from the head of the queue with certain service urgency whose queue length exceeds a specified proportion of the total queue length. Later Perry and Whitt (2013) introduced a six-dimensional fluid model to approximate queue performances of fixed staff levels with time varying arrivals.
Compared to traditional static priority queues, dynamic priority queues may perform better in aligning service accessibilities with service target of different urgent grades of requests for service when there are insufficient staff available (Sharif et al., 2014). In practice, police patrol service systems will experience some moderate under-staffed periods due to staff unavailability. This paper is primarily concerned with system performances of different priority queues in the case of slight overload. The police forces prefer to select a priority queueing system which has a high tolerance for short periods of under-staffing.

3 SIMULATION MODEL FOR MULTI-GRADE SYSTEM

A discrete event simulation model is developed to replicate police patrol service systems for a period of 24 hours by a non-preemptive queue with time dependent Poisson arrivals. Time-varying requests for service arrival rate of all incidents, \( \lambda(t) \), is estimated hourly by historical records from the local police force. Figure 1 plots the estimated time-varying arrival rate.

![Figure 1 Time-Varying Arrival Rate per Hour](image)

There are two urgent grades which have different service targets as defined in real police systems. In this simulation model, service durations are the time spend on dealing with incidents, which are restricted to a common exponential distribution. The elapsed times for patrol officers to travel to the scene of incidents are also modelled by an exponential distribution. According to real patrol activity records, the simulation parameters are listed in Table 1.

<table>
<thead>
<tr>
<th>Urgency Grade</th>
<th>Arrival Rate</th>
<th>Service Time</th>
<th>Travel Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Emergency</td>
<td>30% ( \lambda(t) ) / hour</td>
<td>60 minutes on average</td>
<td>5 minutes on average</td>
</tr>
<tr>
<td>2. Priority</td>
<td>70% ( \lambda(t) ) / hour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A working day for patrol activities start from 7am and finish at 7am. Although police patrol service systems run continuously from one day to the next, the demand for service around 7am in the morning is very low and the system around that time is almost empty. Therefore each day can be viewed as separated and each replication can simulate a 24-hour working day. The service quality is estimated hourly via simulation. The delay in response of the current hour is postponed to the next hour. Assume there are five predefined shifts for staff to take. Some overlaps of shifts exist to cover the rush hours of time-varying arrivals of requests for service. Thus, a working day is divided into several separated periods of different staff levels. With the help of integer programming, the staff level for each shift and each period in the simulation model is set as summarized in Table 2.
At time \( t \), patrol officers have been dealing with existing incidents for a period of \( \mu \), where incidents arrive in the system earlier than \( t-\mu \). Thus the average number of busy patrol officers \( W(t) \), as defined in Jennings et al. (1996), is:

\[
W(t) = \int_{0}^{\infty} \{ \lambda(t - \mu) \cdot \text{Probability(service_time > \mu)} \} d\mu
\]  

Since the average number of busy staff does not take the service targets into consideration, it may slightly under estimate the required staff level but still is a reasonable guideline. Figure 2 compares the input staff level in this simulation model (Table 2) in the straight lines, with the average number of busy patrol officers calculated via Equation (1) in the curve line. It is obvious that the simulation system will experience moderate under-staffing around 10:00, 14:00, 18:00, 20:00 and midnight. The following simulation analysis will mainly focus on the system performance of these understaffed periods.

<table>
<thead>
<tr>
<th>Shift</th>
<th>Period</th>
<th>1 from 07:00 to 11:00</th>
<th>2 from 11:00 to 15:00</th>
<th>3 from 15:00 to 19:00</th>
<th>4 from 19:00 to 23:00</th>
<th>5 from 23:00 to 03:00</th>
<th>6 from 03:00 to 07:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (07:00-15:00)</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (11:00-19:00)</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C (15:00-23:00)</td>
<td>20</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D (19:00-03:00)</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E (23:00-07:00)</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff Level</td>
<td>12</td>
<td>20</td>
<td>28</td>
<td>32</td>
<td>20</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2 Input Staff Levels for Each Replication**

The queueing algorithms to be considered are essentially the accumulating priority queues discussed in Stanford et al. (2014) and the threshold priority queues discussed in Gurvich and Whitt (2010). First come first serve queues and static priority queues are also discussed as special cases in both the queue systems. The successful response probabilities are estimated by performing multiple (1,000) independent replications of simulation.

### 3.1 Accumulating Priority Queues

According to the work of Stanford et al. (2014), an emergency incident which arrives at time \( t_0 \) has a queueing index of \( Q_1(t_0, t) \). Similarly, the queueing index for priority incidents are defined as \( Q_2(t_0, t) \), where:

\[
Q_i(t_0, t) = b_i \cdot (t - t_0), \quad i \in \{1 \text{(emergency incidents)}, 2 \text{(priority incidents)}\}
\]

At the time when a patrol officer finishes the current work and becomes available, the incident which has the highest queueing index will then be attended by this officer. Figure 3 illustrates the operation of an accumulating priority queue.
Figure 3 Example of an Accumulating Priority Queue

The upwards arrows in the lower portion of Figure 3 represent the arrival times of four incidents in this example, where incident 1 and incident 4 are emergency incidents; incident 2 and incident 3 are priority incidents. The accumulation rate for emergency incidents, $b_1$, is larger than the rate for priority incidents $b_2$. The downwards arrows in the lower portion of Figure 3 indicate the time points when there is a patrol officer available to attend an incident waiting for service. Whenever there is an available patrol officer, this patrol officer will attend the incident which accumulates the highest queueing index at this moment. For example, at the time that officer C is available, incident 3 and incident 4 are still waiting to be responded to. Although incident 4 arrived later than incident 3, officer C is assigned to attend incident 4 because incident 4 has a higher queueing index at this moment. Similar to the dispatch of officer C, officer A, B and D are assigned to attend incident 1, 2 and 3, respectively.

Figure 4 Proportion of Emergency Incidents (a) and Priority Incidents (b) Responded within Targets Using the Accumulating Priority Queue When $b_2/b_1 = 0.1, 0.5$ and $1.0$

The accumulation rates, $b_1$ and $b_2$, do not directly influence the service sequence, but their ratio. Assume the accumulation rate for emergency incidents $b_1$ is 1, the priority incident accumulation rate $b_2$ should take values between 0 to 1 since more service resources are dispatched to deal with emergency incidents in practice. This accumulating priority algorithm is applied to simulate police patrol service systems with time varying arrivals. Figure 4 summaries the simulated police patrol service systems when taking different values of $b_2/b_1$. 

107
Due to the staff unavailability at 10:00, 14:00 and 20:00, the simulation fails to meet the set targets (more than 85% emergency incidents are responded to within 15 minutes and 80% priority incidents are responded to within 60 minutes) no matter how the value of $b_2/b_1$ changes as shown in Figure 4. When $b_2/b_1 = 1.00$, as shown in the dot-dashed line, the accumulation rate for priority incidents is the same as the rate for emergency incidents, so this priority queue degenerates into a first come first serve queue. As the value of $b_2/b_1$ decreases, emergency incidents have more advantage in being responded to than priority incidents. Comparing $b_2/b_1 = 0.50$ in the dashed line to $b_2/b_1 = 0.10$ in the dotted line, a conclusion could be drawn that the value of $b_2/b_1$ does not have a big impact on the service for priority incidents, but the service for emergency incidents is largely improved when $b_2/b_1$ decreases. A paired-t test is performed to compare service quality change between emergency incidents and priority incidents. The data for the paired-t test are summarized in Table 3. With a t-statistic of 8.08 and the degree of freedom of 23, the p-value of this paired-t test is smaller than 0.01% which confirms the above conclusion.

**Table 3** Paired-t Test Data for Service Quality Change from $b_2/b_1 = 0.50$ to $b_2/b_1 = 0.10$

<table>
<thead>
<tr>
<th>Time</th>
<th>Respond to Emergency Incidents within 15 minutes</th>
<th>Respond to Priority Incidents within 60 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>from $b_2 = 0.50$ to $b_2 = 0.10$</td>
<td>Percentage Change</td>
</tr>
<tr>
<td>07:00 - 08:00</td>
<td>97.81%</td>
<td>98.30%</td>
</tr>
<tr>
<td>08:00 - 09:00</td>
<td>97.65%</td>
<td>98.48%</td>
</tr>
<tr>
<td>06:00 - 07:00</td>
<td>93.95%</td>
<td>94.50%</td>
</tr>
</tbody>
</table>

### 3.2 Threshold Priority Queues

In queue length ratio based threshold priority queues, choices of service sequence are based on incident queue length ratio of different urgency grades. Incidents of certain urgency grade which exceeds its threshold queue length ratio the most will get the next service resource. Define the priority incident queue length ratio, $r$, as the ratio of number of waiting priority incidents to amount of total incidents waiting for service.

$$r = \frac{\text{amount of priority incidents waiting for service}}{\text{amount of total incidents waiting for service}} \in (0,1)$$  \hspace{1cm} (3)

In the simulation model of patrol police service system, there are only two urgency grades of incidents. The emergency incident queue length ratio can be denoted by $1-r$. Once the queue length ratio of priority incidents $r$ exceeds the threshold value $r_0$, the next available officer will attend the incident at the head of the priority queue; otherwise, this officer will be allocated to the incident at the head of the emergency queue. Figure 5 is a graphical illustration of this queue length ratio based threshold priority queue.
Since 70% of incidents require priority response and emergency incidents should at least have the same priority to access to services, only the queue length threshold ratio $r_0$ of the value no less than 0.70 will be discussed. Figure 6 presents the simulation results. The percentages of successful emergency responses within 15 minutes are shown in Figure 6a and the percentages of successful priority response within 60 minutes are shown in Figure 6b, where $r_0$ is set to 1.00, 0.90, 0.80 and 0.70 using a solid line, dotted line, dashed line and dot-dashed line, respectively.

There is no significant difference in responding to priority incidents when taking different values of the threshold queue ratio $r_0$. The best performance is obtained when the ratio $r_0$ is 1.00, which corresponds to static priority queue. Comparing these four lines, the simulation results indicate that with an increase in the value of the queue length threshold ratio $r_0$, the effect of the under-staffing in responding to emergency incidents decreases. The service target for emergency incidents is achieved when $r_0 = 1.00$ and the percentage of successful responded to emergency incidents is around 85% when $r_0 = 0.90$ regardless of the occurrence of under-staffing periods at 14:00, 18:00 and 21:00. Paired-t tests with all the p-values less than 0.01% confirm that emergency response service quality is more sensitive to the change of $r_0$. Table 5 summaries percentage changes of service quality as the data for the paired-t tests.
The best service for emergency incidents is obtained when performing static priority queue which is a special case of threshold priority queue with the queue length ratio threshold \( r_0 = 1.00 \). The accumulating priority queues with the accumulation rate ratio \( b_2/b_1 = 0.10 \) approximates the behavior of static priority queues by assigning emergency incidents a large accumulation rate. A reasonable assertion can be made that with a large enough accumulation rate ratio \( b_2/b_1 \), the accumulating priority queue will eventually become a static priority queue. The best service for priority incidents is obtained when performing a accumulating priority queue with the accumulation rate ratio \( b_2/b_1 = 1.00 \). It is exactly a first come first serve queue regardless of urgency grades, the simulated service system meets the priority incident service target with scarifying the service quality of the emergency incidents.

Compared with emergency incidents, police service quality of priority responses is not sensitive to the parameter values, static priority queues which always give priority to respond to emergency incidents is preferable for police patrol operations. No matter how the values of parameters change, the service targets for incidents in both the urgent grades cannot be met at the same time because of a lack of staff. More staff should be added into the system for a better service.

### 4 CONCLUSION AND FUTURE WORK

In this paper, we compare two types of priority queues, the accumulating priority queue and the threshold priority queue, for real time deployment of patrol officers in police patrol service system. Although dynamic priority queues seem to perform better in aligning service resources to service targets for a multi-priority system, traditional static priority queue still provides an efficient way to set the service sequence when the service target from the higher urgency grade is hard to achieve in moderate under-staffed periods.

All results are subject to change depending on the simulation input parameters, such as the patrol officer service rates, the travel time and the number of staff on each shift. An obvious extension of current simulation models is that the required service resource may vary with the nature of incidents. For example, a road related incident may require two patrol officers to attend at the same time but to attend a serious street fighting incident may require more patrol officers to work together. Due to differences in patrol policies, in some police area the graphical factors may also influence the service targets. According to Nottinghamshire Police HQ(2015), the service target for emergency incidents is 20 minutes in rural area and 15 minutes in urban area. Since most police cars are equipped with GPS in the UK, with the GPS
information of available patrol officers patrol activity dispatchers in the police force is able to contact the most efficient police officer to attend incidents and an improvement in service could be expected. Combined with integer programming models, the simulation will be able to indicate more effective staffing levels.

REFERENCES


AUTHOR BIOGRAPHIES

HANJING ZHANG is a PhD student in Operational Research at Loughborough University. She is working on an inter-departmental project which focusing on developing combined simulation and optimisation models that can be used to support optimal staff allocation in the local police force. Home page: http://www.lboro.ac.uk/departments/sbe/staff/profiles/phdprofiles/zhanghanjing/zhang-hanjing.html

LISA JACKSON is a Senior Lecturer in Risk and Reliability at Aeronautical and Automotive Engineering Department, Loughborough University. She is also a principal investigator for the ESRC sponsored project
on Demand Modelling and Predictive Positioning of Police Patrol Officers: a project investigating effective use of police patrol officers to maximise availability for crime response. Home page: http://www.lboro.ac.uk/departments/aae/about/staff/dr-lisa-jackson.html

ANTUELA A. TAKO is a Lecturer 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 and conceptual modelling. She is an Associate Fellow of the Operational Research Society. Home page: www.lboro.ac.uk/departments/sbe/staff/profiles/takoantuela/tako-antuelaanthi.html.

JIYIN LIU is a Professor of Operations Management at Loughborough University School of Business and Economics. He received his PhD in manufacturing engineering and operations management from University of Nottingham. His research interests include modelling and optimisation of operations planning and scheduling problems in industrial and service systems. He has published widely. His research papers appeared in academic journals such as European Journal of Operational Research, IEEE Transactions, IIE Transactions, International Journal of Production Research, Journal of the Operational Research Society, Naval Research Logistics, Operations Research and Transportation Research. Home page: http://www.lboro.ac.uk/departments/sbe/staff/profiles/liujiyin/liu-jiyin.html
A LINGUISTIC EXAMINATION OF PROBLEM SOLVING APPROACHES IN SIMULATION

Dr. Roger McHaney
Kansas State University
Manhattan, KS 66506, USA
mchaney@ksu.edu

Dr. Antuela A. Tako
Loughborough University
Loughborough, LE11 3TU, UK
A.Takou@lboro.ac.uk

Dr. Stewart Robinson
Loughborough University
Loughborough, LE11 3TU, UK
S.L.Robinson@lboro.ac.uk

ABSTRACT

Decision makers come from a variety of backgrounds and structure problems according to paradigms reflecting their world views. This phenomenon may result in inappropriate decision support tool usage and/or limit problem-solving approaches. The current research explores how Linguistic Inquiry and Word Count (LIWC) software can be used to better ensure problem/model fit with a focus on the selection of system dynamics (SD) and discrete event simulation (DES). To accomplish this objective, problem statements were analysed using LIWC to uncover linguistic clues that might provide guidance for selection of an appropriate solution methodology. Results indicated problem statements contained linguistic clues consistent with hypothesized characteristics of SD and DES, and these characteristics could be encoded in a LIWC custom dictionary for use by analysts. Subsequent discussion describes how LIWC can be used to help select appropriate decision support tools and problem-solving approaches in applied settings. Future research plans are also provided.

Keywords: Decision Support, System Dynamics, Discrete Event Simulation, Simulation, Model Building, Linguistic Inquiry and Word Count, Simulation Success

1 INTRODUCTION

Linguistic Inquiry and Word Count (LIWC) is a computer software program developed by psychology researchers at the University of Texas. It examines speakers’ and/or writers’ emotional and cognitive processes through their language usage choices (J.W. Pennebaker & King, 1999). The software relies on approximately 80 validated dictionaries to categorize and classify nearly 4,000 words and word stems into broad language choice dimensions (Cindy Chung & Pennebaker, 2007). The LIWC program operates as a data processing system. It opens a series of text files which may contain the contents of essays, paragraphs, advertising text, prose, poems, blogs, or other writings. It then compares each word from the content to the dictionaries. The words are classified into validated dimensions using this approach. According to Tausczik and Pennebaker (2010), more than 100 million words derived from text files in various studies were used to develop the software and ensure the tool’s validity. Further details regarding the creation of LIWC and its validation can be found in Pennebaker et al. (2007).

Using a tool such as LIWC helps overcome inherent limitations of social psychology research which necessitate reliance on observable activities such as behaviours or on the collection of self-report data such as surveys. While traditional approaches are good for validating theories or studying general patterns, many real-world applications such as making marketing choices, developing political campaigns, or selecting...
McHaney, Tako, and Robinson

decision support methodologies require understanding early signals that may influence predictive or pre-emptive choices. Our research investigates correlations of individuals’ word usage with desires to obtain a particular form of information that can influence better decision-making choices. For instance, some problems are more operational in nature while others may be broad and strategic.

In previous usage, Hochberg et al. (2014) conducted research regarding medical field decision making where outcomes may have life and death consequences. Our premise is to expand their suggestion that linguistic markers can be used to determine optimal modes of decision making in other domains. The current study specifically expands their ideas into the business and organizational decision support domain. We contend that observation of underlying word choice selection of decision makers may reveal subtle dimensions that exist in their decision making process. Although a variety of tools can be used to solve most problems, even a slight alignment advantage between problem and solution can have beneficial outcomes. This preliminary study describes how an advantage might be gained by better understanding, as Chung and Pennebaker (2012) put it, the “…Little Words in Big Data.” Like Hochberg et al. (2014), we hope to demonstrate how linguistic data clues are present in problem statements. Illuminating these clues, we contend, will provide additional data for movement toward better modes of reasoning, with particular value to high-stakes decision making in business. Using the best possible tools can only help provide better decision making capabilities.

1.1 Domain for Study
System Dynamics (SD) and Discrete-Event Simulation (DES) are model development approaches used to better understand, examine, fine-tune, and develop models for real world systems spanning a wide variety of areas and applications. Operations management, operations research, business, and industrial engineering and other fields have adopted these approaches as a basic component of their tool kits. SD and DES modelling both provide valuable insight into the operations of real world systems and are employed when experimentation with an actual system is too disruptive, expensive or even impossible to conduct. SD and DES evolved as two separate problem solving approaches with little overlap between users of each technique. Tako and Robinson (2010) recently examined these two approaches and described how each approach focuses on particular problem-attributes. This naturally permits solutions to different types of problems more intuitive in respective areas. For example, analysts using SD more often seek to develop a big-picture model with more concentration on strategic and overarching attributes. In contrast, analysts using DES often follow a linear progression with greater focus on model coding, verification, and validation with emphasis on operational details. Other studies support this view based on anecdotal accounts and industry experience (Brailsford & Hilton, 2001; Morecroft & Robinson, 2005).

1.2 Study Motivations
Having two similar tools with subtle differences provides a good venue for examining whether linguistic differences can be used as a criteria for selecting different problem solving approaches. Due to the orientation and world view of each approach, it follows that particular methods are better suited to solving particular problems. While this seems a sensible way to approach problem solving, reality may be different. For instance, analysts may solve problems using tools with which they have experience or a comfort level, even if that tool is not the best choice. Likewise, problem solvers may not recognize the nuances of the decision maker or problem owner. In other words, a problem description statement might contain subtle clues about the desired outcome that may not be intuitively visible. A tool such as LIWC is well suited to bring these details to light. To illustrate further, this study examines ten problem statements for linguistic clues indicating which approach might be best suited for problem solution development. These problem statements were derived from existing literature and represent experts’ approaches to solving problems. This information will be used to validate the conclusions suggested by an LIWC analysis developed using first default dictionaries and then a custom dictionary specifically developed for SD/DES.
2 BRIEF SIMULATION LITERATURE OVERVIEW

This study is based on earlier theoretical work particularly that of Tako (2015) and uses the SD and DES literature as a basis for creating a custom LIWC dictionary for discriminating between various problem approaches. We believe language dimensions can be relevant in the process of selecting an appropriate modelling approach/tool which is an important step in the simulation lifecycle (Balci, 2012; McHaney, 1991, 2009; Pidd, 2004; Robinson, 2014). To better understand the examples illustrated in this study, a brief overview regarding SD and DES simulation approaches follows.

2.2 SD Modelling Overview

System Dynamics models generally are consistent with systems theory and have been used to understand complex systems behaviour over time. SD is a mathematically based modelling technique that can help define, describe, analyse, and fine-tune dynamic systems through the use of internal feedback loops and time delays ($\Delta t$). Small changes result in larger dynamics that affect system behaviours over time (Forrester, 1961). SD is also considered a methodology for representing complexity that involves feedback loops, stocks, and flows arranged to help explain nonlinearity and other emergent actions. SD models represent causation and simultaneity through interrelated variables updated in small time increments. SD illuminates the impact of time delays and complex relationships which produce system behaviours that are not easily determined through analysis of system components individually. Therefore, SD is well-suited to holistic understanding and strategic decision analysis. According to Sterman (2000), an analyst using SD should take an endogenous viewpoint and consider the system as a closed-loop structure that defines its own internal behaviour. Feedback loops are at the heart of the system and yield interesting outcomes. Individual events and decision within the system are “surface phenomena that ride on an underlying tide of system structure and behaviour. It is that underlying tide of policy structure and continuous behaviour that is system dynamicist’s focus” (Richardson, 2011, p. 861). For the purposes of the current study, SD offers solution potential for strategic, holistic and complex problems.

2.3 DES Modelling Overview

Discrete Event Simulation is often used in the development of representational decision support system models used by engineers, business analysts, and design specialists. DES is particularly appropriate in detail-oriented models used at operational or tactical levels within organizations. In general, DES is distinguished by the passage of blocks of time during which no changes to the system state occur. Events are scheduled and placed on future and current event chains that cause the system state to update. Generally, entity arrivals and service completions impact a network of queues and resources to allow particular system characteristics and operating phenomenon to emerge. Interaction between constrained resources dynamically drives a model. In DES, system events are instantaneously occurring in discrete steps with resulting impact altering systems operation (Robinson, 2014). To the analyst, a DES approach to modelling considers the system being modelled as a process (Borschchev, 2013, p. 45). This process can be modified, experimented upon, fine-tuned, stress-tested, and subjected to a variety of experimental conditions. For purposes of the current study, DES offers solution potential for detail-oriented, operational problems.

2.4 Comparison of SD and DES

A number of studies have compared SD and DES in a variety of ways (Morecroft & Robinson, 2005). Many comparisons take the form of authors’ helpful views of the approaches with a bias toward foundational concepts and authors’ expertise in each area (Brailsford & Hilton, 2001). Dialog between the two communities of modellers was infrequent and this appears to have resulted in the independent growth of various approaches and techniques (Lane, 2000; Sweetser, 1999). Morecroft and Robinson (2005) explored these differences and suggested reasons why the two modelling approaches are different, but with the caveat that application of both can yield complementary insights. In many instances, it appears that choice of a modelling tool may be based on prior experiences, and existing knowledge rather than the most appropriate tool for a particular desired outcome.
Tako and Robinson (2009; 2010) approached the differences from first, a user perception perspective, and later, from an empirical perspective to comparatively explore model development processes. Tako (2015) followed up with an in-depth empirical study of discrete-event simulation (DES) model development reporting empirical results which relate the thought processes of six expert modellers while building DES models. In these empirical studies, Verbal Protocol Analysis (VPA) was used to collect data. The experts were instructed to speak aloud as they conceptualized and developed models. Results illustrated modelling processes unique to SD and DES, and provided insight into the differences between modelling with each perspective. Patterns of modelling behaviour were reflected in respective experts’ articulation of their thought processes.

Based on the existing comparison literature, Table 1 provides a detailed view of the differences together with a categorization scheme that shows how these differences may impact the way analysts view SD and DES (McHaney, Tako, & Robinson, 2014). This information was used to develop a custom dictionary for LIWC for use in discriminating between the appropriateness of DES and SD approaches.

### 3 A LINGUISTIC EXAMINATION OF PROBLEM STATEMENTS IN SIMULATION

As previously mentioned, the impetus for this study is to determine if LIWC or similar tools can be used to help determine the correct approach for developing a solution to a complex problem. For instance, there may be a tendency for analysts to use solution approaches that are familiar or in which they have experience or domain knowledge, rather than select the appropriate method (Lane 2000). A larger question might be: Which approach is best for modelling a problem? For this reason, we believe it is important to examine whether linguistic clues exist in problem statements which can help guide the solution approach taken. Several of the motivating factors for this investigation include: problems may be approached using the wrong analysis technique; techniques might be used because these match the experience of the analyst; an organization may have invested in a particular tool and therefore might tend to shoehorn a solution approach on to a problem.

To investigate this further, we examined simulation study problem statements for patterns that might help determine an appropriate solution approach. In other words, are linguistic clues embedded in the language of those describing problems? If so, vital clues as to the best approach to use in applying an appropriate solution technique also might exist. Linguistic analysis might provide insight into the deeper motivations of those needing information. To do this, we collected problem statements from ten simulation studies. These were edited to place them into a similar formats but wherever possible, the original language of those describing the scenarios to be modelled was retained. The collection of problem statements ranged from 323 to 570 words in length with average length being approximately 467 for published SD solutions and 465 for published DES solutions. Table 2 contains an overview of this information.

<table>
<thead>
<tr>
<th>Problem Structure/Development Philosophy</th>
<th>DES</th>
<th>SD</th>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational</td>
<td>Strategic</td>
<td>Operational</td>
<td>Strategic</td>
<td>Best Problems for Technique</td>
</tr>
<tr>
<td>Discrete</td>
<td>Continuous</td>
<td>Discrete</td>
<td>Continuous</td>
<td>State Changes</td>
</tr>
<tr>
<td>Analytic; emphasis on detail complexity</td>
<td>Holistic; emphasis on dynamic complexity</td>
<td>Analytic; emphasis on detail complexity</td>
<td>Holistic; emphasis on dynamic complexity</td>
<td>Perspective</td>
</tr>
<tr>
<td>Detail Level</td>
<td>Aggregate Level; Big picture</td>
<td>Detail Level</td>
<td>Aggregate Level; Big picture</td>
<td>Model Scope</td>
</tr>
<tr>
<td>No agreed standard</td>
<td>Standard Structures</td>
<td>No agreed standard</td>
<td>Standard Structures</td>
<td>Diagramming Format</td>
</tr>
</tbody>
</table>
### Inputs

<table>
<thead>
<tr>
<th>DES</th>
<th>SD</th>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity durations from probability distributions</td>
<td>Time in reservoir modelled with limited flexibility</td>
<td>System Activity</td>
<td>(Brailsford &amp; Hilton, 2001)</td>
</tr>
<tr>
<td>Individual entities, attributes, decision and events</td>
<td>Homogenised entities, continuous policy pressures, and emergent behaviour</td>
<td>Resolution</td>
<td>(Lane, 2000)</td>
</tr>
<tr>
<td>Features randomness prominently.</td>
<td>Stochastic variability subsumed into an appropriate delay; model structure driven</td>
<td>Randomness</td>
<td>(Coyle, 1985; J. D. W. Morecroft &amp; Robinson, 2005)</td>
</tr>
<tr>
<td>Primarily numerical with some judgmental elements</td>
<td>Broadly drawn</td>
<td>Data Sources</td>
<td>(Lane, 2000)</td>
</tr>
</tbody>
</table>

### Internal Processes

<table>
<thead>
<tr>
<th>DES</th>
<th>SD</th>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many linear</td>
<td>Many non-linear</td>
<td>Relationships</td>
<td>(Morecroft &amp; Robinson, 2005)</td>
</tr>
<tr>
<td>Growth/decay represented as random with discrete steps</td>
<td>Growth/decay modelled as exponential or s-shaped</td>
<td>Relationships</td>
<td>(Morecroft &amp; Robinson, 2005)</td>
</tr>
<tr>
<td>Standard structures rarely exist</td>
<td>Recurring structures</td>
<td>Modelling Structures</td>
<td>(Morecroft &amp; Robinson, 2005)</td>
</tr>
<tr>
<td>Open-process structure</td>
<td>Closed-loop structure with feedback</td>
<td>Problem Structure</td>
<td>(Coyle, 1985)</td>
</tr>
<tr>
<td>Networks of queues and activities</td>
<td>Series of stocks and flows</td>
<td>System Representation</td>
<td>(Brailsford &amp; Hilton, 2001)</td>
</tr>
<tr>
<td>Objects are distinct individuals with characteristics</td>
<td>Objects are continuous quantity</td>
<td>Object Representation</td>
<td>(Brailsford &amp; Hilton, 2001)</td>
</tr>
<tr>
<td>Unequal timesteps; when “something happens”</td>
<td>Finely-sliced time steps of equal duration</td>
<td>Time</td>
<td>(Brailsford &amp; Hilton, 2001)</td>
</tr>
</tbody>
</table>

### Feedback

<table>
<thead>
<tr>
<th>DES</th>
<th>SD</th>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback held in model logic</td>
<td>Explicit feedback</td>
<td>Placement</td>
<td>(Máik, 1993)</td>
</tr>
<tr>
<td>Implicit</td>
<td>Explicit</td>
<td>Transparency</td>
<td>(Morecroft &amp; Robinson, 2005)</td>
</tr>
<tr>
<td>Not emphasised</td>
<td>Vital</td>
<td>Importance</td>
<td>(Morecroft &amp; Robinson, 2005)</td>
</tr>
</tbody>
</table>

### Output

<table>
<thead>
<tr>
<th>DES</th>
<th>SD</th>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opaque/dark grey box</td>
<td>Transparent/fuzzy glass box</td>
<td>Client View</td>
<td>(Lane, 2000)</td>
</tr>
<tr>
<td>Statistically valid estimates of system performance measures</td>
<td>Results considered source of understanding reasons for changes in system performance</td>
<td>Outputs</td>
<td>(Brailsford &amp; Hilton, 2001; McHaney, 1991; Morecroft &amp; Robinson, 2005; Sweetser, 1999)</td>
</tr>
</tbody>
</table>
### Table 2 Cases Used in Current Study

<table>
<thead>
<tr>
<th>ID #</th>
<th>Problem Domain</th>
<th>Solution Technique Used</th>
<th>Length in Words</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UK Prison System</td>
<td>SD/DES</td>
<td>462</td>
<td>(Tako &amp; Robinson, 2010)</td>
</tr>
<tr>
<td>3</td>
<td>Family Practice System</td>
<td>DES</td>
<td>509</td>
<td>(Jacobson, Hall, &amp; Swisher, 2006)</td>
</tr>
<tr>
<td>4</td>
<td>Snack Foods Warehouse</td>
<td>DES</td>
<td>421</td>
<td>(“Walkers Simulation Case Study,” 2015)</td>
</tr>
<tr>
<td>5</td>
<td>Water System Model</td>
<td>SD</td>
<td>527</td>
<td>(Stave, 2003)</td>
</tr>
<tr>
<td>6</td>
<td>Urban Growth</td>
<td>SD</td>
<td>419</td>
<td>(Ghaffarzadegan, Lynees, &amp; Richardson, 2011)</td>
</tr>
<tr>
<td>7</td>
<td>Swamping</td>
<td>SD</td>
<td>477</td>
<td>(Ghaffarzadegan et al., 2011)</td>
</tr>
<tr>
<td>8</td>
<td>Lean Manufacturing</td>
<td>DES</td>
<td>563</td>
<td>(Detty &amp; Yingling, 2010)</td>
</tr>
<tr>
<td>9</td>
<td>Words of Mouth</td>
<td>SD</td>
<td>441</td>
<td>(Wong &amp; Sheng, 2012)</td>
</tr>
<tr>
<td>10</td>
<td>AGV System</td>
<td>DES</td>
<td>478</td>
<td>(McHaney, 1995)</td>
</tr>
</tbody>
</table>

### 3.1 Preliminary Study with Default LIWC Word Dictionary

A preliminary examination of each problem statement was conducted using the default dictionary available with LIWC to determine whether any significant differences between the problem statements existed. We believed this dictionary had the potential to offer useful information with low overhead costs since the automated text analysis operation is fast and easy (Quinn et al. 2010). A variety of dimensions were tested based on categories found to be relevant to SD or DES modelling in an earlier study (McHaney et al., 2014).

In that 2014 study, the language used by DES and SD simulation experts during descriptions of model development was subjected to a linguistic analysis. The underlying premise of that study was to determine if differences in cognitive processing manifested when experts built models in their respective domain. Specifically, were simulation experts more likely to use language that correlated with unique aspects of their modelling approach? The study hypothesized that analysts using a SD approach were more likely to use holistic, policy-level and strategic word choices. On the other hand, analysts using a DES approach were more likely to focus on details, analysis, and individualistic words. Related hypotheses were developed based on theoretical information compiled in Table 1 and then were tested on that data (McHaney et al., 2014).

The hypotheses developed in the 2014 study were retested in the current study using the problem statement data set. As shown in Table 3, the default word dictionary for LIWC yielded only one significant
result for this analysis. Only H4: Affective Processes was significant in a direction consistent with hypothesized outcomes. Affective processes are those dealing with individual actions which would be more consistent with DES modelling of actions at the individual level. Overall, the lack of substantial results in this part of the analysis were not a surprise. We should not expect a difference in the problem statements issued by problem owners because the individuals posing these questions are not generally aware of techniques that will be used to create a solution. Table 3 contains the results of this portion of the analysis.

3.2 Development of a Custom LIWC Word Dictionary

Following the preliminary analysis of the default LIWC dictionary, the focus of the study turned to an examination of the linguistic context of the problem statements through development of a custom dictionary intended to capture nuances of the SD and DES simulation domains. In the past, a number of approaches have been used in similar efforts. Rice & Zorn (2013) describe these approaches as falling into two broad classes: (1) Machine learning that relies on “classifying or scoring a subset of texts (usually documents) on their sentiment, and then using their linguistic content to train a classifier” (p. 2); and (2) “dictionary-based approaches [that] begin with a predefined dictionary of positive and negative words, and then use word counts or other measures of word incidence and frequency to score all the opinions in the data” (p. 2). With the second approach, particularly, validity depends heavily upon quality, comprehensiveness and theoretical foundations of the words used in the construction process.

Table 3 Default LIWC Word Dictionary Categories Theoretically Relevant to SD/DES Comparison for Problem Statements (See McHaney et al., 2014 for details regarding LIWC Dimensions)

<table>
<thead>
<tr>
<th>LIWC Dimension</th>
<th>Hypothesized Relationship</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Words &gt; 6 Letters</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.663)</td>
</tr>
<tr>
<td>H2: Personal Pronouns</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.005)</td>
</tr>
<tr>
<td>H3: Second Person</td>
<td>SD &gt; DES</td>
<td>Not Supported</td>
</tr>
<tr>
<td>H4: Affective Processes</td>
<td>DES &gt; SD</td>
<td>Supported (p=.041)</td>
</tr>
<tr>
<td>H5: Insight</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.544)</td>
</tr>
<tr>
<td>H6: Causation</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.102)</td>
</tr>
<tr>
<td>H7: Tentative</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.624)</td>
</tr>
<tr>
<td>H8: Discrepancy</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.191)</td>
</tr>
<tr>
<td>H9: Certainty</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.089)</td>
</tr>
<tr>
<td>H10: Inhibition</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.689)</td>
</tr>
<tr>
<td>H11: Inclusive</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.218)</td>
</tr>
<tr>
<td>H12: Exclusive</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.668)</td>
</tr>
<tr>
<td>H13: Perception</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.860)</td>
</tr>
<tr>
<td>H14: Seeing</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.450)</td>
</tr>
<tr>
<td>H15: Relativity</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.882)</td>
</tr>
<tr>
<td>H16: Motion</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.347)</td>
</tr>
<tr>
<td>H17: Time</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.603)</td>
</tr>
<tr>
<td>H18: Space</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.943)</td>
</tr>
<tr>
<td>H19: Work</td>
<td>DES &gt; SD</td>
<td>Not Supported (p=.884)</td>
</tr>
<tr>
<td>H20: Achievement</td>
<td>SD &gt; DES</td>
<td>Not Supported (p=.356)</td>
</tr>
<tr>
<td>H21: Question</td>
<td>DES &gt; SD</td>
<td>Not Supported</td>
</tr>
<tr>
<td>H22: Emotion</td>
<td>DES &gt; SD</td>
<td>Not Supported</td>
</tr>
</tbody>
</table>

The current study developed a custom dictionary based on theoretical differences existing between SD and DES approaches as delineated by Tako (2009) and others (as illustrated in the details of Table 1). The current study used the processes described by Pennebaker et al. (2007) to transform the theoretical constructs into a useful dictionary of key domain-specific words. Several iterations of development were conducted during which time overlapping terms, non-discriminating terms and other words were excluded. Table 4 provides examples from the SD-DES custom dictionary. An ‘*’ following a group of characters indicates various forms of the word are counted. For instance, activit* would indicate ‘activities’ and ‘activity’ are both counted.
Table 4 SD-DES Custom Dictionary Examples

<table>
<thead>
<tr>
<th>Dictionary Word</th>
<th>DES Related Word</th>
<th>SD Related Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstract</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>accurate</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>activit*</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>adjustment*</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>aggregate</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>analytic</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>asset*</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>behavior</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>buffer*</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>causal</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>close</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

3.3 Results Using the Custom LIWC Word Dictionary

Following development, the custom SD-DES LIWC dictionary was applied to the ten problem statements described in Table 2. Each problem statement was broken into 3 segments by LIWC which yielded an overall sample size of 30 text snippets. The custom dictionary generated an ‘SD’ and ‘DES’ score for each text sample which was classified as either correct or incorrect based on whether it matched the SD or DES methodology used to solve the problems in the source articles. The outcomes were statistically analyzed using PROC ANOVA in SAS 9.4. The results were significant (F=24.0, p<.0001) indicating the custom dictionary accurately selected the correct approach based on a comparison of scores. Table 5 provides a compilation of the analysis showing the expected solution technique and scores related to the correct choice and the incorrect choice.

Table 5 SD Versus DES Scores and Expected Categories for Problem Statements

<table>
<thead>
<tr>
<th>ID #</th>
<th>Problem Domain</th>
<th>Solution Technique Expected</th>
<th>Length in Words</th>
<th>Expected Technique Score</th>
<th>Score from Other Technique</th>
<th>Difference in Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UK Prison System*</td>
<td>SD</td>
<td>462</td>
<td>0.00</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>2</td>
<td>Hospital System</td>
<td>DES</td>
<td>313</td>
<td>1.84</td>
<td>0.31</td>
<td>1.53</td>
</tr>
<tr>
<td>3</td>
<td>Family Practice System</td>
<td>DES</td>
<td>509</td>
<td>1.33</td>
<td>0.95</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>Snack Foods Warehouse</td>
<td>DES</td>
<td>421</td>
<td>2.84</td>
<td>0</td>
<td>2.84</td>
</tr>
<tr>
<td>5</td>
<td>Water System Model</td>
<td>SD</td>
<td>527</td>
<td>2.46</td>
<td>0.76</td>
<td>1.70</td>
</tr>
<tr>
<td>6</td>
<td>Urban Growth</td>
<td>SD</td>
<td>419</td>
<td>6.40</td>
<td>0.47</td>
<td>5.93</td>
</tr>
<tr>
<td>7</td>
<td>Swamping</td>
<td>SD</td>
<td>477</td>
<td>2.72</td>
<td>0.63</td>
<td>2.09</td>
</tr>
<tr>
<td>8</td>
<td>Lean Manufacturing</td>
<td>DES</td>
<td>563</td>
<td>2.11</td>
<td>0</td>
<td>2.11</td>
</tr>
<tr>
<td>9</td>
<td>Words of Mouth</td>
<td>SD</td>
<td>441</td>
<td>1.5</td>
<td>1.29</td>
<td>0.21</td>
</tr>
<tr>
<td>10</td>
<td>AGV System</td>
<td>DES</td>
<td>478</td>
<td>2.67</td>
<td>0</td>
<td>2.67</td>
</tr>
</tbody>
</table>

* the analysis was not able to conclusively provide a recommended modelling approach for this case.

4 DISCUSSION AND CONCLUSIONS

As can be seen in Table 5, the results of the current study indicate that all but one case exhibits underlying linguistic clues which match the solution technique used by experts conducting the simulation study ultimately used in the problem solution. This preliminary study supports our belief that tools such as LIWC can be useful in capturing domain specific language and making this knowledge available to those seeking insight into problem solving techniques. As discussed previously, decision makers come from a variety of backgrounds and may bring preconceived biases. Therefore, it becomes helpful to add another dimension
to the ‘selection of modelling/tool technique’ step in the simulation lifecycle and gain additional insight.

A further relevant observation coming from Hochberg, et al. (2014) in their study of medical decision making, suggests that “in mission-critical environments, linguistic markers of decision-making style may be used to determine the optimal modes of reasoning for a particular task in high-stakes human factors domains.” This consideration is worth noting in fields that use SD and DES since expensive and potentially life-impacting decisions are being made.

Little insight resulted from our first attempt to detect differences using the default LIWC dictionary. This outcome was consistent with past studies and further indicates that specialty domains are likely to include specialized language. Therefore, custom dictionaries offer the greatest utility for these situations. This was further illustrated by our second approach which developed and utilized a custom SD-DES LIWC dictionary to discriminate between ten sample problem statements. Nine of the ten discriminated as hypothesized with higher scores tabulated for what the study authors considered the best approach to solving the problem. The one exception being the UK Prison Case (#1) which had been specifically worded to appeal equally to both SD and DES modellers (Tako & Robinson, 2008). The other nine cases scored significantly higher in the area expected.

The overall result of the current study is to encourage further research in this area. We believe that linguistic clues exist in problem statements and can be used to help select appropriate tools for arriving at solutions.

Study in this area is continuing with the analysis of additional problem statements to fine-tune and improve the custom dictionary. The ultimate goal of this research is to better understand problem-solving techniques and better determine appropriate decision support tools in applied settings.

REFERENCES


Lane, D. C. (2000). You just don’t understand me: models of failure and success in the discourse between system dynamics and discrete event simulation. *LSE OR*.


McHaney, Tako, and Robinson


AUTHOR BIOGRAPHIES

ROGER McHANEY is a Professor of Management Information Systems and University Distinguished Teaching Scholar at Kansas State University, College of Business. He is the author of books on simulation, education technology and Web 2.0 applications. 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 the use of data-driven simulations.

ANTUELA A. TAKO is a Lecturer in Operations Research at Loughborough University, UK. She holds a PhD in Simulation and an MSc in Management Science and Operational Research from the University of Warwick. Her research focuses on the comparison of simulation approaches (Discrete-Event Simulation and System Dynamics), facilitated and participative simulation modelling and conceptual modelling. She is an Associate Fellow of the Operational Research Society (AFORS), UK.

STEWART ROBINSON is Acting Dean and Professor of Management Science at the School of Business and Economics, Loughborough University. 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 Simulation Workshop conference series. He is Past President of the Operational Research Society. Home page: www.stewartrobinson.co.uk.
IMPACT OF MODEL FIDELITY IN FACTORY LAYOUT ASSESSMENT USING IMMERSIVE DISCRETE EVENT SIMULATION

Alessandro Petti, Windo Hutabarat, John Oyekan, Christopher Turner, Ashutosh Tiwari
Cranfield University
College Road, Bedford MK43 0AL
a.petti, w.hutabarat, j.o.oyekan, c.j.turner, a.tiwari@cranfield.ac.uk

Neha Prajapat, Xiao-Peng Gan, Nadir Ince
Alstom Power
Newbold Road, Rugby CV21 2NH
neha.prajapat, xiao-peng.gan, nadir.ince@power.alstom.com

ABSTRACT

Discrete Event Simulation (DES) can help speed up the layout design process. It offers further benefits when combined with Virtual Reality (VR). The latest technology, Immersive Virtual Reality (IVR), immerses users in virtual prototypes of their manufacturing plants-to-be, potentially helping decision-making. This work seeks to evaluate the impact of visual fidelity, which refers to the degree to which objects in VR conforms to the real world, using an IVR visualisation of the DES model of an actual shop floor. User studies are performed using scenarios populated with low- and high-fidelity models. Study participants carried out four tasks representative of layout decision-making. Limitations of existing IVR technology were found to cause motion sickness. The results indicate with the particular group of naïve modellers used that there is no significant difference in benefits between low and high fidelity, suggesting that low fidelity VR models may be more cost-effective for this group.

Keywords: Layout Design, Virtual Reality, Oculus Rift, Factory Layout, Virtual Environment, Discrete Event Simulation, Model Fidelity

1 INTRODUCTION

Shop floor layout design influences many aspects of productivity and costs. For instance, shop floor material handling tasks accounts for 15-50% of manufacturing costs and a good layout could drastically reduce that expenditure (Cullinane and Freeman 1985). An effective utilisation of available area could allow multiple tasks to be performed using the same equipment and labour force, further reducing costs. An optimised layout also helps to avoid bottlenecks that lead to the decrease of the work in progress and queuing time, thereby speeding up execution of orders (Fu and Kaku 1997). Furthermore, a well laid out facility contributes to effective health and safety implementation and control of operations.

Shop floor layout design can be optimised using Discrete Event Simulation (DES). DES works by modelling system state changes occurring at specific points in time, which are probabilistically determined by historical data. In DES, analysis of the past data is essential as well as the definition of the probability distribution that defines the activity duration. These activities are essential because DES entities are processed by passing through the activities and following defined rules, which can be easy or extremely complex (Brailsford and Hilton 2001). DES is suitable for comparing different layout scenarios of production lines or linear processes as it allows to easy visualization of computer animations, metrics and graphs (Sweetser 1999).
1.1 Discrete Event Simulation and Virtual Reality Visualisation

In the mid-1990s, DES implementations began to take advantage of 3D visualisation techniques. Its increasing popularity and its potential benefits compared to 2D are being recognised (Akpan and Brooks 2014; Akpan and Brooks 2012). Akpan and Brooks (2012) surveyed the opinions about benefits and drawbacks of 3D among the users of 3D and 2D simulation. Users reported that problem definition, model validation, scenario experimentation, error detection can all be improved by using 3D visualisation. Also 3D models are said to be easier to understand, more usable, and have higher credibility, hence the ability to demonstrate models to clients in 3D is highly valued. On the downside, users opined that 3D models are more difficult and take longer to build, require significantly more time and more expensive to produce than 2D.

Recently, due to the fast development of the technology and the new hardware available on the market, Immersive Virtual Reality (IVR) 3D visualisation has become more widespread. In order to test the feasibility of IVR, researchers have used both the Head-Mounted Display (HMD) and CAVE technologies interchangeably. IVR technology gives the user the psychophysical experience of being surrounded by a virtual environment (van Dam et al. 2000), providing new opportunities for visualising systems and their behaviours. There are other reasons why IVR can be better than standard 3D: IVR provides global context due to peripheral vision, it allows more natural exploration of the environment in which the user is immersed, provides more clues useful to investigate data quicker and reveal patterns in it. Also, in IVR, people replicate the same spatial errors as in the real world (e.g. the over- or under-estimation of height or width) (van Dam et al. 2000).

Apart from these generic benefits, IVR can offer specific advantages for DES especially in manufacturing. Korves and Loftus (2000) demonstrated the usefulness of IVR in layout planning by comparing it with non-immersive 3D visualisation. Their experiment was about assessing the position of objects (i.e. tables used for assembly and bolt-feeder), the position of parts and the material flow in a manufacturing workplace. They found the use of IVR and 3D for workplace appraisal do not show significant performance differences, but IVR was much better in the detection of flaws. A research on layout planning of a Manufacturing Cell (MC) shows that by using IVR the planner is able to quickly assess the layout and easily detect collisions between machines (Korves and Loftus 1999). Moreover, IVR can offer an immediate qualitative feedback if a new machine is to be setup in the factory or the consequences of the rearrangement of the pre-existing equipment (Smith and Heim 1999). A few of these benefits are summarised in Table 4.

1.2 Fidelity in Immersive VR

There has been many experimental investigations into the suitability of IVR for creating a sense of presence, dimension and distance perception, architecture evaluation, and training. The first experiments aimed to test the sense of presence in IVR as the one by Hendrix and Barfield (Hendrix and Barfield 1995): the experiment consisted of assessing the effects of auditory and visual display parameters on the level of presence through a subjective questionnaire. The hypothesis was that higher levels of parameters would have increased the sense of presence; it was found that only the visual display had an impact on presence (Hendrix and Barfield 1995).

A related experimentation in IVR has been on the perception of dimensions and the factors that can affect it. One of the first experiments dates back to 1997, when participants had to make judgments about object locations in a room based on observations in the real world, in VR, or in a picture. The results suggested that real world and VR not be significantly different, but in both cases people tend to focus their attention on objects themselves rather than on the location of objects (Arthur et al. 1997).

In a training experimentation with coal mine workers (Grabowski and Jankowski 2015) two different levels of fidelity were used, where fidelity refers to the quality of immersion. The high immersion fidelity IVR consisted of a HMD with 110 degrees field of vision (Oculus Rift DK1) along with head and hand tracking devices to aid navigation. In comparison, the moderate-fidelity IVR setup included a HMD with
45 degrees field of vision, a wired Razer Hydra controller and a joystick for navigation. The results show that the better immersion resulted in a more user friendly and effective training.

### Table 4 Qualitative comparison between 2D, 3D and IVR used in DES (After Akpan and Brooks, 2012; Korves and Lofus, 1999)

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>2D DES</th>
<th>3D DES</th>
<th>IVR DES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Problem Definition</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Fast Model Validation</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Easy Experimentation</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Easy to build</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low cost</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Error Detection</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Easy Demonstration to client</td>
<td>- -</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Workplace Appraisal</td>
<td>- -</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Easy Material Flow Detection</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Easy Collision Detection</td>
<td>- -</td>
<td>+</td>
<td>++</td>
</tr>
</tbody>
</table>

IVR offers the opportunity to recreate a virtual environment that approaches reality and this may be useful for applications in layout design. There is a lack of investigation about the claimed advantages of an increase of fidelity for layout decisions in a manufacturing environment. This work aims to address this gap by experimentation centred on Oculus Rift DK2 technology.

### 2 MODEL DEVELOPMENT

As a prerequisite to experimentation, a shop-floor model needs to be prepared in both high fidelity and low fidelity versions. Lanner’s Witness version 14 was used as the main engine for reproducing an existing manufacturing shop floor and its VR capability was used to create the 3D model of the facility. The WITNESS option to stream the running simulation model to full immersive VR was chosen. This uses a dedicated VR software, Visionary Render from Virtalis, and this was used to add additional animations to the objects and to enable the IVR of the shop floor through an Oculus HMD. A 3D design software, Sketchup Pro 2015, was adopted to re-produce objects and equipment in the 3D environment.

Increase of fidelity refers not only to a better rendering of the quality of graphics and how close it is to reality (Figure 1), but it also concerns a higher degree of animations of objects and how they interact with users, in order to make IVR more similar to reality during experiments. For instance, a user employing top-down view cannot make out detailed movements on the shop floor, while a user walking around at head level will appreciate animations on tools and workpieces on the shop floor (Figure 2).

### 3 USER STUDY

A guiding research question for this work is: “Does a higher level of visual fidelity improves layout decision making?” In literature, investigation into the impact of IVR fidelity led to the discovery that higher fidelity of the environment helps to provide a more precise distance judgment (Phillips et al. 2009; Kunz et al. 2009). Consequently, this study starts with the belief that the results gathered from the high fidelity scenario would be better than the ones in low fidelity.

The design of the user studies was based on a previous research by Akpan and Brooks (2014) in which several experiments were carried out to compare 2D and 3D view in DES by using Witness. This framework tested two important task types in DES: observation of processes and completion of tasks. Along the way it also analysed the effect of fidelity between 2D and 3D.
Despite the fact that Akpan and Brooks’s experiments provided a comprehensive spectrum of investigation in DES, they did not consider the IVR. This work adopts Akpan and Brooks framework and investigates the same task types. However, to accommodate IVR, the tasks themselves were adopted from the research by Korves and Loftus (1999), where they showed the advantages of IVR, as presented in Table 4. Four tasks were performed in strict sequence. The tasks, task types, and timings of the user study experiments is provided in Table 5.

**Table 5 Name, type, and time limits for each task of the user study**

<table>
<thead>
<tr>
<th>Task</th>
<th>Name</th>
<th>Task Type</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Understand the system behaviour</td>
<td>Observation</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Detection of material flow difficulties</td>
<td>Observation</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Spot the error</td>
<td>Observation</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Position a new equipment</td>
<td>Task completion</td>
<td>5</td>
</tr>
</tbody>
</table>

Task 1 was aimed to test the two claimed advantages of IVR, which were (a) faster problem definition and (b) better workplace appraisal. In 10 minutes’ immersion, test participants must comprehend the activities performed within the facility and identify important steps in the process. Task 1 is evaluated through eight questions: four questions on number of elements, in order to understand if a higher/lower degree of fidelity affected the participants’ concentration; then four questions on process sequencing, to elicit participants’ comprehension of the process.

Task 2 was aimed to test the claimed IVR ability to help users detect material flows faster, as cited in research by Korves and Loftus (1999). Participants observes the process for 5 minutes then try to identify potential locations of material handling trouble spot.
Task 3 was related to the stated IVR advantage of detecting errors. An error was introduced on purpose in the model and the participant was challenged to spot it as soon as possible.

Task 4 was about task completion. As this was the last task, it was assumed that the test participant should be quite familiar with the environment, having understood the process to a high degree. They were asked to find the best location to position a new equipment. This models a principal activity in layout decision-making, and is consistent with the advantage of using IVR for collision detection.

Fourteen participants attended the IVR DES user study session and were assigned the tasks above. They were divided into equal numbered sets for high fidelity and low fidelity scenario. It should be noted that the participants were not highly skilled in simulation modelling and also did not have any previous knowledge or experience of the facility under examination.

4 RESULTS

It was discovered that motion sickness caused by the Oculus was a key challenge, with 64% (9 participants) suffering from motion sickness after and during the user studies. Only 36% did not feel sick (Figure 3 left). More precisely, looking at the 9 participants, 67% (6 participants) performed the user study in high fidelity, whereas 33% (3 participants) in low fidelity. Due to the motion sickness, there is participant dropout after almost every task (Figure 3 right). This suggest that there is a need for a more detailed investigation of the causes of nausea in high fidelity. It should be noted that other non-immersive and immersive ways of viewing Witness simulation models in 3D do not cause this motion sickness.

Performance data is collated and normalised based on the average score obtained in each task. Figure 4 shows the comparison between the normalised scores in high and low fidelity scenario. Figure 4 also depicts a trend for the scores obtained in low fidelity scenario to increase over time. Two different hypotheses allude to this: either participants became increasingly familiar with the immersive environment, impacting their performance; or the increase is because the reduction of participant numbers led to an increase in the average score. The latter hypothesis can be immediately discarded as the trend of high fidelity scores do not follow the same principle, even though the number of participants also decreased. From Figure 4, over the execution of the four tasks, there was a trend of improvement in low fidelity and an opposite trend in high fidelity.
The six participants who successfully performed the entirety of the user study were split evenly between participants in low fidelity study and in high fidelity so their views can be directly compared from their answers to the evaluation questionnaire. This is reproduced in Appendix A of this paper. At the end of the user studies two out of three participants in high fidelity felt sick, while nobody in low fidelity reported any symptom of nausea. Analysing the comments of the “high fidelity participants”, they indicated that they were distracted by the motion sickness, causing them to not pay enough attention to the task. In fact, more participants in the high fidelity scenario felt sick than low fidelity ones.

It is also interesting to note and compare the total time spent in IVR between the two groups. Altogether, the high fidelity participants spent 242 minutes in immersive environment, while low fidelity participants spent only 197 minutes.

5 DISCUSSION

The result of the experiments do not support the hypothesis that a higher degree of visual fidelity would aid layout decision making. In fact higher visual fidelity seems to correlate with lower performance. Participants in low fidelity scenario obtained higher scores (+37% on average) than the opposite group. Thus, the additional efforts and time to develop the high fidelity scenario, five times more than low fidelity does not appear to be cost effective. Nevertheless, the lower performance of high-fidelity IVR seem to be due to motion sickness, which is most likely due to the shortcomings of the test equipment rather than the inherent characteristics of high fidelity immersive environments.

Additionally, it was found that the text labels attached to low fidelity parts may have influenced the outcome of the user studies since it is easier to understand which part is which by reading a label rather than looking at details. This labelling was unavoidable due to the lack of other distinguishing marks at low fidelity. This effect is more evident analysing the results of task 3, where participants had to spot one error, done on purpose, in the model. The error consisted in the substitution of the wheels with the bolster in one assembly step of the process; participants in low fidelity both answered faster and achieved a higher score than in high fidelity. This is supported by participant comments, where low fidelity IVR users confirmed that task 3 was easy to carry out because the labels helped them.

Motion sickness greatly affected participants performance and the experimental results, particularly in the high fidelity scenario. Generally, in high fidelity six over seven participants (86%) felt sick. This merits
further analysis. The only participant who did not suffer nausea is a heavy videogame user and has had more than 5 years’ experience in simulation, suggesting that experience may help in combating motion sickness. However, contradicting this, another heavy videogame user participating in high fidelity scenario did suffer motion sickness. After further investigation, it was found that the first participant used a fast mode, while the second opted for a slower setting. The contradictory finding may have been caused by different selection of the simulation rate of movement.

Regarding the increased period of immersion for high-fidelity scenario participants, a possible explanation could be that in high fidelity scenario participants were distracted by the amount of environment details, e.g. losing time to look at the accuracy of the machineries or parts. This increases the likelihood of motion sickness and consequent reduction of performance (Figure 5).

To fully understand the root causes of the motion sickness a detailed analysis of the participants’ comments after the user study is necessary. They reported display problems during the user studies: a continuous flashing along the most complex shapes, some object edges were jagged and a level of jerky object motion. The most likely reason for this is hardware and software limitations. Despite the workstation-class graphic cards used (NVIDIA Quadro K5000M), the load from the high-fidelity scenario caused some motion freezing and consequent choppy movements of objects. The Oculus screen display was not sufficiently detailed causing the aliasing or staircase effect (jagged effect). It is reasonable to attribute this effect to Oculus Rift because the aliasing was not evident on the computer screen at all. On the other hand, after studying various videos of the Oculus Rift on internet and having tried the demo provided with the device, the staircase effect was not evident. Therefore this problem may be possible to overcome with different Oculus drivers and settings.

Finally, the strategy adopted to carry out the tasks is another important factor influencing motion sickness. Two different strategies were observed: walking around in the facility (‘walking strategy’), adopting a stationary position above the facility looking at the process from the top down (‘bird eye strategy’). In one instance, a participant decided to walk around in the first task, stopping one time, while in the second task adopted chose the Bird Eye strategy and finally for the third task reverting to walk again. After the third task, the participant was overcome by motion sickness. Maybe due to BE strategy in the second task he was able to keep doing the user study, as he did not move too much. Looking at his profile, he never played videogames and performed the user study in high fidelity scenario, so the probability of motion sickness was very high.

It was found that participants who adopted the bird eye strategy and minimised their movement in immersive environment did not feel sick. However, all participants who fully adopted the bird eye strategy belong to the low fidelity user study group. It is thus impossible to make any conclusion about which are more likely to reduce motion sickness: the movement-minimising strategy or the low fidelity environment itself.
6 CONCLUSION

The continuous changing of the market and the increase of variability of customer demand is pushing future factories to be more flexible. Among the many analytical tools available, Discrete Event Simulation (DES) and Immersive Virtual Reality (IVR) are among the most innovative and promising.

This research has presented the study of the impact of visual fidelity, which refers to the degree to which objects’ aspect in VR conforms to the objects’ aspect in the real world, on layout decision-making. Two scenarios with different fidelity (high and low) of a real facility have been developed and two Witness models have been created using the shop floor data gathered. Fourteen respondents have attended the user study, where both the scenarios and the simulation model have been used.

It was found that there is a decreasing trend of participants’ performance throughout the user study, particularly in high fidelity due to the increase of nausea symptoms. This include users dropping out even only after performing one task due to motion sickness.

The results of the user studies show that there is not a significant difference between high and low fidelity results and, considering the time and the efforts made to create the high fidelity scenario, low fidelity results to be a more cost-effective solution.

Additional findings indicate that the adopted IVR system caused motion sickness in the majority of the participants. Fidelity is also found to be not the only important concept in layout-decision making. For example, the visualisation of the simulation time, which in this example was not displayed in IVR, has an impact on layout decisions as it helps, and maybe speeds, the understanding of the process and the events.

Therefore, these two findings can be used as a base for future researches in IVR. Firstly, one work should investigate if the suspected causes of motion sickness are correct and if a further investment in a more recent and powerful hardware can lead to different conclusions about fidelity. In the case in which the high fidelity helped to achieve better scores, a new research should investigate the extent to which the details have to be represented, that is the right trade-off between efforts and degree of visual fidelity.

Secondly, there is a need to statistically validate the conclusions of the project thesis and a future research can be focused on this. The sample size used (14 participants) cannot be used to give statistically significant results but it contributes to the growing body of data on the use of VR in manufacturing.

Finally, future investigation may need to ensure that participants do not try to adopt strategies which renders them immobile (‘bird eye strategy’). It is suggested to avoid that behaviour for two reasons: it skewed the results and it is not in keeping with the idea of IVR, because it essentially employed the IVR to provide a glorified 2D plan view.

It should be noted that there are many other aspects of this work that can also lead to further study:

- Model fidelity is a continuous scale and the terms used in this paper of “high” and “low” are relative. There are also many different standards of high fidelity view in Virtual Reality and this study has only used one example. View complexity and fidelity are not the same thing and complexity can radically affect 3D performance. There is much more to learn about the appropriateness of different types of views for this type of model.
- The types of participant in a study may affect results. Users of a model may typically be more familiar with a facility than was the case here. This domain knowledge may affect greatly the interaction and understanding of the model.
- Although this study concentrated on IVR views with Oculus there are many other ways for immersive VR to be viewed, for example on large wall displays with 3D glasses. The results for model fidelity may be different in this type of environment and for individual or multiple concurrent use.

ACKNOWLEDGMENTS

The authors acknowledged the support of Innovate UK and EPSRC, provided through grant number EP/M506813/1 “Towards Zero Prototyping of Factory Layouts and Operations Using Novel Gaming and Immersive Technologies”. All data supporting this study are provided as supplementary information accompanying this paper and can be accessed by contacting researchdata@cranfield.ac.uk.
A APPENDIX: SELECTED PARTICIPANTS’ COMMENTS

- I felt dizzy. The 3D space mouse is ok but the Oculus is very tiring. I saw blur and flashing when I moved. It was not easy to understand the process because I did not see the operations.
- I have the feeling of nausea. I felt distressed. The Oculus is heavy for my nose. After a while, I think the animation affected my sensation.
- I am feeling good. I would like to see the parts from one stage to the other, otherwise it is difficult to understand the right process. I did not like the fact that the computer was freezing.
- I lost time to ensure not to be sick, so I could accomplish the task.
- I could not see all facility because the setting was too slow but it was necessary otherwise I would feel sick.
- I am feeling less good than before. I could not see far from me because of the lens of the Oculus. I would like to be faster but I couldn’t because when I did I felt not so good.
- I feel good. The equipment is fine and user friendly. The label on the parts helped me during the experiment, very easy.
- I am a little dizzy. The Oculus was slow and I cannot move my head because it gave me instability. The animations were too blocky and finishing the task was too difficult.
- I am not good. The mouse is ok, but the resolution of the Oculus was too low may it influenced my feeling. I spent more time not to be sick than memorise and finish the task.
- I am good. The interaction between mouse and Oculus needs improvements: it is unnatural to move the head without moving in the same direction. All stations looks the same, I wanted more details.
- I am ok. At the start was a little difficult and noisy, then I familiarised and everything was perfect.

REFERENCES


**AUTHOR BIOGRAPHIES**

**ALESSANDRO PETTI** received an MSc in Manufacturing Consultancy with first class honours from Cranfield University in 2015 and his first degree (cum laude) from Politecnico di Bari on Management and Industrial Engineering. He has performed research projects with multi-national companies such as Airbus, Alstom Power, Boeing, and Alenia Aermacchi.

**WINDO HUTABARAT** is a Research Fellow at the Cranfield’s Manufacturing Informatics Centre, where he pioneered the idea of applying the innovations from the gaming sector into manufacturing and services that resulted in multiple research and development projects funded by Innovate UK. His interest is in optimisation, simulation, and computational intelligence.

**JOHN OYEKAN** is a Research Fellow in the Manufacturing Informatics Centre at Cranfield University. He has successfully worked on several EPSRC and Innovate UK projects in the areas of computational intelligence, robotics and human-machine interaction with various industrial partners such as Rolls-Royce, Airbus, and Jaguar Land Rover. His interest lies in sensing and computational algorithms for various applications. He is currently investigating the use of low cost gaming devices to improve productivity in manufacturing.

**CHRIS TURNER** is a Research Fellow and Project Manager within the Manufacturing and Materials Department at Cranfield University. He is research active in the fields of manufacturing simulation, manufacturing informatics, business process management and business process optimisation. He was involved in the successful completion of several UK research council funded projects with subjects ranging from business process optimisation to the simulation of product-service systems.

**ASHUTOSH TIWARI** is the Professor of Manufacturing Informatics and the Head of the Manufacturing Informatics Centre at Cranfield University which is internationally recognised for its research in optimisation of high-value manufacturing processes and digitisation of skill-intensive manufacturing. He has developed a strong research track record by leading, as PI, 8 EPSRC projects, 4 Innovate UK projects, 1 AMSCI project, and 3 projects funded by EPSRC Centres for Innovative Manufacturing.

**NEHA PRAJAPAT** is a research engineer in Alstom Power. Neha Prajapat received her BSc in Mathematics from the University of Leicester in 2012. She is currently a Statistical Methods Engineer in the Methodologies for Tools department at Alstom Power in Rugby. She is also pursuing her PhD in Manufacturing Process Optimization at Cranfield University.

**XIAO-PENG GAN** is the Head of Optimization and Robust Design group within the Methodologies for Tools department at Alstom Power. He received his BSc in Mechanical Engineering from the University
of Zhejiang, China. He completed his PhD in Mechanical Engineering at University of Bath. His research interests include: optimisation methodologies, statistical analysis, robust design, and thermal analysis.

**NADIR INCE** is the Head of Methodologies for Tools department at Alstom Power. He attended Technical University of Istanbul where he completed his BSc in Mechanical Engineering. He completed his MSc and PhD at University of Manchester in Experimental and Computational Modelling of Natural Convection. He is a member of the UK NAFEMS steering group committee and chairman of the Optimisation working group. He is also a UK industrial committee member of Ercoftac.
ABSTRACT

Patients presenting with chest pain at an emergency department (ED) in the UK receive troponin tests to assess the likelihood of an acute myocardial infarction (AMI). Until recently, two blood samples separated by at least six hours were necessary to analyse the change of troponin levels over time. New high-sensitivity troponin tests may allow rule-out of AMI on the basis of a single test if the first blood sample is taken at least three hours after onset of the worst pain. This paper presents a discrete event simulation model which shows the likely impact on the numbers of hospital admissions to short stay wards if emergency departments adopted a single test rule-out strategy based on the use of high-sensitivity troponin tests. Data sets from acute trusts in the South West of the UK are used to quantify the resulting benefits.

Keywords: discrete event simulation, emergency department, diagnostic pathway, avoidable admissions

1 INTRODUCTION

Chest pain is one of the most common reasons for patient attendance at UK emergency departments and represents a significant workload for EDs. In terms of presenting complaints, chest pain is responsible for between 2.4% - 6% of visits to EDs in the UK per year (Goodacre et al., 2005). Many such patients do not require hospital admission, because they do not have acute myocardial infarction (AMI) or unstable angina. The National Institute for Health and Care Excellence (NICE) recently issued guidelines which propose serial troponin tests in order to safely rule out myocardial infarction (NICE, 2014). NICE recommends that the first sample is taken upon arrival of the patient at ED and that the second sample is taken at least three hours after the first. It is argued that high-sensitivity troponin tests – either Troponin T or Troponin I – can be used to safely rule-out AMI earlier than with previous tests and they recommend a shortening of the inter-test time from 6-12 hours to three hours (NICE, 2014). However, this practice of serial testing with an interval of at least three hours forces EDs to admit patients to short stay wards to be able to meet the 4 hour target for overall patient length of stay in ED. The second sample is commonly taken on a short stay ward which sometimes is still under control of the ED. In contrast to this, there is evidence that a negative
first test could be a sufficient basis to rule out AMI (e.g. Zhelev et al., 2015; Body et al., 2011; Rubini et al., 2013). Shah et al. (2015) argue that, if the first sample was taken at least two hours after the onset of pain, patients could be safely discharged home. They conducted a large trial and explicitly highlight that patients presenting late enough could be sent home on the basis of one negative sample without affecting their safety because the diagnostic accuracy is very good. The European Society of Cardiologists (ESC) also argues that the new high-sensitivity tests are sensitive enough to discharge patients earlier (Roffi et al., 2015).

The work presented in this paper has been carried out in collaboration with seven acute trusts in the South West of the UK and the South West Academic Health Science Network (SW AHSN). The aim of this study is to model the impact on the time patients spend in ED and on short stay wards. It also aims to assess the likely changes in the numbers of admissions to short stay wards if patients could safely be discharged earlier using high-sensitivity troponin tests. Below, we first introduce the major characteristics of the relevant diagnostic pathway for patients with chest pain. This takes into account both the analysis of historic data records and the mapping of the diagnostic pathway used to develop the simulation model. We then introduce the simulation model, highlight the “what-if” scenarios and present the quantitative and qualitative findings. Finally, we summarise this study and discuss further research potentials in this area.

2 THE DIAGNOSTIC PATHWAY FOR ACUTE CORONARY SYNDROME

Acute Coronary Syndrome (ACS) refers to cardiac conditions most often associated with patients presenting with chest pain. Building a model of the diagnostic pathway for ACS is a key focus of our study. This requires an analysis of available data to understand the processes at emergency departments. The collaborating trusts provided both clinical protocols for the ACS diagnostic pathway and historic data records for the targeted patient group (presenting complaint chest pain, attended ED, received troponin test). Extensive dialogues with practitioners at the different trusts and dedicated pathway mapping sessions were a key input to build a conceptual and a formal model. Historic data records were very important in order to populate the simulation model and to validate it. Figure 6 shows the diagnostic pathway for one of the trusts which participated in our study – we will refer to this trust as Trust A.

There are some differences within the diagnostic pathways for ACS across the different trusts, but most of the procedures are common practice. The standardised version of an ACS diagnostic pathway starts with patients arriving, registering and getting triaged subsequently. A blood sample is generally taken during a first (often nurse-led) assessment before a doctor assessment takes place. This blood sample is then sent to the biochemistry laboratory for analysis. Ideally, the results are available before the doctor assessment takes place and can be reviewed during the assessment. Otherwise, patients wait until the results are available and have been reviewed by a clinician. According to these results and the overall medical history, patients then are either admitted or sent home. Admission to a short stay ward is commonly required, where a second sample is taken ≥3 hours after the first one. Some of these short stay wards or lounges are closely connected to the ED and are managed by ED staff. However, referring patients to one of these wards counts as an admission and does affect the current workload for ED staff, unlike discharges. High risk patients with conditions such as STEMI are not included in this pathway – they are usually referred to the catheterisation
laboratory for specialised treatment. As previously mentioned, recent research suggests that part of these admissions could be avoided since the new generation of high-sensitivity troponin tests can be used to safely rule-out ACS on the basis of only a single test.

Analysing historic data records

The collaborating trusts provided historic data records for a period of at least one year. Depending on the size of the trusts, these data sets included between 1,000 and nearly 9,000 patient episodes. Inclusion criteria for the research project were all patients having presented with chest pain at registration and having received at least one troponin test in addition to clinical history, examination and other investigations such as ECG. The patient population used to parameterise the various trust models had slightly more male than female patients, however female patients were on average older than male patients. Based on the historic data records, we estimated process times of the various activities. The vast majority of process times showed daily variation (e.g. arrival to triage, triage to consultation with doctor, etc.). This is explained by varying workload and the numbers of staff allocated to different shifts. We estimated the model parameters in intervals of 4 hours as highlighted in Figure 7 which visualises inter-arrival times for Trust A. Inter-arrival times typically ranged from 60-180 minutes for patients arriving by ambulance (dashed line, grey) and 120-420 minutes for self-presenters (dotted line, black). The accompanying solid lines divide the day into 4 hour segments and, in our opinion, sufficiently capture the overall trends suggested by the continuous plots.

Figure 7  Inter-arrival times for self-presenting patients and ambulance arrivals by hour of arrival.

Similar effects occur across the different trusts which participated in this project. The number of arrivals by ambulance is about 75% out of all arrivals which suggests that patients are generally aware that chest pain may indicate a potentially life-threatening illness. Arrivals of self-presenting patients, whilst only accounting for 25% of the total, are up to four times more frequent during typical working hours. This could reflect the fact e.g. that some of these patients do not think their symptoms have a serious underlying cause, that access to public transport is much reduced outside daylight hours, that patients are aware that EDs generally carry only limited stocks of medications, and that pharmacies are not open. In addition, working patients may be encouraged by employer policies to seek help sooner rather than later. By contrast, ambulance arrivals typically show less variation over the day. A comprehensive analysis of historic data records together with detailed mapping sessions held at each of the trusts formed the basis for the discrete event simulation (DES) model. The model for Trust A will be highlighted in the following section.

3  DISCRETE EVENT SIMULATION MODEL

A large number of studies are available focusing on simulation models for EDs – see e.g. Gul and Guneri (2015) for a recent overview. Only few of these publications provide evidence that the conducted studies have an impact on the actual situation at the hospitals’ emergency departments as pointed out in Brailsford et al. (2009). Thus, our work fills this gap nicely as it is originally driven from the practitioners’ point of
view and is determined to initiate a phase of implementing pathway changes. The DES model used in our study served as a tool for clinicians to assess the impact of changes. Based on a generic DES model which served as a prototype we developed individualised models for each acute Trust. All models were built using the software package Simul8 (Simul8 Corporation, Version 2015). The DES model is divided into the sections Arrivals, Pathway through ED, Sample processing at Laboratory, Admissions and Discharges – Figure 8 exemplifies the structure for Trust A. Steps of the process through ED, the wards patients get admitted to and the corresponding processes at the laboratory were individualised according to local settings.

The underlying flow of patients in this DES model is based on the generic pathway (section 2). All process times depend on the time of day and were aggregated in six blocks of four hours length according to the analysis of historic data records. The decision about further treatment, i.e. whether and where to admit or to discharge a patient, is made during the results review. The likelihood that the patient’s symptoms may have an underlying cardiac cause is estimated by the attending doctor through a combination of their clinical history and examination, electrocardiogram (ECG), a formalised risk score (e.g. GRACE, TIMI), and the serum troponin measurement. Patients whose ECG findings and/or troponin results suggest a diagnosis of either STEMI/NSTEMI or unstable angina, are admitted to cardiac wards for specialised care. Other risk categories include other cardiovascular causes of chest pain, a variable subset of identifiable causes, and unknown causes. Where appropriate, all such patients will be assessed on the basis of their level of troponin. If single testing is possible, patients with a negative first test outcome will be discharged whereas a positive first sample leads to an admission. Should serial testing be required, patients will be admitted to a short stay ward to have a second sample taken, regardless of the outcome (i.e. positive or negative) of the first sample. The time taken to reach the short stay ward depends on various factors such as bed/chair availability, staff workload and overall medical condition of the patient.

What-if scenarios
We conducted simulation studies looking at different time thresholds between onset of worst pain and first sample taken at ED in order to facilitate early discharge of patients. We will refer to these scenarios indicating the time threshold between onset of worst pain and the time point the blood sample is taken. The Status Quo at most EDs is 360 minutes. For each of the scenarios we performed trials of 20 runs each, using independent and different random number seeds. This ensures that 95% of the simulated key performance measures are within 5% of their actual mean. The primary outcome of this study is the change in number of patients able to be discharged with respect to a different time threshold, for a single test rule-out strategy. The time between onset of worst pain and arrival at ED is a key input factor for this analysis. This time interval might vary significantly among different regions within the UK and in an international comparison.
Also, the way this data is currently recorded does not allow to use it for analytic purposes – if at all recorded, it is commonly handwritten on patient records and therefore not accessible. Thus, we performed a basic sensitivity analysis using different input distributions for the interval between onset of worst pain and arrival at ED. We particularly highlight the resulting differences when using a distribution based on a study reported in Hong et al. (2011) and using a distribution adjusted for Trust A. Hong et al. (2011) list key percentiles of the underlying distribution and we identified this time to follow a log-normal distribution (see Appendix). A major drawback is that distributions reported in the literature have considerably large variation which would affect the output of simulation studies. Therefore, the proposed sensitivity analysis enables to identify overall beneficial effects while taking into account that the time from onset of pain to arrival at hospital could vary significantly.

4 PRACTICAL IMPLICATIONS

Focusing again on Trust A, we will first point out likely changes on the number of admissions, particularly to short stay wards. Lowering the threshold for single test rule-outs stepwise from six hours down to two hours has a remarkable impact on admissions to short stay wards. This effect can be seen clearly in Figure 9. The solid (Trust A) and dashed (Hong et al., 2011) lines represent different assumptions on the time from onset of worst pain to arrival at ED as discussed in the previous section. A reduction from 360 to 120 minutes could reduce the number of admitted patients by 30-40%. At the same time, the number of discharges based on a single test can be increased by more than 50%. Already small reductions of the time limit, e.g. by some initial 60 minutes, significantly contribute to the number of early discharges as these increase by at least 25%.

The differences between the solid and the dashed lines indicate that the positive impact on the number of early discharges – when using a single test rule-out strategy – becomes more apparent in case of longer pre-hospital delays. This must not imply recommendations to deliberately delay seeking for help at the hospital. Changing the threshold time leads to similar consequences in terms of increase/decrease in absolute numbers of discharges and admissions to short stay wards respectively. Again, the benefits of lowering the threshold for single test rule-outs become more apparent where longer pre-hospital delays are more common. This can be seen in the overall number of additional discharges which is 2,616 for the adjusted distribution for Trust A and 2,286 for the distribution suggested in Hong et al. (2011).

The positive impression of lowering the time threshold for single test strategies is strengthened by highlighting the impact on average time in ED and the percentage of 4 hour breaches if the threshold is lowered. Figure 10 indicates that a significant reduction of the time spent in ED, of between 38-44 minutes, could be achieved. According to the data sets provided by the trusts, it takes less time to discharge a patient after the results have been reviewed than admitting a patient which requires additional preparatory and administrative work. Modelling the process of admissions into hospital was beyond the scope of this study.
In addition to the foregoing, the percentage of patients predicted to breach the ED 4 hour target decreases by approximately 2\% for every hour that the troponin test threshold can be brought forward (figure 6), using a single test rule out protocol.

The key findings of this study are that first, a reduction in the time threshold for single test rule-out positively influences the number of admissions and the number of discharges respectively. Second, our sensitivity analysis highlighted that the time from onset of worst pain to presentation at hospital is a highly influential parameter (in our model). There are currently only few good estimates of this parameter in the literature, and it is especially unclear how this parameter might vary according to the underlying cause of disease. However, even if these delays could be substantially reduced, our results would still indicate beneficial effects of single-test rule out strategies in terms of reducing the number of admissions to short stay wards.

5 \hspace{1em} CONCLUSIONS AND OUTLOOK

We presented a discrete event simulation model which is part of a bigger study looking at ED pathways for patients presenting with chest pain. This paper originates from a project which analyses the benefits of using high-sensitivity troponin at emergency departments in the South West of the UK. Based on a generic DES model, the presented simulation model was tailored to represent the situation at a dedicated trust. We performed a number of what-if-analyses in order to investigate the impact of introducing a single test rule-out strategy where the time threshold was lowered from 6 hours down to 2 hours. The currently applied time limit of six hours for single test rule-out strategies is historical and there is evidence that the new high-
sensitivity troponin tests can reliably rule-out AMI with shorter waits. The findings presented in this paper suggest that, given that the reduction of the threshold can be done without additional risk, already small reductions of the time thresholds for single test rule-outs can significantly increase the number of discharges based on a single (negative) test. However, we also pointed out that the process at ED is not the critical part of this study. The main input factor is the time between onset of worst pain and arrival at the hospital. Based on this, it can be summarised that the positive effects of avoided admissions become more obvious in settings where there are longer pre-hospital delays. There are thought to be numerous factors driving patients’ decisions to seek help, and these are beyond the scope of our study. Furthermore, the high-sensitivity troponin test is just one of many factors by which patients are directed from ED, and many such patients can be managed without recourse to it. Therefore, our conclusions do not amount to a recommendation to delay hospital attendance when experiencing chest pain. Rather, a central aim of our study is to help Trusts to expedite high-sensitivity troponin testing for those patients who will benefit.

Finally, the model developed in this study can easily be tailored around trust-specific settings which makes it widely applicable, especially when looking at ACS pathways. It serves as a major evaluation tool in follow-up projects which accompany the implementation of pathway changes at some of the collaboration trusts. Independently, quantifying the main drivers to decide whether or not a patient is admitted could be further investigated using Data Mining or Machine Learning techniques. This would then allow to refine the simulation model with its decision logic and explore how additional knowledge about input factors would affect the accuracy of the model output. Together with an analysis of the time since onset of pain and its effects on the outcome, this could underline the positive impact of our study.

ACKNOWLEDGMENTS

The authors would like to thank clinicians and administrative staff at the collaborating acute trusts in the South West region for their valuable comments and support during the process of developing our simulation models. This work is part of a research project founded by the Academic Health Science Network (AHSN) South West. This work is also funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula at the Royal Devon and Exeter NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

APPENDIX

Hong et al. (2011) report the following percentiles for a distribution describing the time between the onset of worst pain and a patient’s arrival at ED. They particularly report a median of 131 min and inter-quartile range of 70 min – 261 min. The following R code was used to identify the best-fit distribution:

```r
library(rriskDistributions)
q <- c(70,131,261)
get.lnorm.par(p=c(.25,.5,.75), q, show.output=TRUE, plot=TRUE, tol=0.0001)
```

The resulting graphical output in terms of a cumulative probability function \( F(x) \) is highlighted in Figure 12 where the black dots indicate the data input. The package has chosen the fitting procedure \( \text{Nelder-Mead} \) and confirmed that the best fit could be achieved using a log-normal distribution with LN (4.89279, 0.97597). Figure 13 and Figure 14 show distributions used in order to model the time between onset of worst pain and arrival at the hospital. The solid lines highlight the data provided in Hong et al. (2011) and the dashed lines represent the adjusted distribution according to the data set provided by Trust A.
Figure 12 Fitting data provided in Hong et al. (2011) to a log-normal distribution using R.

Figure 13 Probability distributions modelling time between onset of worst pain and arrival at ED.
Figure 14 Cumulative distribution functions modelling time between onset of worst pain and arrival at ED.

REFERENCES


Rachuba, Salmon, Zhelev, and Pitt


AUTHOR BIOGRAPHIES

SEBASTIAN RACHUBA received his doctorate in Management and Economics from Ruhr University Bochum (Germany) in 2013. He completed his diploma studies in Management and Economics at the same university in 2008. He is currently a post-doctoral researcher at University of Exeter Medical School working on various projects to improve healthcare systems and service delivery in the South West. Before he joined the University of Exeter he worked at a medium-sized German municipal hospital as a consultant to the Medical Director.

ANDREW SALMON qualified as a doctor/clinician in 1995 and worked in acute and emergency medicine before spending some years as a principal medical investigator for phase I clinical trials. He subsequently completed undergraduate and postgraduate degrees in mathematics and statistics before joining the University of Exeter Medical School as a cost-effectiveness modeller.

ZHIVKO ZHELEV received his master degree in clinical psychology from the University of Sofia (Bulgaria, 1994) and completed an advanced clinical training in systemic family therapy at the University of London (UK, 2001). He worked in a variety of settings as a psychologist and psychotherapist and was involved in a range of health and social care projects that naturally led to his doctoral research on disability discourses in contemporary society. He received his PhD from the University of Plymouth (UK, 2009) and is now appointed as a Research Fellow in Diagnostics at the University of Exeter Medical School (UK, since 2010).

MARTIN PITT is Associate Professor of Healthcare Modelling and Simulation at Exeter University Medical School and Director of PenCHORD – The Peninsula Collaboration for Health Operational Research and Development. He has a long-standing and wide ranging experience in healthcare modelling ranging from economic modelling in Health Technology Assessment to Discrete Event Simulation for operational models for service re-design. He has a specific research interest in promoting the effective implementation of models in healthcare and the use of visualisation techniques to improve the communication of outputs to key user groups.
ABSTRACT

Stroke is a major cause of death and long-term disability world-wide. To improve functional outcome treatment with intravenous tissue plasminogen activator (tPA) is the most effective medical treatment for acute brain infarction within 4.5 hours after the onset of stroke symptoms. Unfortunately, tPA remains substantially underutilized. Acute stroke care organization is among the dominant factors determining undertreatment. Recently, simulation has been suggested and successfully implemented as a tool for optimizing stroke care pathway logistics. Starting from a number of pioneering simulation studies challenges in simulation application and simulation methodology are identified. The definition of a domain specific modelling framework for acute stroke care is advocated to master system complexities, facilitate joint team work in solution finding, organize model data collection and make a further entrance to the field.

Keywords: Simulation modelling methodology, ischemic stroke, stroke systems, hyper acute pathway

1 INTRODUCTION

Acute ischemic stroke is the second leading cause of death and a leading cause of long-term disability world-wide (Truelsen et al. 2005; WHO, 2012). The acute brain infarction is caused by a blood clot. As a result of the blocked blood vessel downstream brain tissue is deprived from oxygen and starts to mortify. The longer the blood clot is existent the more damage is done, resulting in (severe) disability and possibly death.

Treatment with intravenous tissue plasminogen activator (tPA) is the most effective medical treatment for acute brain infarction. Essentially, tPA restores blood flow in the brain by dissolving the blood clot at the root of infarction. tPA has shown to be effective, i.e., improving patient functional outcome, within 4.5 hours after the onset of stroke symptoms. Of all patients worldwide suffering a stroke, 1–8% (Wardlaw et
al. 2009; Adeoye et al. 2011; Singer et al. 2012) are currently treated with tPA, whereas 24–31% (Waite et al. 2006; Boode et al. 2007) would be attainable in optimized settings. The sooner the treatment is started the better functional outcome is i.e. TIME = BRAIN.

Main causes of undertreatment found are the narrow therapeutic time window, patient and bystander unfamiliarity with stroke symptoms and how to act, and stroke care organization. Notably, the benefit of tPA depends strongly on time since stroke onset (The National Institute of Neurological Disorders and Stroke rtPA Stroke Study Group, 1995; Hacke et al. 2008; Lees et al. 2010; Wardlaw et al. 2012; Emberson et al. 2014), which in turn negatively affects the chance of administering tPA treatment as time since onset increases (Hamann, 2004).

In recent years many researchers made proposals for improving stroke care organization, thereby attempting to reduce patients’ delay along the stroke pathway, and – hence – increase their chances for favourable outcomes. Respective research efforts have primarily relied on the use of Randomized Controlled Trials (RCTs) as a main research vehicle. RCTs are meant to establish potential benefits of alternative set-ups of the stroke pathway, by comparing real-life outcomes for two groups of patients: those that traversed the existing pathway and those for whom the pathway has been adapted according to proposed interventions. Although the merits of RCTs for use in health research concerning simple or solitary interventions such as pharmaceuticals or devices are clearly established, complexity of the acute stroke pathway – entailing a sequence of interrelated care services, spanning both the pre-hospital and the hospital phase – may hinder its applicability. For example, two recently proposed improvement programs reported disappointingly low nonsignificant increases in tPA treatment rate of 1.0–1.5% (Dirks et al. 2011; Scott et al. 2013). Clearly, disadvantages of real-life testing in terms of time, costs and efforts involved in study set-up and experimenting, and project organization, management and lead time become apparent here (Law, 2007).

Recently, several studies have shown how simulation may be used as an efficient alternative or precursor to clinical trials (Monks et al. 2012; Pitt et al. 2013; Churilov et al. 2013; Lahr et al. 2013a; Lahr et al. 2013c; Jacobson et al. 2015; Komenda et al. 2015). Whereas clinical trials are limited to testing a seemingly arbitrarily selected set of interventions along the stroke pathway, the efficiency of computer simulation models allows for a far greater set of interventions to be put to test at minimum efforts and means. Moreover, where efforts in setting up clinical trials often restrict studies to just a part of the stroke pathway, simulation can easily cope with interventions addressing the entire stroke pathway. Furthermore, simulation efficiency does not only allow for redesigning existing stroke pathways but also rethinking the way care chains may be set-up for a region. For example, should stroke care be offered at every community hospital (i.e. decentralize facilities), or should care be concentrated in a comprehensive stroke center (i.e. centralize facilities)?

Apart from aforementioned efforts simulation hardly made an entrance in the field. In this article we identify and study modelling challenges faced in addressing the field, thereby relying on our experiences in doing some of the pioneering studies for the field (Monks et al. 2012; Pitt et al. 2013; Lahr et al. 2013a; Lahr et al. 2013c). Our findings are meant to accelerate studies in the field (i) by identifying hurdles – and ways to overcome these – for those who consider undertaking simulation studies, and (ii) reveal research requirements for simulation modelling methodology.

The remainder of the paper is organized as follows. In Section 2 we shortly typify the hyper acute stroke pathway. In Section 3, we address challenges faced in simulation application, concerning complexities in system design, innovations impacting future system design, and project organization. Next, in Section 4 we address methodological challenges faced in simulation use for stroke system optimization. Finally, we will discuss and summarize main findings (Section 5 and 6).

2 STROKE SYSTEMS – THE HYPER ACUTE PATHWAY

Stroke is categorized in two sub types: ischemic and hemorrhagic, relating to about 85% and 15% of the patient population respectively. Ischemic strokes occur when a blood vessel is blocked due to a clot and
disrupts blood circulation to the brain, whereas hemorrhagic strokes boil down to a bleeding. Only ischemic strokes are eligible for tPA treatment, see Section 1.

The stroke pathway entails several phases. Here we only consider the initial phase, the so-called hyper acute (emergency) phase, see Figure 1. The hyper acute phase spans care services from stroke onset to tPA treatment. Patients either arrive at the hospital by Emergency Medical Services (EMS) after making a call to their GP or the emergency number (112) or by self-referral. Next, a patient’s eligibility for tPA treatment depends on his/her delay along the pathway (treatment must be administered within < 4.5 hours of the onset of symptoms) together with the results from the neurological examination, laboratory evaluation (blood testing) and neuroimaging examination (CT-scanning). tPA treatment influences patient outcomes in terms of the severity of their disabilities, ranging from no disabilities, to major disabilities or death. Note that a small number of the patients faces a stroke while being hospitalized.

Figure 1 Hyper acute stroke pathway – dominant set-up

3 CHALLENGES IN SIMULATION APPLICATION

Until a few years ago hardly any simulation study had been performed for optimizing the hyper acute stroke pathway. Below we address modelling challenges encountered in our recent studies (Monks et al. 2012; Pitt et al. 2012; Lahr et al. 2013a; Lahr et al. 2013c). Challenges will be linked to system complexities, system innovation, performance measurement and project characteristics.

3.1 System Design

Where RCTs only allow for studying a small set of interventions along the stroke pathway, simulation efficiencies enable elaborate testing of more complex systems. Observed opportunities for simulation use are in decision support for operational decision making – focusing at the overall chain optimization of existing treatment chains, and strategic decision making – concerning the choice of the regional stroke system network topology. The first opportunity is in line with a growing awareness in the field that a concerted effort of all parties involved in the hyper acute pathway results in best chain performance, i.e., highest treatment rates and best outcomes. Whereas earlier research efforts concentrated on optimizing the intra-hospital phase, relevance of including the pre-hospital phase in overall chain optimization has been widely acknowledged in recent years (Fassbender et al. 2013). The second opportunity relates to the observed need for a regional organization of care, for reasons of health economics and health quality, i.e., patient outcomes. For example, it has been shown how concentrating treatment of acute stroke in comprehensive stroke centres, instead of attending to patients in the nearest community hospital may benefit patients (Lahr, 2013; Lahr et al. 2013b). Clearly, such choices with respect to regional organization of
stroke care suggest a trade-off between transport delays in the pre-hospital phase vs. potentially better care and shorter lead times for the comprehensive stroke centre (Monks et al. 2014).

3.2 System Innovation

Whereas opportunities sketched in Section 3.1 essentially do not question pathway set-up in terms of the nature of care services, recent innovations do so. We observe two avenues of future change, concerning health technology employed in patient diagnosis, and new treatments. As far as technology is concerned we mention the following examples:

- The Point of Care device, allowing for a quick analysis of patient blood samples – as a replacement of classic blood testing as done in a lab (Rizos et al. 2009).
- Telemedicine, suggesting to exploit communication devices for, for example, consulting stroke expertise at a distance for use on scene or in a hospital lacking suchlike expertise, or pre-notifying patient arrivals at the hospital (Levine and Gorman, 1999).
- Mobile scanning technology available in the ambulance, allowing for CT-scanning at the patient scene. Note that this concept is exploited by so-called Mobile Stroke Units, i.e., dedicated ambulances that allow for tPA treatment at the patient scene (Wendt et al. 2015).

Note that all aforementioned examples stress reduction of delays along the stroke pathway.

Furthermore, we mention emerging new treatments for acute stroke. Recently, the use of endovascular thrombectomy (mechanical clot retrieval using a medical device) has been shown to improve patient outcomes (Berkhemer et al. 2014). Although its benefits are clear, it comes at a price, by setting specific demands concerning stroke expertise, and scanning technology (CTA-scanning). This may imply that in the near future it may only be offered by comprehensive stroke centres.

3.3 Performance Measurement

Interestingly, recent stroke studies have shown the linkage of treatment effectiveness in terms of patient outcomes and logistic performance, i.e., patient delay from stroke-onset to his/her treatment. Recent research shows how outcomes for acute stroke patient in terms of his/her chances of being treated (Lahr, 2013a), disabilities (Lees et al. 2010), and additional life years gained (Meretoja et al. 2014), may be estimated as a function of the patient treatment lead time. In turn, patient outcomes may be used to assess cost-effectiveness of alternative set-ups of the hyper acute stroke pathway.

3.4 Project Characteristics

Simulation studies on stroke systems tend to put high requirements on project team composition. Typically, the team hosts parties involved in the pathway, i.e., neurologists, EMS and GPs, health economists, health system engineers and simulation modellers. Profound insights on the way (interlinked) care services influence patient outcomes and the way chain logistics may be best organized, modelled and analysed are the key stones of a successful simulation study. Project efforts should be concentrated on creating a mutual understanding of stroke system set-up to facilitate its joint (re)engineering. Such an understanding is not guaranteed as parties, and disciplines represented in the team, may not be familiar with such co-operation.

4 CHALLENGES IN SIMULATION MODELLING METHODOLOGY

How to facilitate the simulation modeller and his team in setting up and doing the study? Here we explore methodological issues as they relate to the problem situation, and current means for addressing them.
4.1 Problem Situation

Optimization of the hyper acute stroke pathway is considered a world-wide issue. This is due to the number of patients involved (estimated at 15 million per year worldwide), and the severity of health consequences faced by those patients not being treated (as they arrive not within the therapeutic time window or the hospital is unable to treat them within 4.5 hours of onset) or who could have been treated earlier within the therapeutic time window – thereby improving their chances on favourable outcomes. In principle, care services along the hyper acute pathway required for effective tPA treatment are rather well-known. However, their implementation and facilitation (staffing, resources, and their co-ordinated use) may differ from country to country or even from region to region, due to the way local health infrastructures have been set-up. Nevertheless, world wide similarities in both choice of care services and system set-up suggest a high potential for simulation model re-use.

The hyper acute pathway is a complex system. Relevant parties in set-up and operation of the hyper acute stroke pathway include not only care givers, such as GPs, EMS, and neurologists, but also those who fund or regulate stroke care, such as insurance companies, (local) governments, and professional societies, and – last but not least – (representatives of) patients. This clarifies the need for problem structuring, to get a joint hold of system set-up and problems faced in optimizing it. In turn, this may also provide initial guidance on (conceptual) modelling – by revealing possibilities for model simplification.

For a simulation study to be successful (historical) data on patient lead times, diagnostic results, and their treatment and its outcomes are crucial, both to facilitate model set-up and its validation. Ideally, the build up of patient lead time along the hyper acute pathway since stroke onset can be explained by sufficient data on relevant care services. In turn, diagnostic results, like patient’s choice of first responder, EMS urgency level etc. will inform patient routings. Finally, data on patient treatment captures their likeliness of being treated, and their expected outcomes. Unfortunately, in many cases respective categories of data may (i) not be easily accessed as they are dispersed over separate parties involved in the hyper acute pathway or (ii) may not be available altogether, and have to be obtained at high costs.

Typically, as simulation is new for the field, aforementioned parties involved in optimizing stroke systems are not familiar with simulation. This implies a need to familiarize them with the tool, its application, and its potential for decision support on stroke system design.

4.2 Guidance Available

How should we address the challenges faced in the problem situation, as identified in Section 4.1? Many challenges link to the initial phases of a simulation study, i.e., conceptual modelling. So far, model coding, and its analysis, seem to be less of an issue. Note that this does not imply that they may not become a future issue.

Three basic approaches may be distinguished for guiding the analyst in specifying a conceptual model for simulation (Robinson, 2008a). Principles of modelling advocate the benefits of aiming for simple models through incremental modelling. Their application may, among others, entail the good use of metaphors, analogies, and similarities in model creation (Pidd, 1999). Methods of simplification work the other way around by suggesting a reduction of model scope and detail. Gains with respect to modelling efforts or computational efficiencies in doing experiments may be realized by, for example, combining model elements, leaving them out or adapting their attributes (Innis and Rexstad, 1983). Modelling frameworks suggest a step wise approach for detailing the conceptual model in terms of its elements, their attributes and their relationships. Typically, proposed steps are supported by guidelines, methods, and good practices.

The main differences among modelling frameworks concern their intended field of application, scope, and process support. Modelling frameworks developed so far tend to address rather broad fields of application, like operations systems (Robinson, 2008b), supply chains (Van der Zee and Van der Vorst, 2005), health systems (Kotiadis, 2007), and the military (Pace, 1999; Pace, 2000). Whereas some frameworks restrict scope to the specification of model content only (Arbez and Birta, 2011), other
frameworks include problem understanding, modelling objectives, experimental factors and model responses (Kotiadis, 2007; Robinson, 2008b). Chwif et al. (2013) and Kotiadis et al. (2014) address process support by suggesting formats for workshops, conceptual model documentation and the way model data are to be collected. For overviews of modelling frameworks, see Robinson (2008a), Karagoz and Demirors (2011), and Van der Zee et al. (2011).

Clearly, aforementioned approaches indicate relevant support for simulation conceptual modelling. However, they usually do not inform and guide the analyst in addressing modelling needs that are specific for a branch of industry or a domain of health care.

5 DISCUSSION

Challenges faced in simulation-based optimization of the hyper acute stroke pathway indicate that developing conceptual frameworks may be worthwhile, in an attempt to master system complexities, facilitate joint team work in solution finding, organize model data collection and make a further entrance to the field. Problem scale, and similarities in system set-up suggest that efforts put in defining such frameworks are relevant and feasible. Moreover, a domain specific framework would present an interesting – potentially viable – example of simulation model re-use. Re-use would build on the identification of generic model components, solution directions (model inputs), and relevant performance measures (model outputs), and ways of exploiting such means in process set-up and data collection. Clearly, defining such a framework is no easy job, already because it involves the involvement of many parties and disciplines.

6 CONCLUDING REMARKS

In this article we identify and study simulation modelling challenges faced in addressing the hyper acute stroke pathway. In doing so, we rely on our experiences in performing some of the pioneering studies for the field. We found how many challenges seem to address simulation conceptual modelling. The definition of a domain specific modelling framework for acute stroke care is advocated in an attempt to master system complexities, facilitate joint team work in solution finding, organize model data collection and make a further entrance to the field. Apart from being highly relevant for the field, given patient numbers, and potential for improving their health, the framework may present an interesting example of simulation model re-use. In our future work we will address the definition and use of such a framework.

ACKNOWLEDGMENTS

Research findings reported in this work build on authors’ involvement in several simulation projects on stroke system optimization (Monks et al. 2012; Pitt et al. 2012; Lahr et al. 2013a,c), and initial efforts in setting up simulation modelling frameworks for stroke care (Monks et al. 2015; Van der Zee et al. 2015).

This project was supported by the Netherlands Organization for Health Research and Development (ZonMw); Ref.: 80-82800-98-104.

One of the authors (Monks) is funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) Wessex. The views expressed in this publication are those of the author and not necessarily those of the National Health Service, the NIHR, or the Department of Health.
REFERENCES


Van der Zee, Monks, Lahr, Luijckx, and Buskens


AUTHOR BIOGRAPHIES

DURK-JOUKE VAN DER ZEE is associate professor of Operations at the Faculty of Economics and Business, University of Groningen, The Netherlands. His research interests include simulation methodology and applications, simulation & serious gaming, manufacturing planning & control, and health care logistics. He is a member of the INFORMS-SIM. His email address is d.j.van.der.zee@rug.nl, and his webpage is http://www.rug.nl/staff/d.j.van.der.zee.

THOMAS MONKS is a Senior Research Fellow in Operational Research in the Faculty of Health Sciences, University of Southampton where he leads NIHR CLAHRC Wessex’s methods research group. He holds a BSc in Computer Science and Mathematics, MSc in Operational Research and PhD in Simulation Modelling. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His specialty is simulation of healthcare systems of unscheduled and emergency care. He is co-chair of the UKs Simulation Workshop 2016 and co-chair of the UKs simulation special interest group. His webpage is http://www.southampton.ac.uk/healthsciences/about/staff/tm3y13.page

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 e-mail address is m.m.h.lahr@umcg.nl, and his website is http://www.rug.nl/staff/m.m.h.lahr/.

GERT-JAN LUIJCKX is a neurologist at the department of Neurology, University Medical Center Groningen. He is head of the Cerebrovascular Unit, and Principal Investigator Research School of Behavioural and Cognitive Neurosciences Groningen (BCN). Research topics addressed by him include clinical research prognostic factors acute ischemic stroke therapy, clinical research on neuroprotective determinants in acute stroke and clinical research on organization and implementation of acute stroke care. His email address is g.j.luijckx@umcg.nl.
Van der Zee, Monks, Lahr, Luijckx, and Buskens

ERIK BUSKENS is Professor of Health Technology Assessment and program director Healthy Ageing at University Medical Center Groningen, 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. His email address is e.buskens@umcg.nl.
OPENING THE BLACK BOX: COMBINING AGENT BASED SIMULATION AND REALISM IN INTERVENTION DEVELOPMENT

Dr. Sean Manzi
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
s.s.manzi@exeter.ac.uk

Dr. Daniel Chalk
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
d.chalk@exeter.ac.uk

Dr. Mark Pearson
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
mark.pearson@exeter.ac.uk

Dr. Jo Day
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
j.k.day@exeter.ac.uk

Prof. Ken Stein
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
ken.stein@exeter.ac.uk

Prof. Martin Pitt
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
m.pitt@exeter.ac.uk

Prof. Iain Lang
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon, EX6 7ST
i.lang@exeter.ac.uk

ABSTRACT

Interventions in healthcare seek to make a change to service provision or behaviour in order to improve health of individuals and populations. An intervention is usually designed and developed by experts using the available research findings in a given area. Prior to the piloting phase of intervention the effectiveness of the intervention design is not normally tested. Simulation methods provide a way to test and refine changes to a system before they are implemented in the real world.

The use of Agent Based Simulation (ABS) using a realist approach is presented as a method for the rapid development of complex behavioural simulations which can be used to inform the design and development of interventions. The complementarity between ABS and realism will be discussed and an example of this approach being used presented.

Keywords: Healthcare, agent based simulation, intervention development, realism
INTRODUCTION

Currently missing from the intervention development process is a platform on which to test and refine intervention effectiveness prior to any implementation in the real world. Intervention, process change and behaviour change are the labels given to the use of research findings to make some change to a system and/or process. These systems and processes are often individual, group, or physical (Argyris, 1970), although the differences between them are often blurred. An intervention is typically developed using the available research literature, piloted on a small scale and then implemented on a larger scale. Classic interventions were often described as ‘black box’ interventions. The change or new process would be implemented and observed, but the explanation of the mechanisms responsible for the change would be limited simply to a description of the intervention (Harachi et al., 1999, Tulsky et al., 1998).

The process of moving from using the literature to inform the intervention design to piloting and then large scale implementation challenges the researcher in terms of evidence synthesis, theory development, research design and evaluation. All of these stages and aspects of the intervention process must be considered while trying to produce and implement a useful and effective intervention (Craig et al., 2008).

The realist approach is being increasingly applied to intervention development (Pawson et al., 2005, Connelly et al., 2007). The realist approach is driven by theory and seeks to understand the interplay of context, mechanisms and outcomes in systems and processes (Chen and Rossi, 1980). Theory is developed from data and previous literature to produce an explanatory account of the system or process under investigation, Pearson et al. (2015) have used realist review techniques for understanding collaborative care for offender mental health prior to the development of an intervention. As demonstrated by Pearson et al. such explanatory accounts can be used to develop interventions, it is also possible to use realist methods to inform the implementation of interventions and evaluate them (Pawson et al., 2004, Pawson, 2013). This is the process of ‘opening the black box’ of an intervention so that the individual mechanisms responsible for producing a change in an intervention can be studied, understood and replicated.

Simulation methods can provide a test bed for intervention development which would support the design process. Agent Based Simulation (ABS) is a method by which the behaviour of individuals can be simulated within a given environment using sets of rules applied to the agents. The resulting behaviour and outcomes emerge from the interaction of the agents within the bounds of their rules and environment (Salamon, 2011).

This paper will argue that realist methods of inquiry can be used to inform simulation models of the proposed intervention, which in turn could improve the overall effectiveness of an intervention. Brief introductions to realist inquiry, ABS and interventions will be given in order to set out the scope of this approach. The complementarity between realist inquiry and ABS will then be examined and the similarities between the two approaches discussed. An example of the development process for a complex behavioural simulation which uses a realist approach will be given to demonstrate how this process can be practically implemented. The final discussion will look at the advantages and disadvantages of this approach to intervention development and highlight the technical difficulties that will need to be overcome.

REALISM

Modern realist inquiry has its roots in the works of Roy Bhaskar and Donald Campbell such as “A Realist Theory of Science” (Bhaskar, 1979) and “Methodology and Epistemology for Social Sciences: Selected Papers” (Campbell, 1988). When broken down to its most simple tenements, realism espouses the use of all types of data. Whether qualitative or quantitative, all data relating to a phenomenon aids in its description and our understanding of it (Campbell, 1988). The goal of realist enquiry is normally explanation, it seeks to go beyond description and provide insight into the cause. This is achieved by a focus on the context, mechanism and outcome of the phenomena under investigation (Pawson, 2013).

Through the utilisation of multiple data types detailed information can be collected on the context in which the phenomena occurs, mechanisms which are thought to be at work within that context and the outcomes that can be observed, enabling the development of explanatory theory which is what makes the
realist approach so powerful. Two modern applications of the realist approach which have become somewhat formalised in their application are realist synthesis and realist evaluation.

Realist synthesis and evaluation both seek to use evidence, whether from the literature or primary data sources, to develop explanatory accounts of the phenomena under investigation. When utilising one of these realist methods or a more general realist methodology, the realist approach brings a depth and breadth of evidence use which can inform the creation of an agent based simulation when developing an intervention which has a large behavioural component.

3 AGENT BASED SIMULATION

Agent based simulation (ABS) is a method which uses the individual as the main building block of the simulation. Typically it is the complexity that arises from the interaction of multiple individuals over time which is of interest, which represent the emergent properties of the system (Chen et al., 2007). This simulation method has been successfully applied in many different disciplines most prolifically ecology for the study of animal behaviour (Chalk, 2009, Bryson et al., 2007) and epidemiology for the study of disease outbreak and transmission (Eubank et al., 2004). In operational research (OR) this method is beginning to gain traction as its use for the study of behaviour in relation to behavioural OR becomes apparent (Escudero-Marin and Pidd, 2011), along with its usefulness in multimethod modelling to support discrete event simulation and system dynamics modelling (Flynn et al., 2014).

The agents in a simulation are programmed with a set of rules which often operate stochastically to mimic the variation seen in the real world. In this respect the behaviour and experiences of each agent differ over time. Agents can be programmed to learn and adapt based on their interaction with the environment and/or other agents. It is these properties of variation, individuality and learning which make ABS a suitable method for modelling social phenomena (Šalamon, 2011).

4 INTERVENTIONS

Interventions are programmes of change designed to improve the functioning of a system. They are commonly associated with health sciences where the system undergoing the change can vary from a ward or department in a hospital to national scale health process and from the individual person to whole populations. The main guidelines for the development and evaluation of complex interventions comes from the United Kingdom Medical Research Council (Craig et al., 2008) and as highlighted by (Anderson, 2008) the use of new methodological advances is encouraged especially those of realist synthesis and evaluation. In general there are four stages to an intervention: Development, feasibility and piloting, evaluation and implementation.

The development process of an intervention seeks to understand the existing literature around the process to be changed, a theory about how the change can be achieved is developed and the expected processes and outcomes mapped out (Richards and Rahm Hallberg, 2015). Realist synthesis can be used to aid in the process of developing interventions as it provides a relatively systematic approach to achieving each of these development steps (Pawson, 2006). The second stage is feasibility and piloting, this is where the intervention procedures are tested on a small scale. The feasibility and pilot studies carried out are evaluated to refine the intervention process and understand the mechanisms responsible for causing the changes in measured outcomes observed during the studies (Craig et al., 2008). It is at this stage that the use of ABS could be most effective.

The evaluation stage of the intervention process seeks to understand how the intervention was or was not producing the desired change. The implementation phase of an intervention is more long term than the feasibility and piloting phases. Once the intervention has been refined to the point that confidence is high in the positive value of the changes produced by it, the intervention can be rolled out on a larger scale.
5 COMPLIMENTARITY OF ABS AND THE REALIST APPROACH

There are a number of similarities between ABS and realism which make them suitable for use together.

5.1 Emergent behaviour

The outcomes of an ABS model are the result of the emergent behaviour produced by the behavior of individual agents and their interactions with their environment and other agents. This can be likened to Archer’s (1995) realist social theory of behaviour, in which behaviour is the result of the individuals’ intentions. The emergent behaviour of the system does not confirm the intentions of any one individual in the system but instead produces “Unavoidable, unplanned, self-generated, morphogenetic change” in behaviour (Archer, 1995). In this way, ABS produces the same naturally generated behaviour from system component interaction as seen in the real world through the realist lens.

In the realist approach, reliability is determined through testing and re-testing of hypotheses. When modelling, various scenarios can be explored using “what if” analysis and simulations are run many times to determine the variation within the system. It is often difficult to test and re-test hypotheses developed through realist synthesis and evaluation due to the complexity of the studies that would be required, the cost of running such studies and the resource use that would be required. The ethical issues associated with running such studies over and over again are also prohibitive. ABS provides an ethical and cost-effective environment in which to test the hypotheses arising from realist synthesis and evaluation used for intervention development and evaluation.

5.2 Theory driven approach

The realist approaches to evidence synthesis and evaluation are both driven by the use of existing theory and the development of theory (Chen and Rossi, 1983). For an ABS, the rules which are ascribed to the agents and the environment are derived from existing theory or the simulation is being used to test a theory about the functioning of a system. The need to at once develop and test theory is common to both ABS and realism for without it both approaches have no substance or use.

The role of context in the development and testing of theory in the realist approaches is also a central tenant of the realist philosophy. The context in which behaviour takes place can impact on the mechanisms responsible for producing the behaviour (Elster, 2007). In an ABS the environment in which the agents interact impact the behaviour of the agents. Therefore, the environment of an ABS can be used to model the context of real world behavior which is central to the realist approach.

5.3 Use of data

The realist approach is one of the few philosophical schools of thought that does not place restrictions of the type of data that should be used under its banner (Campbell, 1988). Many methodological approaches require the use of either qualitative data or quantitative data but do not support the integration of both data types (Bryman, 2012). To enable the researcher to capture data about the context, mechanisms and outcomes in the situation under investigation both qualitative and quantitative data are necessary and valid. Qualitative data adds description, detail and insight to the study while quantitative data adds specificity, objectivity and comparability (Olsen, 2010).

ABS models can be informed by both quantitative and qualitative data. Behaviour in ABS models is represented using a series of rules, typically following an “if-then” format. Qualitative data can be used to inform the development of such rules (unlike in purely mathematical models) allowing behavioural aspects to be captured even where quantitative data is lacking.
6 AN EXAMPLE OF BEHAVIOURAL SIMULATION DEVELOPMENT – POLYPHARMACY MEDICATION ADHERENCE

When people are taking more than one medication on a regular basis this is referred to as polypharmacy. An ABS tool was developed as part of a wider programme of work which is investigating peoples adherence to multiple medications. When taking multiple medications people have been observed to change their prescribed drug regimen, but this process is not well understood. Understanding how people change and self-optimize their medication under conditions of polypharmacy is a topic of interest in relation to person-centred healthcare and medication prescribing practice (Anthierens et al., 2010).

Many different factors influence whether or not somebody will take their medication as prescribed, and this represents a complex behavioural phenomenon where a change in behaviour and practice is required. The number of medications, how often they have to be taken and the complexity of taking them all add to the treatment burden the patient has to deal with. In the case of polypharmacy, one set of medications for a condition may also adversely interact with the medications for another condition. Both high treatment burden and adverse drug interactions are more likely to result in divergence from prescribed treatment regimens.

The ABS for this project was developed using a realist inspired approach which is outlined in Figure 1 and described in detail below.

The first stage of the project was to determine the research question, this was achieved through collaboration between experts in the area of polypharmacy research and the simulation team. By working together the scope of the project could be constrained to make it practical for simulation while still being useful for the intervention development process. The research question was: “Can ABS be used as a means of determining the optimal set of medications to prescribe to adults with type 2 diabetes (T2D) and asthma, such that their treatment burden is minimised, the clinical benefit of the medication maximised and any wasted cost from unused medication or additional hospital visits minimised?”

To narrow the scope further for a preliminary exploratory investigation, the team focused on a population of interest of adults with both Type 2 Diabetes and asthma. The key outcome measures were the percentage of total medications prescribed being taken by the patient and their average state of health, as if a patient takes fewer medications than prescribed there are potential wastages involved, along with potential safety issues. The population of interest was chosen because insulin medications for diabetes are known to produce negative side effects when taken with some asthma medications. In addition, diabetes and asthma are common co-morbidities found within the UK population making this a useful study population. The wider project is seeking to understand polypharmacy from a patient perspective with the idea that more effective medication use can be achieved by understanding how the patient uses their medication, which in turn could help to inform prescribing practice.

Once the research question and study population were determined the next stage of the process was to map the relevant potential proximal and distal factors which could influence someone not to take their medication as prescribed. To draw comparison with the realist approach, this is the stage at which the theoretical framework, on which the simulation would be based, starts to be developed. The mapping process included engaging with current patients who have multiple comorbidities, experts in the field of polypharmacy, pharmacists and a General Practitioner. Through extensive discussion, an influence map was developed (see snapshot in Figure 1).

The influence map enabled a focused search of the literature to find both quantitative and qualitative data that could be used to represent and parameterise such relationships within the mathematical modelling environment of an Agent Based Simulation. Relevant journal articles were identified and entered into a table of the relationships with a brief summary of the useful findings. While this was a labour-intensive process with over 150 relationships to find evidence for, by creating the influence map first the search of the literature could be focused only on those relationships which had been identified.

The next stage of the simulation development process was to extract the data and develop rules based on that data to parameterise the simulation. Both qualitative and quantitative findings were included in the literature search; the qualitative data was used primarily to develop rules for the simulation while
quantitative data was used to assign probability distributions and weightings within the rules from the qualitative data. The use of a realist approach to this project allowed multiple research methods (focus groups and evidence synthesis) and data types (qualitative and quantitative) to be used. It also enabled these data to be interpreted and used in the context of each other to produce a more complete picture of the phenomena under investigation. The process of mapping the relationships between the factors influencing adherence gave a basic structure to the programmatic rules for the agents. This process made the inclusion of the evidence informed rules easier as they could be fitted into the structure of the relationships.

Figure 1 Snapshot of the polypharmacy influence map

Once all of the information required to build the simulation had been collated, the build process could begin. The simulation was built using Anylogic 6.71 multimethod simulation software. The model was structured such that each general factor that directly influenced adherence resolved to a boolean proposed action – to either adhere or not to adhere. In addition, a modified Linear Operator Learning Rule was implemented to model the impact of patient wellness and perceived wellness on the self-optimisation process. Each patient maintains a perception of how well they feel with their current medication, as well as a memory or estimate of how well they would feel for each other combination of medications. The perception of their wellness with their current medication combination is updated each simulated day based on how well they actually are, their sensitivity to the effects of the medication, and the importance they
place on how they feel now compared with how they felt before. If their perception of how well they are with their current medication falls a user-defined percentage below that of an alternative medication combination, the proposed action will be for them to switch to that medication combination. The action that a patient undertakes each day (to stick with their current medication or switch to an alternative combination) is selected at random, with different weightings for the proposed action derived from the patients’ wellness, and the proposed actions derived from the other more general factors.

Currently, the simulation is structured as a tool that allows users to explore the impact of different prescribed medication combinations and different population demographics on adherence to prescribed medication and levels of health, with treatment burden of a medication combination judged by the user. In time, it is hoped that this tool can be used to aid the development of an intervention to improve prescribing practice for those with Type 2 Diabetes and Asthma.

This simulation development process was facilitated by the use of information from a range of sources including; expert patients, expert researchers, primary research and secondary/review literature, to develop a theory of polypharmacy medication adherence behaviour on which to base the simulation. Being able to bring together qualitative and quantitative data not just in the theory development stage of the project but also in the creation of the rules and parameters for the model enabled a more complete picture of the relevant behavioural systems to be understood and implemented in the simulation.

The relative speed with which this type of simulation can be developed would likely not interfere with the normal intervention development process and provides a way to more completely study complex behavioural phenomena which maybe disparate in the literature, but can be studied in conjunction with one another in the simulated virtual environment.

7 DISCUSSION
The aim of this paper was to make the argument for the use of a realist approach to the design of ABS models to support the intervention development process. The level of complementarity between the realist approach and ABS would enable each approach to inform the other leading to more rigorous, reliable and valid theory. Both approaches acknowledge social behaviour to be greater than the sum of its parts, as it is the emergent quality of the behaviour that both approaches seek to understand and capture in the theory from which they are developed and seek to develop.

The data types required by both approaches and their use of this data to find explanatory sequences of behaviour are also unified. Realist methods of enquiry can provide the rules and theory for developing a model of phenomena of interest while ABS provides a technical platform in which to test that theory and hypotheses. Being able to study complex and even ethically difficult situations in detail beyond that which could realistically be captured in the real world provides researchers and particularly social scientists with a new and potentially useful approach to scientific investigation.

With the integration of realist approaches and ABS comes the opportunity to apply their complementarity to an important area of research; interventions. As described in Section 6, ABS modelling informed by a realist approach can be used to support and improve intervention development. Financial support for research is becoming increasingly scarce and competition increasing, and it would be timely to support intervention programmes with ABS to reduce the pressure on pilot studies which could reduce costs and potentially improve the effectiveness of the intervention and increase its subsequent impact.

There are challenges to be met when using ABS and realist approaches in support of intervention development. As with all interdisciplinary research, the level of understanding between ABS practitioners, social scientists and healthcare researchers about the language being used and the capabilities of each approach will have to be addressed. This is something that takes time and communication, but being aware of this issue from the start of a project will aid in overcoming any difficulties (Wear, 1999). A limitation of ABS as with all simulation and modelling approaches is that models and simulations are simplifications of reality providing only an estimate of what might occur in reality (Law, 2006). However, as long as the results are communicated and understood in this context, the outputs of models can still be very useful in providing some evidence to help inform a decision.
A final challenge will be to ensure that ABS and realist approaches are truly applicable to the intervention being planned. Successful application of this approach will come from ensuring that ABS is used in situations where a key component of the proposed intervention is individual-level behaviour, and that the realist approach is correctly applied to produce sufficient data on which to build a model.

The integration of ABS and possibly other modelling and simulation approaches along with realist approaches in the support of intervention design has huge potential. These methods work best when applied to real world problems, and it is our hope that impactful applications of these methods will help to prove the effectiveness of this combined methodology.

ACKNOWLEDGMENTS

This research was funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

The polypharmacy project work was led by Dr Daniel Chalk and carried out in collaboration with Prof Nicky Britten and Dr Janet Heaton (NIHR CLAHRC for the South West Peninsula (PenCLAHRC)), Bettina Kluttgens (South West Academic Health Science Network (SW AHSN)), Rati Magura (Royal Devon and Exeter NHS Foundation Trust) and the Peninsula Patient and Public Involvement Group (PenPIG).

Thank you to Prof Rob Anderson for his comments, support, insight and advice on this topic and the writing of this paper.

REFERENCES


**AUTHOR BIOGRAPHIES**

**SEAN MANZI** received a BSc (Hons) Psychology from Plymouth University in 2011. He completed his PhD at the same university in 2014. He is currently an Associate Research Fellow with the Operational Research team (PenCHORD) of the South West Peninsula Collaboration for Leadership in Applied Health Research and Care (PenCLAHRC). Dr Manzi is using his multidisciplinary research experience to improve healthcare provision and develop new research methods to support this process. [http://clahrc-peninsula.nihr.ac.uk/staff/sean-manzi](http://clahrc-peninsula.nihr.ac.uk/staff/sean-manzi)

**DANIEL CHALK** achieved a BSc Computer Science at Exeter University in 2005. He completed his PhD in Biosciences at the same university in 2010. He is currently a Research Fellow within the Operational Research team (PenCHORD) of the South West Peninsula Collaboration for Leadership in Applied Health Research and Care (PenCLAHRC), where he applies modelling, simulation and analysis techniques to help inform decision making in the local NHS, and has a particular interest in modelling behavioural aspects within health and social care systems.
MARK PEARSON attained his PhD from Plymouth University in 2008 and is currently Senior Research Fellow in Implementation Science with the NIHR CLAHRC South West Peninsula (PenCLAHRC) at the University of Exeter Medical School. Dr Pearson’s research focuses on the interplay of social, psychological and organisational factors that occur when evidence-informed changes in policy and practice are implemented in health services, and in particular the role that realist research can play in improving understanding of these processes.

JO DAY received a BSc (Hons) Social Psychology from the University of Kent, a MSc in Applied Criminological Psychology from Birkbeck College, University of London and completed her PhD in Sport and Health Sciences at the University of Exeter in 2012. She is a registered and chartered psychologist and is currently a researcher with the NIHR PenCLAHRC. She contributes to applied research projects that seek to develop in-depth insights and understanding of how to implement, disseminate and share research evidence and knowledge in health and social care settings with the aim of improving patient and public outcomes.

KEN STEIN qualified in medicine from the University of Bristol in 1987 and later specialised in general practice and Public Health. His higher degree (MD) is in health economics and health technology assessment (Bristol, 2007). Ken is currently Professor in Public Health at the University of Exeter Medical School and deputy director of the Peninsula Collaboration in Health Research and Care (PenCLAHRC) and oversees the work of PenCHORD for the collaboration.

IAIN LANG is a Senior Lecturer based in the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care for the South West Peninsula (PenCLAHRC) at the University of Exeter Medical School and is also an NIHR Knowledge Mobilisation Research Fellow. He is interested in knowledge use in both applied and research settings.

MARTIN PITT is currently Associate Professor of Healthcare Modelling and Simulation at Exeter University Medical School and Director of PenCHORD - The Peninsula Collaboration for Health Operational Research and Development. He has a long-standing and wide ranging experience in healthcare modelling ranging from economic modelling in Health Technology Assessment to Discrete Event Simulation for operational models for service re-design. He has a specific interest in promoting the effective implementation of models in healthcare and the use of visualisation techniques to improve the communication of outputs to key user groups.
A FACILITATION WORKSHOP FOR THE IMPLEMENTATION STAGE: A CASE STUDY IN HEALTH CARE

Dr. Kathy Kotiadis  
University of Kent  
Canterbury, CT2 7NZ, UK  
K.Kotiadis@kent.ac.uk

Dr. Antuela A. Tako  
Loughborough University  
Leicestershire, LE11 3TU, UK  
A.Takou@lboro.ac.uk

ABSTRACT

Research on facilitation in discrete event simulation (DES) is gathering pace but there is still a need to put forward real examples to explain the process to newcomers. Most of the research has focussed on facilitation in the initial stages of the simulation modelling process. In this paper we focus on one of the postmodel coding stages. More specifically we focus on the implementation stage, the final stage in the modelling process. The primary contributions of this paper are the description of the process followed and the introduction of tools that can be used during this stage to support workshop activities. A real case study is provided describing the sequence of the interactions undertaken in the workshop. Extracts from the transcripts are also included, with the view to bringing evidence of the stakeholders’ involvement and their mood during the workshop. The paper concludes with a discussion on the process followed and the importance of using tools in this stage.

Keywords: Facilitation, Implementation, Methodology, Workshops, Health Care

1 INTRODUCTION

This paper complements existing work on facilitated DES (Adamides and Karacapilides, 2006; den Hengst et al, 2007; Barjis 2011; Robinson et al 2012; Robinson et al 2014; Tako and Kotiadis, 2015) by describing a post model coding facilitation workshop and the tools used to support it.

A little over 10 years ago Taket (2002) noted that facilitation was emerging as a term and listed a number of existing books (e.g. Taket and White (2000)) and articles (e.g. Huxham (1991); Phillips and Phillips (1993); White and Taket (1994)), none of which included facilitated discrete event simulation, although Group Model Building (facilitated System Dynamics) had already established itself (Vennix 1996). The term facilitation has only been adopted in the last few years by the DES community (van der Zee 2007; 2011; Tako et al 2010; Barjis 2011; Tako and Kotiadis 2012a,b, 2015; Kotiadis et al 2014; Robinson et al 2014) but nevertheless ahead of other hard OR approaches (such as linear programming, combinatorial optimisation, etc.) that appear to have not yet explored the opportunities that facilitation has to offer.

Facilitation in discrete event simulation (DES) offers an alternative mode of engagement compared to the traditional expert mode of undertaking DES where the focus is on an individual client rather than on a group of stakeholders. The expert mode encourages the operational researcher(s) to use the simulation model to undertake an objective analysis of the client’s problem and then recommend optimal or quasi-optimal solutions (Franco and Montibeller 2010). In facilitated DES the aim is for the operational researcher(s) to use the model in a workshop(s) with several stakeholders to enable a subjective analysis, where the solutions are viewed as feasible and desirable, whilst taking into account environmental constraints.
Facilitated DES authors are in unison over the need for more real examples and/or methodological developments (Tako et al 2010; Tako and Kotiadis 2012a,b, 2014; Kotiadis et al 2014; Robinson et al 2012; 2014, Adamides and Karacapilides, 2006; den Hengst et al, 2007; van der Zee 2007;2011; Barjis 2011). We use a real case study in health care to describe the implementation workshop process and put forward tools to support the workshop. Our implementation workshop is about how the modellers engage with a group of stakeholders to narrow the solution space so feasible and desirable action can be taken. We have used extracts from transcripts to demonstrate the mood and interaction in such a workshop. Hence our contribution is twofold. Firstly, we contribute towards creating a larger pool of real examples of facilitated DES. Secondly, we contribute towards developing a methodology of facilitated DES by focusing on the process followed and tools used for the implementation stage. Indeed this paper complements our previous work where we have concentrated on premodel coding stages (Kotiadis et al, 2014) or on the overall PartiSim framework for facilitation (Tako and Kotiadis, 2015). To give the reader a an overview of the PartiSim framework, it is formed of six key stages: Initiate simulation study, Define system, Specify conceptual model, Model Coding, Experiment with model and Implement Findings (Tako and Kotiadis, 2015; Kotiadis and Tako, 2010).

The paper is structured into five further sections. The following section explores some of the existing literature relevant to post model coding facilitation in DES. Next we describe the case study followed by a description of the process followed in the implementation workshop in section four. A discussion follows on the contribution of facilitation to the stage of implementation in DES and the need for dedicated post model coding tools such as the ones we put forward. Section six concludes the paper.

2 POST MODEL CODING FACILITATION IN DES

In this section we will focus initially on the studies contributing to facilitation in the post-model coding stages and then we will explore the use of tools in these stages. Den Hengst et al (2007) reports on a collaborative simulation study for a Dutch airline carrier, that combines group support with simulation modelling. Post model coding workshops involved the management team and the equivalent implementation stage considered alternatives and choosing a direction for the future. Holm et al (2013) describe a study which involves developing a DES model of a surgical hospital unit within the SSM’s seven stage process where postmodel coding facilitation focused on the desirability and feasibility of changes. Despite embedding workshops with stakeholders, the study does not consider in depth the aspect of facilitation, as it focuses primarily on developing a multi-paradigm multimethodology for combining soft and hard methods. Robinson et al. (2014) provides empirical evidence of carrying out facilitated modelling with a group of healthcare professionals at an outpatients eye clinic, as part of a lean improvement workshop, called SimLean. The post model coding workshop involved presenting a simple model developed beforehand to discuss lean improvements. The authors put forward the steps followed during the process, however they focus mainly on developing simple models that can be used to help understanding.

We next focus our review on the use of tools in facilitation, particularly to support post-model coding facilitation in DES. Robinson et al (2014), also echoed by Barjis (2011), identify the need for premodel coding tools to assist the process of facilitation. Similarly, den Hengst et al (2007) suggest the need for developing aids and tools that can support the facilitation process and stakeholder engagement in the workshops. Kotiadis et al (2014) have put forward tools that aid the pre-model stages, that is conceptual modelling. These tools are not suitable for post-model coding stages because the outputs differ between the pre- and post- model coding stages. PartiSim tools have been designed to fit the outputs of the intended stage. Hence we will distinguish PartiSim tools from general facilitation tools used to record and enable general debate that could be used in theory at any stage. For example, Group Support Systems (GSS) used by den Hengst et al (2007) offer anonymity, parallel input and group memory. GSS are said to support five different patterns of collaboration: divergence (e.g. brainstorming), convergence (clarify and reduce), evaluation, organization, and building consensus (Briggs et al, 2006). For more information on GSS we

Unlike the pre-model coding stages (Kotiadis et al, 2014), the postmodel coding stages in PartiSim have not been explained in detail so it is not clear how they compare to the stages that other facilitated studies have put forward. Furthermore, no tools have been put forward that can specifically support the postmodel coding stage. Barjis (2011) makes the point that facilitated DES would benefit from the development of tools to support the whole process.

3 CASE STUDY

The case study context is the treatment of patients with morbid obesity for an obesity service that provides lifestyle, pharmacotherapy and surgery treatment options for the UK’s National Health Service. For confidentiality reasons we will not refer to the centre by its name. At the time of this research (early 2010) the centre was just about meeting the demand. However, in the long term, they recognised that they would be running the risk of building long waiting lists, with patients experiencing long waits and risk breaching government directives, such as the 18-week target (patient maximum wait time from referral to first treatment) set by the Department of Health in the UK (Department of Health, 2004).

A stakeholder group of around 12 had accepted the invitation to participate in the implementation workshop. The same group stakeholders had participated in previous workshops focussed on other stages of the modelling process. This workshop was organised in a 2 hour slot. The stakeholder group consisted of a wide representation of different parts of the obesity care system, including healthcare professionals (surgeons, doctors and nurses) of different seniority from a range of specialties such as general surgery, chemical biochemistry, anaesthetics and endocrinology as well as members of the senior management team. The modelling team at the workshop included three analysts, who took on either the role of the facilitator or recorder (note keeper) during the workshop.

It should be noted that in our interactions with the stakeholders as part of the pre-model coding stages of the study (Kotiadis et al, 2014), it was agreed that the aim of the study was to identify the impact that an increase in resources (surgeons and physicians) and/or a reduction of patient referrals (lower referral rates) into the service, would have on the 18-week target. The agreed simulation study objectives were:

Objective 1: To explore reducing the waiting list for the surgical clinics, pharmacotherapy clinic and patient education by incrementally increasing the number of surgeons and physicians to a maximum of three and two respectively as well as reducing first time referrals.
Objective 2: To explore reducing the percentage of patients that breach the 18-week target by incrementally increasing the number of surgeons and physicians to a maximum of three and two respectively.
Objective 3: To explore reducing the percentage of patients that breach the 18-week target by managing demand through a reduction in patient referral rates into the service.

The readers should note that in this paper we do not provide a lot of detail of the solution space as our aim is to describe the workshop experience and explain the tools used to support facilitation. Instead we provide some snippets of the workshop transcripts to provide readers with a sense of the interactions. However further information on the actual scenarios and the associated findings can be found in Tako et al (2014). The intervention took place over a period of about 6 months with the implementation workshop scheduled in the final month. We next describe the postmodel coding process followed.

4 IMPLEMENTATION WORKSHOP

The workshop was structured around three key aspects: 1) review of learning and changes implemented (during the study), 2) risks analysis and feasibility of change, 3) Agree action trail.
4.1 Review of learning and changes

The workshop started with a reminder about the aims of the study, refreshing stakeholders’ memory on what had been already accomplished. Robinson (2014) suggests that one of the main benefits of DES studies comes from the learning generated during the modeling process; yet the modelers/facilitator may need to intervene in creating awareness of the learning achieved. If the clients understand their problem situation and are given support in developing actions to address this, then they are more likely to implement the proposed solutions.

The facilitator referred briefly to the problem statement, the objectives, ran the simulation model to remind the stakeholders of it and provided them with a table with the feasible and desirable scenarios, including the experimental factors (inputs) and the final model results. The briefing was only aimed as a warm-up to the workshop. A report had been compiled following an earlier workshop on experimentation, which had already been circulated to the stakeholders and most of them had already read it. Nevertheless, comments and extensive discussion took place during the workshop because the report described only the scenarios and model results but did not explain the behaviour of the performance measures. The stakeholders delved deep into the reasons for which these results were achieved. For example:

*Stakeholder A*: So how can this happen? I’m trying to understand. {referring to a result}  
*Stakeholder B*: How does that work? I don’t understand how that works {referring to a result}.  
*Project Champion*: It works by the number of referrals, when you cut your referrals down. So if you outsource [number purposely deleted] but you keep your referrals coming in at the same rate, you don’t change that.  
*Facilitator*: I completely understand what you’re saying. It’s because of how you introduce the resources in the second scenario. If you introduce the physician earlier by the amount that we’ve introduced it, what the physician does is they push loads of patients forward.  
*Stakeholder A*: Oh I see.

Churchman and Schainblatt (1965, p73) in their seminal paper emphasise the need for such an understanding to be reached ‘For the proper communication to take place, the manager must understand what the scientist is trying to do and why he does what he does. Here the problem of implementation is the education of the manager. After a successful implementation, the manager himself becomes “more of a scientist”’. Some of the stakeholder discussions about the results included assertions about their expectations and beliefs about the system. By getting the stakeholders to articulate these, the modelling team gauged the impact of the study and the learning gained by the participants. Some examples include:

*[The numbers are]* Slightly worse than we expected. {[workshop participant]}  
*I was very surprised that we actually, if we outsourced ... [number purposely deleted] patients, it actually makes it worse. I think that was a real solution. I thought actually taking ... [number purposely deleted] people out of the system would actually make the system better. Because it’s ... my starting point has always been if we can clear the backlog and we can get the system in balance, that’s the solution. {[project champion]}*

Next the stakeholders were prompted by the facilitator to report on any changes that might have already occurred in the system since the study started. From our communications with the stakeholders during the three month gap between the previous workshop and this one, we were made aware that additional surgical slots, equivalent to the addition of one surgeon, had been already introduced into the system. The model results demonstrated that this isolated implementation was found to be a poor decision because it led to bottlenecks when not combined with increases in other types of resources. The participants explained that

---

1 The project champion is a stakeholder that champions the process and has more involvement in the project compared to other stakeholders.
it was a decision taken prior to the study and that the study would in fact influence the next decision they made. However prior to the study we were told that there were no imminent changes to the resources. Learning normally occurs gradually and the subjects themselves may not be aware of it happening as it changes the system of beliefs and attitudes, used to make judgements (Ajzen, 1991). In hindsight, a more indirect way of identifying change of attitude would have been more appropriate, such as administering before and after questionnaires, recording stakeholders’ plans or any additions in their knowledge/learning as the study went on (Monks et al 2014). When discussing the impact of the change already made to the system and in light of the scenarios previously shown, stakeholders commented that this was a pretty much quick fix of the waiting list for surgery:

... we can now pretty much meet our steady state capacity, ... so that is enabling us to, based on monthly referrals, keep the backlog as it is rather than grow larger. But we still have obviously this big balloon after ... [clinic name deleted for confidentiality purposes] at the moment (meaning a large waiting list further up in the system). [workshop participant]

Stakeholders recognised that there was a large waiting list further up in the pathway with patients waiting to be referred for surgery. The discussion that followed indicated that stakeholders understood that a more sustainable change needed to be implemented. This was clearly a learning point from the stakeholders’ point of view during the review session, which provided evidence of the impact of such a decision (adding more surgeons) on the rest of the system.

4.2 Risks and feasibility of change

This part of the workshop focused on the most desirable scenario, and aimed to explore the factors that may hinder implementing the changes required. For example earlier in the workshop physical space was identified as an issue for implementation of any 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>

The aim here is to narrow the solution to ideally one scenario that could be implemented. Out of the scenarios explored the third scenario was the best performing scenario for most performance measures. This was also the most preferred scenario by all stakeholders. The facilitator asked the stakeholders to consider how this scenario could be put in place and the inhibiting factors were discussed. It is recognised that factors such as psychological perceptions may hinder the stakeholders from taking action (Ajzen 1991). Ajzen (1991) maintains that communication that attacks believed constraints can produce changes in attitude towards a behaviour. Hence debate and discussion is considered important to challenge attitudes and perceptions towards change. To add to this line of argument for debate leading to implementation, Schultz et al. (1987) explain that debate and the unveiling of the diversity of opinions is likely to change future management strategy because discussions help to change management’s own values, personal beliefs, and attitudes. Debate and discussion is important to challenge attitudes and perceptions towards a change, and communication and involvement can provide further support for change.

The facilitators used the “Feasibility and risks scale” Tool (Figure 1) to identify the reasons for which this scenario was feasible and the reasons for which it was not feasible. The outcome was to weigh up its feasibility. The tool was designed prior to the workshop by the authors and the facilitator followed the process to construct it with the stakeholders. 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 slopping line, dipping in this case on the not feasible side of the scale. This particular scenario 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. As a result of this analysis, it was accepted that this scenario was not feasible mainly due to timing issues. A number of staggered scenarios that would ultimately lead to the same resources in the longer term were subsequently discussed using the same tool and process.

4.3 Agree Action Trail

The participants next concentrated on other indirect resourcing issues such as: improving referrals so they are appropriate; introducing dedicated space with additional surgical theatres; outsourcing some part of their work elsewhere (to a different hospital) and the financial impact of such a decision. The stakeholders were handed the following form to record their thinking and actions for change (Figure 2). The forms were theirs to keep and take forward in a move to hand over implementation and action back to the stakeholders.

The workshop came to an end with the facilitator asking the participants to comment on the modelling approach and process as well as fill in a brief survey. An extract from the conversation follows:

**Project Champion:** We’ve had good involvement.

**Facilitator:** It doesn’t have to be good! [the extract follows from a series of positive comments so the facilitator is suggesting here that other less positive views can be expressed]

<Laughter>

**Stakeholder A:** I agree with ... [name removed] in that I think we knew there was a problem, we knew where the problems were. What you’ve done is you’ve actually put it in black and white and we can actually see that it is clear, it’s there, and that we need to do something about it. But I think what it’s shown is every time we correct something, actually the problems work in.

**Stakeholder B:** It’s the quantification and the clarification of the problem, quantified and clear. This I would say will increase, you can put numbers, it’s quite an important thing to plan the resources... this process is proper process, this is the standard, proper process. You have a pathway and then you have a model and you validate the model in the workshop and see where the model ends up, so this process is a good process. There’s a good process there....
As an immediate outcome of this study the Trust decided to add more surgeons to the service instead of adding physicians alone. Following the workshops, the Centre involved and the Primary Care Trust, engaged into discussions about changing the local eligibility criteria for this type of surgery, which eventually led to a reduction in the number of referrals to the centre. A decision to build a new operating theatre was also made as the management team realised that additional capacity was needed in order to achieve aspired service levels and operation volumes.

5 DISCUSSION
Implementation of findings in PartiSim is quite different to expert mode DES where the aim is to go as far as documenting, presenting and using the results (Law 2007). Using the results in expert mode is explained as ‘Results are used in decision-making process if they are both valid and credible’(Law 2007, p70). Elsewhere it is explained as implementing the findings; implementing the model and implementation as learning (Robinson 2014). In PartiSim, this stage is undertaken in a workshop setting, where a learning process is undertaken. The process aims to move the stakeholders away from the model and its findings towards gaining an understanding of the present and future implications of each scenario, so that both feasible and desirable solutions can be identified, in order to enable action to be taken.

We will now explore what facilitation aims to bring to the stage of implementation. To do this we use Franco and Montibeller’s (2010) four underlying assumptions of facilitated OR. The first assumption is that problems are socially constructed entities rather than real entities. The workshop venue is able to support the different opinions of the participating group through discussion or brainstorming. The second assumption is that subjectivity is unavoidable and that the facilitator should try to externalise these and represent them in a model. The workshops focus is on the opinions of the stakeholders about the problem and system rather than on our (‘modellers’) view or opinions of the real system. The third assumption is that clients want satisficing solutions rather than optimal solutions. Indeed during PartiSim’s implementation workshop the stakeholders explore the feasibility of the preferred scenario rather than simply focusing on the improved performance measures. The final assumption is that participation increases the commitment for implementation rather than believing that implementation of scientifically-based analysis is straightforward. A dedicated workshop is aimed at discussing implementation which intends to get the stakeholders to move away from the simulation model towards identifying an action trail for change. Separating out the workshop, among other things, is meant to demonstrate to the stakeholders the importance of implementation and generating outcomes.

Other facilitated DES approaches use Group Support Systems (GSS) (den Hengst et al, 2007) that could be used at any stage of the modelling process. PartiSim post-modelling tools like GSS are aimed at
supporting patterns of collaboration such as divergence (e.g. brainstorming), convergence (clarify and reduce), evaluation, organization, and building consensus that will ultimately lead to each workshop output. However these do not always offer anonymity or parallel input because that would require equipment and/or software for each participating member, which may not be available. The idea in PartiSim is that tools should be accessible. They are designed so that facilitators and participants are not reliant on expensive equipment, to ensure that the facilitative approach is widely adopted.

6 CONCLUSION

PartiSim (Kotiadis and Tako 2010; Tako and Kotiadis, 2015) is formed of six key stages. The key stages include: Initiate simulation study; Define system; Specify conceptual model; Model Coding; Experiment with model and Implement Findings. The least explored aspect of PartiSim are the post-model coding stages, which include experimentation, i.e. searching the solution space, and implementation, i.e. establishing action to be taken. This paper focused on the latter of these two stages. Extracts from the workshop were provided to illustrate how a workshop was structured around three key aspects: 1) review of learning and changes already implemented (during the study), 2) risks analysis and feasibility of change, 3) Agree action trail. The extracts have captured the mood of the workshop and the engagement of the stakeholders. Additionally we proposed two simple paper based workshop tools dedicated to this stage of implementation. The aim of this paper is ultimately to encourage debate and more research to improve the implementation of DES findings through workshop participation and the development of dedicated tools for this stage.

ACKNOWLEDGMENTS

This study was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/E045871/1. Kathy Kotiadis additionally acknowledges the support and sponsorship of the Daphne Jackson Trust Fellowship and the University of Kent in the dissemination of this study.

REFERENCES


Kotiadis and Tako

Kotiadis and Tako


AUTHOR BIOGRAPHIES

KATHY KOTIADIS is a Daphne Jackson fellow at the University of Kent. She is a former assistant professor at the University of Warwick and has a PhD from the University of Kent. Her research interests include the process of modelling, conceptual modelling, discrete-event simulation, Soft Systems Methodology (SSM), facilitated approaches in healthcare and developing Self Adaptive Discrete-Event Simulation. Her publications can be found on https://www.researchgate.net/profile/Kathy_Kotiadis

ANTUELA A. TAKO is a Lecturer 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 and conceptual modelling. She is an Associate Fellow of the Operational Research Society (AFORS). Home page: www.lboro.ac.uk/departments/sbe/staff/profiles/takoantuela/tako-antuelaanthi.html.
AN OBJECT-FUNCTIONAL MODELLING PLATFORM TO ENABLE SEMANTIC WEB ANALYSIS OF CAMPAIGN MODELS

Mr. Gareth D. Toomey
Defence, Science Technology Laboratory (DSTL)
DSTL Portsdown West, Portsdown Hill Road, Fareham
gdtoomey@dstl.gov.uk

ABSTRACT

Historically, Dstl has produced large models in order to represent and analyse the defence domain. These monolithic modelling structures have become difficult to adapt and maintain in an ever-changing defence landscape. The Generic Aggregator Model Valuator (GAMOV), is an approach to model development that has been produced to allow models to be rapidly constructed from a set of reusable libraries. GAMOV’s code-base embraces the principles of an emerging design paradigm called Object-Functional. This paradigm facilitates the tracking of data flows around the subsystems of a simulation model in a new way. Research is currently underway to exploit this potential through developing an approach called the Semantic Web Examiner of Emergent Phenomena (SWEEP), which we believe will enable us to systematically understand emergence within our modelling. The first-stages of this research are taking place over the coming year. It is hoped that at a first-pass these initial experiments shall enable us to develop the SWEEP approach so that model behaviours can be categorized in order to better target our analytical resources.

Keywords: Semantic Web, Ontology, Object-Functional, REST, Defence

1 INTRODUCTION

Historically the complexity associated with the analysis of military systems has resulted in the production of large and complex models for which the fitness-for-purpose is reassessed on a case-by-case basis (Moffat et al, 2004; Taylor and Lane, 2004). However, due to the emergence of new joint operating concepts and new military capabilities, coupled with constrained operating budgets; there is an on-going requirement for a more agile modelling solution to understand these systems of systems as they evolve. This challenge is being addressed through the Generic Aggregator Model Valuator (GAMOV) modelling approach (Glover and Toomey, 2012), which has enabled models to be constructed much more rapidly than previously possible. GAMOV was originally built upon the concept of a modelling approach that is driven by both data and functionality as opposed to the previous models that were purely data driven using hard-coded functionality. Although the main ambition underpinning GAMOV was to produce a more agile modelling approach, as our understanding of the implementation evolved over the course of its development, we identified that GAMOV opens up the potential to explore other avenues of research for the analysis of emergence and complexity. The underpinning architecture and organisation of the software components are such that GAMOV adheres to an emerging software paradigm, referred to as Object-Functional, which is a hybrid of the Object-Orientation (OO) and Functional programming paradigms. As a result, the components within GAMOV are loosely but explicitly coupled, enabling the flow of data around the system to be understood, moving us towards a white-box approach to verification and validation of the models.

At present, the resultant outcome spaces produced by a model constructed using GAMOV are very large, often requiring the exploitation of High Performance Computing (HPC) resources in order to generate
them. We also currently lack any automated means of being able to interrogate these large volumes of data to understand what the driving forces are within our modelling in a fully transparent and systematic way. This challenge has been recognised in the literature (Sterman, 2000), but a solution has yet to be developed by the modelling and simulation community. Whilst a solution to this problem is undeniably important, there are a number of key precursor stages that firstly need to be addressed. The first step requires some form of mechanism in order for analysts to more easily understand and make queries of the system. In consequence, research effort is currently under way to exploit the Semantic Web technology stack, which we believe shall enable us to realise the required mechanisms. This project is referred to as the Semantic Web Examiner of Emergent Phenomena (SWEEP), which shall draw on the mechanisms and theory briefly outlined in this paper in order to perform an automated analysis of models produced in GAMOV.

It is believed that the GAMOV approach, coupled with SWEEP, shall in due course enable us to begin to explicitly understand causation within our modelling. The remainder of this paper shall explain how the architecture of GAMOV and the organisation of its software components enable these ambitions to be realised.

2 THE OBJECT-FUNCTIONAL DESIGN OF GAMOV

Object-Functional is a software design paradigm that has progressively emerged over the last decade as a result of developers exploiting the capabilities of both OO and Functional programming (Odersky, 2014). Fundamentally the paradigm promotes approaches that decouple functionality from the data entities within the software. In many regards it is similar to RESTful design (Fielding, 2000), which is a design approach used for the development of web services across a client-server architecture. Under the Object-Functional paradigm, there are objects that hold data (Clients Objects) and function objects that provision services to these client objects. This is an adaption of classical OO design, where objects are normally constructs (often representing a real-world concept) that hold both data and functions. Because functions are their own object under Object-Functional (often referred to as first-class objects), they can be reused by multiple types of objects, including being passed as parameters to other components of a larger system provided the necessary interfaces are developed. Figure 1 below illustrates the layout of components.

![Figure 1: Object Functional interactions / GAMOV Entity Movement Example.](image-url)

GAMOV adheres to this principle by organizing its components into two groups. Firstly there are Entities, which are the client objects that are used to represent every actor within the simulation as a data object. Secondly there are Mediators, which are the function objects. All entities are fundamentally the same in terms of their attributes, although the data values will obviously differ. For example, both an entity representing a vehicle and a location have a movement attribute, however, the location entity will have a movement value of zero to reflect that it is a static location. In this way entities are only given purpose at runtime in a GAMOV model through the functions that their parameters enable them to interact with. In
order for two entities to have an effect upon one another, such as a combat encounter, the applicable
information from both entities is fed into the mediator within the function object that handles combat and
the result is then reflected back through changes to the entities data values, which in this case would be
damage. This allows for a plug-and-play approach to modelling as the function used to calculate combat
outcomes can be very easily swapped for another, in order to represent different interaction processes. This
enables models to be constructed iteratively, which aids in both verification and validation of the overall
model and supports studies operating under constrained budgets. This is enabled through models being built
in meaningful stages, from functions that are inherently reusable under the Object-Functional paradigm,
enabling studies to produce customer focused insight at each stage of the development of a model produced
in GAMOV. The concept of GAMOV is similar, but not influenced by, the work undertaken by Buss (2004)
on the development of SimKit. However, SimKit was written in Java which at the time of SimKits
development did not have any concept of a first-class object. As a result it is very likely that there will be a
higher degree of coupling between its components when compared to GAMOV written in Python.

In order to realise the benefits listed above, the flexibility offered by Object-Functional systems requires
a certain degree of discipline, guided by good practice concerning the design of GAMOV models. Whilst
developers using the GAMOV approach are able to shape a model both through changes to data and
function Object definitions, our recent experience has noted that some work remains in order to educate
users on how to think in these terms. This applies to both the analysts who will be designing and using new
models, as well as other developers who are external to the GAMOV team. For example, our initial
engagement with a near-term study has shown that analysts will instinctively think about the systems they
are modelling from the data perspective. Thus some of the early designs that have been proposed were
framed around modifications to the entity data structure in order to represent a new capability. This would
break one of the principles behind how a model is constructed using the GAMOV approach, specifically
that all entities use the same structure, which could otherwise create unnecessary coupling between the
components of the model. As explained in previous sections, a new capability should be framed in terms of
its mediator requirement within the system, which shall apply its effects onto the entities.

However, this is not a unique problem for the GAMOV approach; rather it is endemic to the current
understanding of Object-Functional design, which can be a difficult paradigm shift for people who are used
to expressing their problems in terms of OO. As the paradigm began to emerge, efforts were made to
produce a programming language to support it (Odersky, 2008). This resulted in the production of Scala, a
language that supported both OO and Functional whilst providing unique constructs that supported
hybridization of the two. However, one of the myths surrounding the usage of Scala is that the language is
‘hard to use’ (Erikson, 2012). Whilst the additional power and sophistication of the language undeniably
may present some unique learning opportunities for new users, it can be argued that much of the learning
curve is actually associated with understanding Object-Functional design as opposed to the syntax and
features of the language. Scala offers the capability to program either purely OO or Functional; however
when using a hybrid approach, if a developer does not express their problem correctly in terms of Object-
Functional design, they will find themselves at odds with some of the features of the language.1

3 THE SWEEP APPROACH

As previously stated the Object-Functional implementation of GAMOV has given us a loosely coupled
environment for modelling, this provides us the transparency to now understand the interactions between
its components. However, GAMOV alone does not provide any mechanisms for engaging with the model
components at any meaningful level beyond the traditional post-mortem analysis from the output files.
There is now a requirement for an automated mechanism that will allow the analyst to systematically
interrogate the model. In 2010 we identified that the Semantic Web technology stack could provide the

1 While the first version of GAMOV has been produced in Python, a second version (GAMOV 2) is being
developed as part of the SWEEP research in Scala, but designed so as to be backwards compatible with
GAMOV, enabling python entity and functions from GAMOV to be re-used in GAMOV 2.
Toomey

capabilities that we would need in order to interrogate the components of a GAMOV simulation through the use of relational mapping and ontologies.

The entity/mediator structure within GAMOV supports the ideas presented by Hoffman, Palii and Mihelcic (2011) whereby in order to extract any meaningful insight from using semantic web ontologies, there must be two form of Ontology applied to a model. Firstly there must be a ‘Methodological Ontology’, which underpins the simulation structure. The Entity structure within GAMOV meets the requirement for a methodological ontology as its base structure provides the consistent basis for every interaction with the mediator in each function object. The second is the ‘Referential Ontology’, which supports the analyst’s perception of a simulation. Within GAMOV these are the mediators, which embody our understanding of the defence domain.

The concept of SWEEP is built upon these ideas resulting in an approach that when coupled with GAMOV could perform the automated analysis capability as outlined in the introduction. It is envisaged that at a high-level SWEEP shall provide an iterative means of testing and understanding the scenarios built within a model, as outlined in figure 2 below.

Figure 2 The anticipated SWEEP workflow diagram

As per Figure 2, users shall submit queries through SWEEP as inputs to the model setup processes in the ‘Prepare’ phase. These queries shall be structured in terms of the vocabulary used to define the characteristics of the models. The exact format of these queries shall be determined following the initial experiments that shall be outlined in section 4. The key requirement here is that the model has been produced according to the GAMOV methodology and the necessary analytical criteria are described through a referential ontology, supported by appropriate semantic mark-ups. The data that satisfies the query is then produced by a set of agents as part of the ‘Process’ phase along with the normal model data and logs. These results are then visualised to the analyst using the most appropriate mechanism. The visualisation to be used shall be defined as an input as part of the ‘Prepare’ phase. As new insight is gained through the ‘Interpret’ phase, the entire process shall be re-run, through a combination of asking a new query or by changing the semantic mark-up enabling new understanding to emerge from the extant query set. Initially the interpretation shall be undertaken manually but as our understanding of this work develops, it is expected that the agents shall progressively undertake aspects of this phase. At present I am looking at how behaviours can be categorized correctly, which I have outlined in the next section.
Toomey

4 THE NEXT STAGE

The next stage in this research is to demonstrate the utility of the SWEEP approach on a simple GAMOV model. The model in question shall be a simple logistics model that uses an optimizer, called the Mission Planner, which it uses to select the optimal solution for transporting logistics between two points on a node-arc network. The mission planner\(^2\) is a genetic program that uses genetic algorithms and simulated annealing in order to optimize its solution space. The mission planner takes in the perceived state of the model with respect to the relevant commanders it is representing, including aspects such as decision policy. The mission planner then optimises a solution in order to meet the objectives it has been assigned to plan against. In the case of logistics, the mission planner will devise a solution for moving a set number of logistics across a node-arc network using a variety of movement channels (i.e. land, sea or air) in the most efficient manner. There shall be a number of transport units available with set capacities and a limitation on which logistics types it can carry. There may also be the presence of enemies on the node-arc network.

The first set of experiments shall look to prove whether the SWEEP approach enables us to quickly and automatically categorize behaviours within the model; specifically the distinction between problems that are as a result of poor model setup or data validity problems, as opposed to issues of analytical interest. For example, the mission planner has the concept of a ‘maximum order length’ which is the maximum number of actions the planner can use in order to achieve its objectives. This is important since it constrains the mission planner to solutions that can be optimized within reasonable timeframes for study use and prevents the solution space from expanding beyond the available computing capabilities. However if the maximum order length is set too low, the mission planner can optimize towards an incorrect solution. For example if the node-arc network is predominately land based, the mission planner may discount using any available air and water arcs, because the unloading of logistics onto a plane or ship and then subsequently offloading back onto a land vehicle at the end may constitute too many orders and results in exceeding the maximum order length. As a result the mission planner may order a logistics convoy to select routes that should otherwise be discounted, at least from the analyst’s perspective, because they are dangerous due to an enemy presence or are inefficient in terms of time. However, the balance of risk may allow for certain dangerous routes to be used if operational necessity demands. Obviously the mission planner is not explicitly aware of the concept of risk although its fitness function seeks to weigh operational risk so as to achieve militarily acceptable solution. Being able to distinguish between behaviours that are as a result of model limitations or data inputs and those which are of analytical interest will be a significant step forward for operational analysis, allowing better targeting of analytical resources, since tracking down the cause of a single model behaviour such as the one just explained can take a significant amount of time for an analyst through conventional post-mortem analysis. Being able to identify which behaviours to focus effort towards will thus help us to direct our analytical resources much more efficiently.

However, in order to achieve this, we need to be clear on our own understanding of what is permissible within the model so that the mission planner can make decisions that are closer to our understanding of reality and appropriately balanced against the risks. The experiments that I am conducting will test the current capability of mission planner to optimize logistic solutions against a progressively more complicated network of routes, threats and loading capacities in order for us to understand the bounds of its decision making. This will then enable a vocabulary to be constructed with Semantic Web technologies in order to help track the decisions being made by the current logistics implementation of mission planner. It is expected that this will enable us to quickly identify the decisions that are being made as a result of modelling features such as exceeding the maximum order length. This work shall also serve as an important first-step to understanding the degree to which the current logistics implementation of the Mission planner represents how real commanders assess risk in logistics operations, which shall serve as an important indicator concerning how to improve function objects such as the mission planner so as to make better automated decisions.

\(^2\) The mission planner is capable of optimizing a variety of different problem types; however in this instance it shall be focused upon logistics.
5 CONCLUSIONS

Within this paper, the benefits of Object-Functional model design and the potential that it unlocks have been outlined, based on our experiences of developing the GAMOV approach. Whilst using the Object-Functional paradigm improves our capability to deliver modelling, careful design following emerging best practice must be employed so as to avoid unintended coupling between model components. Comprehensive documentation to support the progressive broadening of the GAMOV modelling community is on-going in order to ensure that the benefits of the GAMOV approach are in practice realised. This research is now focused upon exploiting the potential that the Object-Functional paradigm has unlocked through the development of the SWEEP approach to support the model designer, model implementer and the analyst. In particular the research with SWEEP will serve as a solid foundation to build our understanding of the interactions between the components within our models through delivering automated analysis capabilities. The first step to realizing this capability shall be the experiments using the current logistics implementation of the mission planner, which shall be taking place throughout the remainder of 2015 and much of 2016.

REFERENCES


AUTHOR BIOGRAPHY

GARETH TOOMEY received a BSc (Hons) Intelligent System from the University of Portsmouth in 2007. He continued on at the University to receive his MSc (Merit) in Forensic Information Technology in 2008. He has been working at Dstl for 7 years, primarily focussed on the development of rapid modelling and analytical techniques. He is currently undertaking a part-time PhD, due to conclude in 2016, focussing on the development of a Semantic Web approach to automated analysis of large campaign models.
TOWARDS SELF-ADAPTIVE DISCRETE EVENT SIMULATION (SADES)

Dr. Kathy Kotiadis

University of Kent
Canterbury, Kent CT2 7NZ
K.Kotiadis@kent.ac.uk

ABSTRACT

Systems that benefit from the ongoing use of simulation, often require considerable input by the modeller(s) to update and maintain the models. This paper proposes automating the evolution of the modelling process for discrete event simulation (DES) and therefore limiting the majority of the human modeller’s input to the development of the model. This mode of practice could be named Self-Adaptive Discrete Event Simulation (SADES). The research is driven from ideas emerging from simulation model reuse, automations in the modelling process, real time simulation, dynamic data driven application systems, autonomic computing and self-adaptive software systems. This paper explores some of the areas that could inform the development of SADES and proposes a modified version of the MAPE-K feedback control loop as a potential process. The expected outcome from developing SADES would be a simulation environment that is self-managing and more responsive to the analytical needs of real systems.

Keywords: Reuse, Automation, Real-time, Self-Adaptation.

1 INTRODUCTION

The practice of simulation has been categorised into three modes (Robinson, 2002) simulation as software engineering; simulation as a process of organisational change; and simulation as facilitation. The first mode, supports models with a wide range of goals requiring a very accurate representation of the real world with models requiring years to develop through software engineering and requiring ongoing maintenance. This mode, unlike the other two modes, would aim for a long term use of a model. The second mode models typically have well-defined goals, aim for a reasonable representation and are developed using ‘off the shelf” software. In this mode the maintenance costs of the model soon outweigh the benefits associated with its continued use, making them useful in the short to medium term. The third mode, involves using a model to support the involvement of a group of stakeholders in the modelling process (Tako and Kotiadis, 2015). It would almost certainly involve an ‘off the shelf” package and the model would be ‘shelved” following the intervention. An inspection of recent simulation literature (Taylor et al, 2013) reveals that another mode of practice is emerging that aims to support ongoing rapid decision making but does not necessarily require or aspire to the extensive software engineering development process required for mode 1. Indeed, research that fits with this mode of practice includes the development of generic models for reuse (e.g. Bartholet et al, 2005; Pidd, 2002), real time simulations (e.g. Tavakoli et al, 2008) and automation of various aspects of the modelling process (e.g. Huang et al 2011). However to a large extent these research areas are being developed independently from each other. This paper proposes bridging these areas in order to develop Self Adaptive Discrete Event Simulation (SADES) which could potentially be more effective at meeting the aim of this emerging mode of practice.

SADES would sense the need to run a simulation, from a stream of real time data, and then automate the evolution of the modelling process leading to a solution space. The aim of this mode would be the ongoing use of DES models in order to continually support decision making, ideally with very limited support by the modelling team.
The aim of this paper is to provide a vision of SADES and a focus on the research areas that could be bridged to develop SADES, such as the ones mentioned earlier and areas outside of simulation and OR, such as autonomic computing and software engineering for self-adaptive systems. The paper also considers how SADES might work and puts forward some preliminary ideas based on the MAPE-K adaptation control loop originally put forward by IBM (Kephart and Chess 2003). The impact of SADES would be to enable systems that benefit from simulation to be even more responsive to changes and enable ultimately the practice of simulation to become more sustainable.

The paper is organised into four further sections. The next section puts forward the vision of SADES, though an example, in order for the reader to appreciate how the various streams of research, subsequently described, can support the development of a SADES. Next a number of research areas are explored that could support the development of SADES. In the following section the MAPE-K loop is used to explain how SADES might work. The paper concludes with some proposed areas for further research.

2 ILLUSTRATING THE VISION OF SADES

In discrete event simulation (DES), a modeller will typically use a model of a system to explore various ‘what if’ scenarios. For example, a model simulating a UK hospital Accident and Emergency (A&E) department might be used to explore the number of doctors needed to meet the ‘four hour’ waiting target set by the Department of Health. Once an optimal or desired scenario is identified, it is implemented leading to changes in the actual system. Even in cases where a simulation model is used for understanding, so that no particular scenario is implemented as such, the system or the system’s goals are likely to evolve over time. If a simulation model is to be reused in such instances it would require some updating to reflect the changes in system of interest or its goals. Given the costs of developing a model and its usefulness as a decision aid, it would seem sensible to reuse a model by updating and maintaining the model to support future decisions. However, that is often not the case because of the associated costs, such as, involving a modeller. One way of extending the lifecycle of DES models is to automate their evolution, so that the role of a modeller is largely restricted to the development stage of the simulation. For example, the application of a SADES to a UK A&E (Accident & Emergency or Emergency department in US) would enable a new simulation to be triggered whenever real-time data indicate that the ‘four hour’ target has been breached (Figure 1). This could be achieved through sensor technology attached to patients or devices within an A&E as well as connecting to databases and the internet. Sensors could also be attached to staff and other A&E resources. A smart management system, in the example described for simplicity as Monitor & Control, would automatically go through the modelling process so that a current model of the system is automatically generated, which is then automatically run to produce a range of feasible scenarios that fit with current goals. This would allow the stakeholders to respond quickly to a range of problematic situations without incurring further modelling costs.

Figure 1 An example of an A&E Self Adaptive DES model
3 RELEVANT STUDIES

This section will explore some areas mentioned in the introduction that could initially drive the development of SADES. These areas are: DES reuse or building generic models, automation of DES processes, real time simulation and Self Adaptive Software Systems.

3.1 DES Reuse

Central to the philosophy of SADES is the idea of a long term use of a model or some aspect(s) of a model, which can be described as a form of reuse. Reuse is a concept that was discussed in the 60s in software (McIlroy, 1968) and as early as 1986 (Sargent, 1986) in simulation. Software reuse is the isolation, selection, maintenance and utilisation of existing software artefacts in the development of new systems (Reese and Wyatt, 1987; Robinson et al, 2004). Reuse in simulation spans a spectrum that ranges from full model reuse, to component reuse, to function reuse, to code scavenging (Pidd, 2002; Robinson et al, 2004). The benefits of reuse can be found in many forms such as saving effort, reducing development costs and enabling a quicker development time than developing a model from scratch (Waziruddin et al 2003).

More recently the idea of reuse has evolved to the creation of generic models that are built to fit a particular context. For example the District General Hospital Performance Simulator (DGHPSim) (Gunal and Pidd, 2011) is a (healthcare) domain specific example of reuse where a generic and restrictive model is built using data that would be found in a hospital’s database such as a UK’s Patient Administration Systems (PAS) and national databases such as the UK’s Health Episode Statistics (HES). In another example, Pinto et al (2015) put forward a generic model for an ambulance system and even propose a generic method to develop such models.

Another aspect of reuse that has attracted interest is constructing complex systems from off-the-shelf components (Bartholet et al, 2005). Similar to whole model reuse, a simulation component that can be reused several times can save a great deal of time, money and human effort (Davis and Anderson, 2003). Indeed the benefits are even greater when reusing more than one component. Composability is about ‘the capability to select and assemble simulation components in various combinations into valid simulation systems to satisfy specific user requirements’ (Petty and Weisel, 2003). However reusing any component is not straightforward as it may only be valid under certain experimental frames e.g. a specific context (Zeigler et al, 2000). Spiegel et al (2005) explain that simulation composability and reuse will require comprehensive identification of constraints and where components are reused they must be designed with that in mind.

3.2 Automation in DES

A key aspect in the development of SADES is the automation of the evolution of the modelling process. The modelling process includes conceptual modelling, model coding, experimentation, implementation as well as validation and verification that takes places throughout the simulation lifecycle (Arthur and Nance, 2007). In principle there is reasonable agreement over what takes place during these stages but it is very hard to pin point these in detail for every study. Particularly for a process such as conceptual modelling, that is still considered by many to be an art. In this section we will explore some work on automation that addresses some of the activities that take place during the modelling lifecycle but for reasons of simplicity the real time simulation literature has been separated out, to be explored in the next section.

Data collection and analysis is largely an activity that relates to the initial stages in the modelling lifecycle. Padilla et al (2015) put forward a semi-automated initialization of input variables that are challenging to quantify and require additional processing to be assigned their initial values. The study supports the retrieval of data from structured and unstructured data sources and generates input data. They propose that further work could include exploring data from different sources such as data from social media and whether the assigned values yield more accurate simulation results (validation and verification). Schoogh et al (2010) also explore this area of DES automation and particularly so for manufacturing so that simulation models can be more frequently updated and integrated in the daily work of production engineers.
Their study explores MTConnext tool for the collection of raw data and GDM-Tool for data processing (Bengtsson et al, 2009). They compare their approach to manual data handling and find that just for processing raw data they were able to reduce their work from half a day to 1 hour and conclude with future work to include some additional refinements for the tools.

Huang et al (2011) propose a data driven approach to automatically generate models using prebuilt and validated model components. In fact they view their approach as an automated reuse of model components. They provide an example of the automated generation of a rail network model using available data sources. They conclude that future research should focus on an automatic model calibration and different data analysis methods and techniques to estimate the models parameters.

A recent paper (Batarseh et al, 2015) contributes ideas on automation supporting model coding. The paper explores the use of model driven architecture (MDA) from software development to support the automatic creation of simulation models. The approach empowers domain stakeholders with tools for defining the problem to be solved and formally integrating the definition with the simulation tool. A key feature of the approach is expressing the semantics of the domain which is captured in a library. The authors suggest that further work on addressing warm up issues and transients is needed so that domain experts do not misuse the simulation tool.

The stage of experimentation involves model calibration as well as searching the solution space. Hoad et al (2010a,b) put forward an automatable algorithm to select the number of replications and automated the process of estimating the warm up length for steady state systems which have since been incorporated within the SIMUL8 simulation package (www.simul8.com). Hoad et al (2011) put forward the AutoSimOA framework explored automating the analysis of simulation output. It includes generating replications, warm up and run length estimation.

Waziruddin et al (2003) put forward a study that could inform both automated model coding and automated experimentation (searching the solution space). Specifically they explored the process of ‘coercing simulations’ that involves a combination of code modification and simulation behaviour optimisation with the goal of driving the behaviour of a simulation so that it satisfies a new set of requirements. They propose that some aspects can be automated so that ultimately the transformation of a simulation to meet new requirements can be performed more efficiently.

### 3.3 Real time simulation and dynamic data driven simulations

One characteristic that SADES simulations should have is the ability to sense the need to run a new simulation. This would require the ability to automatically monitor real time data and process that data so that it can be used in the simulation. Hence it involves automation of processes that fit within conceptual modelling and model coding. Two areas have been identified as relevant in the development of these capabilities: real time DES and dynamic data driven simulations.

Real time DES are essentially simulation models using real or near to real time data rather than historical data and are meant to overcome shortcomings of traditional simulation (other modes). The shortcomings non real time DES (Tavakoli et al, 2008) are: a. being time consuming because of the manual collection and analysis of data, b. time depended on historical data that soon enough become obsolete, c. inaccurate for prediction because of issues with historical data and d. costly because of effort and expertise required for the data collection. Real time simulation modelling has been predominantly been explored within a number of contexts such as manufacturing and healthcare (Tavakoli et al 2008; Mousavi et al 2011), construction (Song and Eldin, 2012; Vahdatikhaki et al, 2013; Vahdatikhaki and Hammad, 2014) and road and traffic management (Henclewood et al, 2008; Henclewood et al, 2010; Henclewood et al, 2012; Jaoua et al, 2012a,b).

Most of the studies mentioned explain the functionality of their real time simulation using a framework or architecture. We will now explore some of these:

- An early study by Lee and Fishwick (1999) put forward a multimodeling methodology called OOPM/RT (Object Oriented Physical Modeler for Real Time Simulation) that minimises the
modellers interventions, through a semi-automatic method, and hence supports the modelling of real time systems. The approach considers the question of how to determine the optimal model that simulates the system by a given deadline while still producing valid outputs at an acceptable level of detail. They identify some issues that were unresolved in OOPM/RT such as optimality (of the level of abstraction) and validation challenges.

- Tavakoli et al (2008) propose an architecture for data acquisition and data restructuring called Flexible Data Input Architecture (FDILA) which they illustrate with examples in manufacturing and healthcare. In a later study Mousavi et al (2011) put forward the SIMMON architecture that allows extensive data acquisition technologies that prepares the information for the real time simulation package (restructuring) and it subsequently translates the simulation results into easy to comprehend business performance parameters. The extension to this work mainly related to further testing. Particularly insightful for real time simulation is their work on input variable selection methods using a case study in the deep drilling industry (Tavakoli et al 2013).

- Vahdatikhaki and Hammad (2014) put forward a framework for near real time simulations of earthmoving projects that use location tracking technologies. The framework among other things captures different levels of sensory data, then processes it, analyses it and filters it. The authors suggest that future work will include enabling distribution fitting from the data captured and updating the functionality of the framework to include the possibility of updating the logic of the model and the sequence of activities when required. Similarly, Song and Eldin (2012) propose an adaptive real time tracking and simulation framework for heavy construction applications with look ahead scheduling. The idea here is to react to changes in real time or near to real time. They conclude with future work to include to finding a way to collect and transmit data only when a meaningful change occurs in the real system and to enable an automated generation of construction simulation models from scratch using extracted knowledge about key model parameters and precedence logic.

- Henclewood et al (2010) use real time sensor data e.g. from loop detectors or video cameras, to better manage traffic operations. In subsequent research, Henclewood et al (2012) conclude that further work is needed to develop new calibration procedures as models for traffic simulations are often calibrated using data from a single time period.

We now explore some ideas from dynamic data driven application systems (DDDAS) (Darema, 2004; Darema, 2005; Gaynor et al, 2005; Farhat et al, 2006; Kennedy and Theodoropoulos, 2006) given their potential to extend real time DES simulations and ultimately inform SADES. ‘The vision of DDDAS, goes beyond the current concepts of real time control, in terms of the concept of interaction of the application simulation with the measurement system, the range of applications, the streamed input data and the scope of systemic approach to addressing the ensuing challenges: assessment of data quality, incorporation of uncertainty, ability to combine different kinds of data taken at different times, application algorithms with stable response to the streamed data, distributed systems (for sensing, for computation and for output, and for data storage), development of effective interfaces of applications with measurement processes, supporting the execution of such applications with dynamic resources requirements’ (Darema, 2004, 663-4).

DDDAS entail the ability to incorporate additional data (archived or real time) and for applications to dynamically steer the measurement process (Darema, 2004). DDDAS has been explored within a number of contexts, not always connected to a simulation, and is meant to be particularly useful in supporting crisis situations where rapid decision making is essential (Gaynor et al, 2005; Farhat et al, 2006, Yan et al, 2009). We could say that a simulation model supports situations when a problematic situation emerges (i.e. crisis) so a symbiosis is perfectly reasonable and has been explored within other types of simulation such as agent
based simulation (Kennedy and Theodoropoulos, 2006). Indeed Henclewood et al (2008) propose a methodology for a real time (dynamic) data driven simulation. A key part of the research is the development of dynamic data driven application (agent based) simulation (Fujimoto et al, 2006).

3.4 Self-Adaptive Systems

In order for DES to self-adapt it needs to be able to automatically go through the modelling process and to automatically sense the system and its wider environment and generate a model of the same system, albeit in its new (possibly changed) state. This idea of self-adaptation for DES is taken from the field of self-adaptive (software) systems. Self-adaptive systems have been explored within the different areas of software engineering, as well as other related research communities, such as control engineering and autonomic computing (Brun et al, 2009). These communities use common language and ideas and the self prefix indicating the autonomous nature of the system. ‘The essence of autonomic computing systems is self-management, the intent of which is to free system administrators from the details of system operation and maintenance and to provide users with a machine that runs at peak performance 24/7.’ (Kephart and Chess 2003, p42). An analogy to this is the autonomic nervous system in the human body, which among other things enables the pupils to adjust to sunlight, enables sweating to keep cool, adjusts breathing rate etc. with no conscious recognition or effort from our part for all this to take place. Autonomic computing grew out of a need to prepare for the next generation of computing systems with increased interconnectivity and integration. Some of the original characteristics of autonomic computing include self-management, self-configuration, self-optimisation, self-healing and self-protection. Brun et al (2009) and more recently Macias-Escriva et al (2013) put forward a number of challenges with a particular emphasis on the design of a self-adaptive system and the importance of feedback loops.

Feedback loops provide the generic mechanism for self-adaptation and typically contains four key activities: collect, analyse, decide and act (Brun et al, 2009). The data typically collected from environmental sensors and other sources reflecting the current state of the system is cleaned, filtered, pruned and stored for future reference to reflect past and current states. The system subsequently analyses the data and then a decision is made about how to adapt the system so it reaches a desirable state. The final activity in the loop is to act by implementing the decision. The properties of the control loop will ultimately affect the system design, architecture and capabilities (Brun et al, 2009). Therefore it is important that the control loop and its properties are adequately explained.

One of the key feedback loops in software engineering is the MAPE-K which stands for Monitor, Analyse, Plan and Execute over a Knowledge base (Brun et al, 2009) as shown in Figure 2. In the feedback loop the autonomic manager and the managed element correspond to the controller and the process respectively. A key addition to the generic mechanism is the knowledgebase which is shared by the MAPE engine. The monitor senses the managed element, filters and stores the event data in the knowledgebase for future reference. The analyser compares event data against patterns found in the knowledgebase to diagnose and store symptoms for future reference in the knowledgebase. The symptoms are interpreted by the planner that subsequently devises a plan to execute a change in the managed element through its effectors. The autonomic manager, with the feedback loop at its core, will adjust the managed element if necessary according to its control objective (Brun et al 2009). In the subsequent section we will explore how this MAPE-K adaptation loop can be used to support the design of SADES.

4 A FEEDBACK STRUCTURE FOR SADES

This section will explore how the MAPE-K loop might be re-interpreted as a structure for SADES, on occasion referring to the A&E example. In the vision of SADES the evolution of the modelling process is automated. One view of how MAPE-K could be adapted for SADES can be seen in Figure 3 with the managed element being the simulation model. The SADES in Figure 3 differs mainly with respect to the loop in MAPE-K (Figure 2) because the simulation modelling process is not a linear activity. Activities relating to analysis for example, take place before the model is determined (e.g. analysis of raw data) as
well as on completion of a model run (e.g. analysis of warm up time). Hence the interpretation of the
MAPE-K loop for SADES takes the form of a centrally controlled structure (autonomic manager) rather
than a loop.

![Figure 2](image)

**Figure 2** The MAPE-K loop reproduced from IBM (Keller, 2005)

![Figure 3](image)

**Figure 3** Re-interpreting the MAPE-K structure for SADES

The control of the sequence of the elements in the structure and the simulation model, assigned to the
autonomic manager, is based on the information held in the knowledgebase (Control with Knowledge in
Figure 3). The knowledgebase would be a storage for the data and information that is important to the
evolution of the modelling process, such as the simulation objectives. The content of the knowledgebase
could be updated either by the elements in the SADES structure, the autonomic manager or by human
intervention. For example, the expectation for the objectives is that these could be automatically or
manually altered over time to reflect changes to the needs of the system of interest e.g. the A&E
department. We will now explore how the some of the activities in the DES modelling process could
potentially be allocated to the four elements of the SADES control structure.

The **Monitoring** element would undertake the activities of sensing information about the environment
and system of interest to be simulated. This corresponds to the data collection in the DES modelling process.
This data would be filtered and stored in the knowledgebase for future reference. For example for a SADES
of the A&E department, it might sense and collect data about the available human resources for each shift,
the arrival, queuing and throughput times for each patient, the processes involved and information about
the categorisation of need/severity of the emergency (e.g. low, medium, high). The data would need to be
sorted to match up to shifts, days etc, possibly translated into a different time unit or format and stored in
the knowledgebase for future reference. Data to be used for Validation and Verification could also be sensed
and stored separately for that purpose in the knowledgebase. Other data sensed might be sourced from the
web, for example relating to targets (e.g. 4 hour target) or epidemiological data relating to the target
population of that A&E. The autonomic manager could also be designed to have capability to request new
data to be sensed by this element if the objectives are altered and therefore alter what is monitored.

The **Analyse** element could undertake analytical activities on the data collected and held in the
knowledgebase. For example, the analysis might involve calculating a moving average for the queuing time
cross each shift and compared to data held in the knowledgebase on the acceptable levels of the relevant
KPIs. KPIs would be matched up to the objectives of the A&E, held in the knowledgebase. The Analyser in
the loop would also determine parameters for distributions or the distributions themselves for the simulation
model. Additionally it could analyse the data for the warm up time and run length time (if necessary) and
the number of replications. It would also analyse the output data from the resulting scenarios and
statistically compare the scenarios. This element could also perform some comparisons of the analysed
V&V data to the model output. The analysed data, distribution selection and parameters, calibration
parameters, V&V comparisons, would be stored for future reference in the knowledgebase for planning and
execution. Hence this element of the loop replaces DES analytical activities that are typically found in all the different parts of the modelling process.

The main roles of the planner or planning element could be that of: a. interpretation of the analysis, by consulting the knowledgebase, b. to determine if a simulation should be undertaken and c. to schedule subsequent modelling activities. For example, any breaches of the system (e.g. four hour target in A&E) KPIs would be interpreted by the planner, that would devise a plan to execute in the managed element (the simulation model). The planner would determine the model content, the experimental design by setting out the inputs and variation. The plan to be executed could include changes to: inputs (e.g. number resources, distributions & parameters), model content (e.g. type of resources, processes, rules etc.) which would be deposited in the knowledgebase. Hence this element of the structure replaces the DES modeller decision making activities that are typically found in all the different parts of the modelling process.

The execution element or executor could consult the plan held in the knowledgebase and translate the plan into sequential or concurrent actions to generate the model content, inputs and assign the calculation of KPIs for each scenario and ensure that the planned activities are executed in the correct order. These activities could also detail the interaction of the simulation output with the elements of the SADES structure. The executor instructions would be deposited in the knowledgebase for action.

The autonomic manager would initialise the simulation by following the execution instructions. The data from the simulation runs would be stored in the knowledgebase for subsequent analysis. It is envisaged that the autonomic manager might revisit elements of the SADES structure (possibly forming a loop) several times before the process is completed. At the end of the process, the solution space would have been captured in the knowledgebase. This in turn would be reported to the stakeholders by the autonomic manager. Ideally some data from the V&V comparisons would also be provided alongside the solution space.

5 CONCLUSION

Brun et al (2009) when referring to the studies undertaken in the many communities exploring self-adaptation describes these as isolated and from their own perspective. This is also true for the studies undertaken in DES aiming for ongoing and rapid decision making, but this study has set out to change this and provide a common focus in developing SADES. SADES could potentially follow the automation-based paradigm for DES which aims to reduce the cost and duration of software development; achieve maintainability; develop reusable software; increase user involvement and reduce software portability issues (Balci and Nance, 1987). Balci and Nance (1987) when proposing the paradigm suggested that it would require further technological developments. Advancements in software engineering and in the state of the art of modelling and simulation over the last two decades, lead us to predict that this mode is feasible providing the simulation community come together to bridge the existing research.

Indeed, studies on DES reuse or building generic models, automation of DES processes, real time simulation and Self Adaptive Software Systems could feed into the development of SADES. For example the studies in real time simulation reveal that a great deal has been accomplished with regard to collecting and preparing data for simulations that could feed into the monitoring element of the SADES (Figure 3). However some existing advancements are context specific and there are still other hurdles such as the automatic generation of models, sensing meaningful change in the data, calibration of models for different time periods etc. Similarly, a lot of the research from reuse could feed into SADES because it is entirely possible that previously generated models of that system could be reused. For example it could be reuse of code or components or some data. An aspect of reuse that will support the development of SADES is understanding how to develop or design components or models that will be reused in a modified state. Another aspect that will be very important in the development of SADES is composability, which is about combining and recombining components to meet different sets of requirements without substantial integration efforts (Bartholet et al, 2005). To sum up there are a considerable number of studies that could be used to inform the development SADES but this will be easier to achieve as part of a more collaborative research agenda. The envisaged impact of SADES on future systems that will be more connected (think of
Internet of Things), is to offer a more sustainable and responsive mode of simulation. The expectation is that by bridging the relevant research we can better understand SADES, design it and ideally construct examples of it.

ACKNOWLEDGMENTS

Kathy Kotiadis acknowledges the support and the sponsorship of the Daphne Jackson Trust, the University of Kent and Simul8 in undertaking this research.

REFERENCES


Kotiadis


Kotiadis


**AUTHOR BIOGRAPHIES**

**KATHY KOTIADIS** is a Daphne Jackson fellow at the University of Kent. She is a former assistant professor at the University of Warwick and has a PhD from the University of Kent. Her research interests include the process of modelling, conceptual modelling, discrete-event simulation, Soft Systems Methodology (SSM), facilitated approaches in healthcare and developing Self Adaptive Discrete-Event Simulation. Her publications can be found on https://www.researchgate.net/profile/Kathy_Kotiadis
WS-PGRADE WORKFLOWS FOR CLOUD-BASED DISTRIBUTED SIMULATION

Nauman R. Chaudhry, Athar Nouman, Anastasia Anagnostou, Simon J. E. Taylor

Department of Computer Science, Brunel University London
Kingston Lane, Uxbridge, Middlesex, United Kingdom UB8 3PH
nauman.chaudhry@brunel.ac.uk

ABSTRACT

Modeling and Simulation (M&S) is used for systems analysis and decision making in existing or new systems. Modeling large and complex organizations produces large-scale simulations that are difficult or impossible to run on a single computer. Such experiments execution requires high computation. Distributed Simulation (DS) allows modeling of large systems as smaller submodels that execute on different nodes of a computer network and interoperate with each other in order to compose larger systems. Furthermore, cloud computing offers on-demand access to multiple compute resources. Consequently, being able to run DS on cloud resources allows for more experimentation with large-scale simulations in a cost-effective way. However, deploying DS and in fact Cloud-based DS presents significant technical challenges. This paper proposes a framework for deploying DS on the cloud in a transparent manner using the CloudSME Simulation Platform based on WS-PGRADE workflows. A healthcare case study is used to demonstrate our approach.

Keywords: Distributed Simulation, High Level Architecture, cloud computing, MSaaS

1 INTRODUCTION

Modeling and Simulation (M&S) techniques are used to study changes in existing or new system’s behaviour to help in decision making, where it is too expensive, impractical or even impossible to implement in the real world. M&S is now widely recognised in the areas of defense, computer & communication (networks), transport (traffic control system), healthcare, behavioural sciences, ecology & environment, biosciences, manufacturing & production, services (e.g., Banks) and economy (Taylor et al 2012). M&S not only requires skilled domain expertise and programming skills, but also high computational resources to simulate large and complex models. Usually, there is a memory resource limitation for a single execution unit. Furthermore, the simulation execution run time could be substantial. Therefore, large-scale models could be distributed into smaller models running on several processors, which could remarkably reduce the execution time (Fujimoto, 2000). Distributing a simulation model, however presents some technical challenges both in technical skills and the availability of distributed computing infrastructures. One of the technical challenges is to develop interoperable independent models running on different networked computers, while the distributed computing infrastructure has to provide a configured environment of networked computers (in some cases geographically dispersed). Having Distributed Simulation (DS) models running as a service on cloud resources significantly reduces the latter constraint and also increases reusability. The cloud computing paradigm attracts increasing numbers of M&S practitioners wishing to perform their simulations on the cloud. Developing solutions for cloud computing is also difficult without expertise due to complex technologies. The CloudSME Simulation Platform (CSSP), based on CloudBroker (provided by CloudBroker, CH) and WS-PGRADE/gUSE (provided by MTA SZTAKI, HU) (Kacsuk et al 2012), was developed to support Modeling & Simulation as a service (MSaaS) in manufacturing and engineering and simplify the deployment of M&S software and applications on various cloud resources (Taylor, 2014b).
In this paper, we propose a framework that utilizes the CSSP to run an HLA-based distributed simulation as a service and investigate how we can automatically configure and use a DS on cloud resources. The paper first introduces the distributed simulation and cloud computing concepts. It then presents the CloudSME Simulation Platform followed by the proposed framework for cloud-based distributed simulation using WS-PGRADE workflows. A case study is then presented to show how HLA-based distributed simulation can be easily run on cloud resources, i.e., the academic cloud provided by the University of Westminster (UoW).

2 DISTRIBUTED SIMULATION ON THE CLOUD

Distributed Simulation can be defined as “the distribution of the execution of a simulation program across multiple processors” (Fujimoto, 2000). DS techniques, therefore, make it possible for a single model to be divided and simulated over several processors or multiple models running on different processors to be joined together.

In DS systems, the participating simulation models are able to interoperate with each other. The simulation models can send information to and receive information from other simulations and be able to operate effectively together, i.e., sending the right information to the right destination and at the right time, also without adding prohibitive communication time overhead. Hence, data and time synchronisation is essential in DS systems.

The IEEE 1516 High Level Architecture (HLA) (IEEE, 2010) standard for DS is used to achieve interoperability and reusability between simulation models. The standard was originally developed by the US Department of Defense (DoD). IEEE 1516 HLA specifies a set of standard rules and processes to support DS.

For the DS interface and the implementation of the HLA standard, there are several Run-Time Infrastructure (RTI) implementations presently available, some are commercial RTIs, such as Pitch pRTI, MAK High Performance RTI, RTI NG Pro etc., and others are open source RTIs, such as Open HLA, CERTI, poRTIco etc. Figure 1 illustrates the HLA IEEE 1516-2010 standard, where each federate represents a simulation model and all the participating federates joined together represent a federation, communicating through the RTI.

Figure 1 High Level Architecture

DS is typically used to model large and/or complex systems which require a significant amount of computing resources. It however requires domain specialist to not only model the system, but also in some cases manage the distributed computing infrastructure. Having access to required distributed computing infrastructure can sometimes be challenging. Therefore, running DS as a Service on cloud eliminates the simulation modelers’ challenges of maintaining and accessing the infrastructure.

Computing power as a utility was introduced in the 1960s (Hill et al 2013). “Cloud computing” refers to accessing internet-based computing resources and as a term first appeared in the mid-2000s. The National Institute of Standards and Technology (NIST) has made efforts to standardize the terminology of cloud computing (NIST, 2013). Mell and Grance (2011) and NIST (2013) define cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing
resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”. Cloud computing provides flexibility to the users for its computational resources, applications, access, etc. that could be tailored according to the user needs. Cloud services can also be deployed according to the security requirements of the organizational structure. Cloud computing offers three defined service models, i.e., Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

Cloud computing can support parallel and distributed simulation as a service by providing the required high performance computing infrastructure and its maintenance. Along with the benefits of cloud computing there also some concerns such as security of data, reliability of services and slower execution of code than code execution on cluster node (Fujimoto et al 2010). There are also difficulties for DS used with optimistic synchronization protocols. An approach was proposed to accelerate the execution speed of federates at comparable speed (Li et al. 2013). A PaaS RTI-supporting architecture was proposed by combining RTI and web services to run on central servers, so as to provide DS platform to simulation users on WAN (Feng et al 2010; Zhang et al 2012). The latest work is more focused on the associated problems and usage of public clouds (Vanmechelen et al 2013; Yoginath et al 2013). A multi-agent system approach was proposed for model partitioning between the execution nodes on the cloud (D’Angelo et al 2014)

In order to support reusability and interoperability specific cloud services need to be developed to support M&S. M&S as a Service (MSaaS) is considered as a separate cloud service model as it allows users standardized access to build their own simulation models by specialized configured software to run simulation experiments (Taylor et al 2014).

3 CLOUDSME SIMULATION PLATFORM

The CloudSME Simulation Platform was developed by the Cloud-based Simulation platform for Manufacturing and Engineering project (www.cloudsme.eu) funded by the European Commission. The main aim of the project was to develop cloud-based simulation services and platforms that enable Small and Medium Enterprises (SMEs), mainly in manufacturing and engineering, to access simulation software and services, and speedup simulation experimentation by using cloud resources and other Distributed Computing Infrastructures (DCIs) such as grids, HPC clusters, etc. CloudSME supports MSaaS by combining PaaS and SaaS and allowing simulation software and services providers to build SaaS solutions for end-users and ultimately deliver MSaaS.

The simulation software accessing mechanism should hide entirely potential heterogeneity and complexity of cloud platform, as well as permit the usage of various clouds that may use different cloud middleware. CSSP achieves this by the combination of WS-PGRADE/gUSE workflow deployment and development services (provided by MTA SZTAKI, HU) (Kacsuk et al 2012) and Cloud Broking services (provided by CloudBroker, CH). In other words, it acts as PaaS that supports simulation services deployment, as well as access, in user-transparent way, to various cloud and HPC resources.

CloudSME has developed MSaaS solutions for two types of end-users. That is, a) simulation software providers and b) end-users who use simulation for their business processes improvement or as part of their business offering. Currently, eight simulation software providers are involved in the project, including 2MORO (FR), INGECON (ES), ASCOMP (CH), SIMUL8 (UK), SIMSOFT (TR), CMCL (UK), DHCAE (DE), and OUTLANDISH (UK) that offer different simulation tools. Moreover, different end-users are included CUTTING TOOLS (UK), EUROBIOS (FR), PODOACTIVA (ES), SAKER SOLUTIONS (UK), PROYFE (ES), HOBSONS (UK), BASEPRO (IT), G-VOLUTION (UK), PROCENG (CH), TIDYBOOKS (UK), OZDEKAN (TR), GOKDOGAN (TR), and IOR (IT). Furthermore, the Repast Simphony open source simulation software was deployed as MSaaS in CSSP (Taylor et al 2014b). The latter, Repast MSaaS, is used in this study for the implementation of the case study. CloudSME is led by the Centre for Parallel Computing at the University of Westminster (UK). The cloud-based product development and end-user adoption is managed by Brunel University (UK), UNIZAR (ES), University of Westminster (UK), SZTAKI (HU), and CloudSigma (CH) provide cloud resources.
Figure 2 illustrates the CSSP components organized into three layers reflecting the cloud computing stack. These layers include the Simulation Application Layer (SaaS), the Cloud Platform Layer (PaaS), and the Cloud Resource Layer (IaaS). Moreover, the Cloud Platform layer consists of two parts, i.e., WS-PGRADE/gUSE and CloudBroker Platform (CBP).

CBP provides adapters for the available cloud resources that enable the creation, running, and managing of simulation software images on cloud virtual instances. The simulation software in order to be offered as MSaaS must have been deployed on the respective cloud resources. The deployment process varies and depends on the operating system. Generally, a simulation software executable and runtime scripts are stored in a repository and are called for instantiation and job configuration when requested. The MSaaS Application Patterns and Deployment configurations are formed by these scripts and executables. The CBP services can be accessed directly using its API or via the WS-PGRADE web-based workflow management system. WS-PGRADE can be accessed using two different APIs, described in detail (Taylor et al. 2014a).

WS-PGRADE is used for workflow creation. Workflows are directed acyclic graphs that can be configured and stored on a workflow server (gUSE). They consist of nodes, job functionality, and arcs, channels for file transfer among the nodes. A node may have one or more input and output ports (green and grey squares, respectively) and each port is denoted by an integer number. A workflow determines the sequence and dependencies of a simulation run. An example can be seen in Figure 3. The specific workflow consists of three nodes that perform different jobs. Initialize Node has one output port (port 0) linked with the input port (port0) of the next node Execution Node. Execution Node will start performing its assigned task only after receiving the required input from Initialize Node. Once this task completes execution, it will generate an output file which will pass through its output port (port1) to the next node. Results Node receives the Execution Node’s output through its input port (port0) and then starts executing its task.

For flexibility, workflows are “agnostic” and therefore has the functionality to run on any DCI as long as the interface is supported by the workflow manager via the DCI Bridge. CBP is seen as a separate DCI.
4 CLOUD COMPUTING FRAMEWORK FOR DISTRIBUTED SIMULATION

This section explains the proposed framework for DS execution on the cloud. As mentioned before, it is important to access cloud resources transparently and therefore hide the underlying complexity from the end-user. To do so, the proposed framework enables the execution of DS on the cloud in a simplified way by using workflows in the CSSP. The main requirement for the implementation of the framework is the cloud deployment of the RTI and the simulation software as well as their dependencies. In our case, these are the poRTIco RTI implementation of IEEE 1516-2010 and the Repast Simphony Toolkit. Both are Java-based software. Therefore, Java runtime also must be installed. Figure 4 illustrates the workflow in order to execute the DS which consists of four nodes plus as many model nodes as the number of federates in the federation. For example, “Initialize”, “Manager”, “Model-1”, “Model-2”, … “Model-n”, “Execution” and “Collect Results” are all job nodes. Each job node has its own input and output ports. “Initialize” node contains one input port (port 0) and one output port (port 1). This job takes a text file in input port (port 0), containing the list of federates names to be part of the DS federation, as an input and pass it to “Manager” node through output port (port 1). “Manager” node has three inputs (port 0, port 1 and port 2) and one output port (port 3). This node represents the Manager federate that ensures that all federates has joined the federation when starting the execution of the DS. “Manager” node prepares the Manager Federate executable and pass it to the “Execution” job node. “Model-1”, “Model-2”, … “Model-n” job nodes contain two inputs (port 0 and port 1) and an output port each. Users provide the model files (model.tar file) and required data (input.tar file) in the input ports 0 and 1, respectively. “Execution” job node takes all the required files on input ports (port 0, port 1 etc.) in order to run the DS. Finally, “Collect Results” job node collects the final simulation results for further analysis and decision making. This framework also supports scalability. More federates can be added by adding Model job nodes in the workflow.

5 CASE STUDY

To demonstrate the applicability of the proposed framework, we developed a simple federation of an Emergency Medical Services (EMS) model (Anagnostou et al 2013). EMS consists of two main organizations, the Ambulance Services and the Emergency Departments in the area of coverage. Each organization is modeled as an independent federate using the appropriate simulation technique. For example, the Ambulance Services model includes objects that should be able to interact with each other and the environment, namely the call operator should be able to interact with the ambulance crew and allocate an ambulance to an emergency call, the ambulance crew should be able to find the appropriate hospital and select the best route. Therefore, for the Ambulance Service model, Agent-Based Simulation (ABS) was utilized. Emergency Departments usually are process-oriented, where the entities, in this case, patients do not interact and are not aware of the environment; they are just driven by the process through
different activities. Hence, Discrete Event Simulation (DES) was selected for modeling the Emergency Departments. For demonstration purposes, the EMS federation consists of two federates; one Ambulance Service and one Emergency Department. Federates interoperate via RTI implementation that coordinates the data exchange and synchronizes the time in order to maintain causality. Emergency Departments have both ambulance and walk-in arrival points. Ambulance patients are received from the Ambulance Services federate while walk-in patients arrive locally in the Emergency Department federate. Therefore a patient agent from the Ambulance ABS federate is transferred with all its attributes to the Emergency Department DES federate and becomes an entity. Patient arrivals in a hospital affect the availability of resources. Emergency Department resource availability is updated locally in the DES federate. Emergency Department availability is communicated with the ABS federate via the RTI. The Ambulance Services federate uses this information in order to find the appropriate hospital for a patient transfer and therefore it is essential to have the most up-to-date value. The whole federation is developed with open source tools. Federates are developed in the Repast Simphony Toolkit (repast.sourceforge.net) and the RTI in the poRTico IEEE 1516-2010 implementation (www.porticoproject.org).

5.1 Cloud Deployment
To enable the execution of distributed simulation on the cloud, we need to deploy the RTI implementation on the cloud. poRTico, an open source, fully supported, cross-platform RTI which implements the HLA standard IEEE 1516-2010 is deployed on UoW public academic cloud (OpenStack – provided by University of Westminster, UK) supported by the CSSP. These cloud resources only support Linux applications. In order to be able to run the DS on the available cloud resource, we developed a Shell Script executable for poRTico deployment. The other requirement was to have the relevant simulation package, i.e., Repast symphony, as well as their dependencies, namely the Java runtime installation.

5.2 Workflow Creation
The workflow creation is supported by the WS-PGRADE. The first step is to create the workflow using the graph editor, a Java Network Launch Protocol (JNLP) application, which provides drag-and-drop user-friendly environment, where the structure of the workflow can be created. Figure 5 illustrates the graph editor, where the bigger squares denote jobs and the attached smaller squares denotes input and output ports (green and grey squares, respectively). Each port will be denoted by a number. Each job can have more than one input and output. Initialize, Manager, Ambulance, Hospital, Execution and Results are job nodes. Each job node has its own input and output ports. Initialize node will receive the federatelist.txt file, containing the name of federates to join the federation, on the input port (port 0). This file will be forwarded through output port (port 1) of Initialize node to Manager node on its input port (port 0). Manager, Ambulance and Hospital nodes will receive the model files (model.tar) and its parameters (batch_params.xml) for each model/federate. Ambulance node also requires data files (input.tar) and these are received through the input port (port 0). Manager, Ambulance and Hospital nodes will generate the output, which are then fed to the Execution node in its input ports 0, 1 and 2 respectively. Execution node will run the federation and generate the results file which is then transferred to the Results node for further analysis.

5.3 Workflow Configuration
The graphical workflow is configured at the WS-PGRADE web portal (see Figure 6). In each of the job nodes, the first step is to select the type of DCI. As mentioned earlier, CSSP’s cloud resources are managed by CB. CB account authentication is needed in order to be able to access the portal. Then the software or own executable that will execute in the workflow and the executable (shell script file) are selected. From the available cloud resources, once a selection of the instance type is made, an estimation of the job cost is displayed. Initialize.sh executable will be selected for Initialize node which will confirm that the federatelist.txt is received. Manager.sh will be selected for Manager node. Model.sh will be selected for
Ambulance and Hospital modes. Finally the software that will execute the DS in the Execution node (in this case Repast_multi 1.0) and the executable (shell script file) will be selected. The next step is to configure the input and output ports. Initialize node contains one input and one output port. This job takes the text file, containing the list of federates names to be part of the federation (distributed simulation), in the input port and pass this file that is required for Manager job node through the output port. Manager node has three inputs and one output port. This node represents the Manager federate. It will be ensured that all federates have joined the federation during the DS execution. Manager node prepares the Manager Federate executable and passes to the Execution node. Ambulance and Hospital nodes contain three and two inputs respectively, and one output port. Users provide the required model.tar, batch_params.xml and input.tar on the input port 0, 1, and 2 respectively. As explained earlier, model.tar contains the simulation model source code, batch_params.xml contains the input parameters and input.tar contains the data files which required for the execution of the model. Execution job node takes all the required files in the respective input ports in order to run the DS and will export the resulting files from the job run in the output port. The output files can be downloaded after the completion of the job.

![Graph Editor](image.png)

**Figure 5** WS-PGRADE graph editor

### 5.4 Workflow Execution

Execution job node contains three inputs, i.e. Manager, Ambulance and Hospital tar files and produce one output after execution. In this experiment, we run the distributed simulation on the single instance (see Figure 7). OpenStack Nova University of Westminster UoW has the following hardware specification for instance:

- Small (CPU: 1, Cores: 1, Memory: 2 GB)
- Medium (CPU: 1, Cores: 2, Memory: 4 GB)
- Large (CPU: 1, Cores: 4, Memory: 8 GB)
- Extra Large (CPU:1, Cores:8, Memory: 16 GB)

We can run distributed simulation on any of above hardware specification instances by specifying the instance type at the time of job configuration. Multiple CPU cores will improve the performance. In this experiment, distributed simulation federation has three federates. Each federate runs on a different core. So, we will select the instance type which has three or more cores. In our case, we run this simulation on Large instance.

### 5.5 Experiment Results

The speed of the simulation execution depends on the number of federates and simulation time period. In our example, we had three federates running on three different cores, a manager federate and two
Emergency Medical Services (EMS) federates, namely Ambulance federate and Hospital federate. A smaller 24 hour simulation time was selected for our experiment. It was observed that in case of small jobs the preparation time is much higher than the actual simulation execution time. But as the simulation gets bigger or the simulation time is increased then the preparation time will be insignificant. Figure 8 presents a breakdown time of Execution node and illustrates the execution time from the job submission stage till the final stage. It was noted that in our simulation run one third of the total execution time was for the simulation run while the rest was for the preparation and finalising of cloud instance. However, as the simulation time or the complexity of the simulation will increase, the running time percentage will also increase. Although the assembling time and preparation time depends on the availability of the cloud service, but in an ideal situation it will remain approximately similar.

**Figure 6 Workflow Configuration**

**Figure 7 Workflow execution on the cloud**
6 CONCLUSIONS

This paper has presented a framework for running HLA-based distributed simulation using WS-PGRADE workflow on the CloudSME Simulation Platform. By doing so, it hides the deployment complexity of cloud-based distributed simulation from users. Using this framework, we are able to execute varied size federations on the cloud. This framework offers great opportunities for the distributed simulation community to make the technology widely accessible for running large scale distributed simulations. Future work involves performance testing of the framework on different clouds and cloud-based clusters.

ACKNOWLEDGMENTS

Special thanks to Dr. Tamas Kiss and Mr Hannu Visti, Centre for Parallel Computing, University of Westminster for supporting in this study.

REFERENCES


AUTHOR BIOGRAPHIES

NAUMAN RIAZ CHAUDHRY is a PhD candidate at Department of Computer Science, Brunel University, UK. His email address is nauman.chaudhry@brunel.ac.uk.

ATHAR NOUMAN is a PhD candidate at Department of Computer Science, Brunel University, UK. His email address is athar.nouman@brunel.ac.uk.

ANASTASIA ANAGNOSTOU is a Research Fellow at the Department of Computer Science, Brunel University London. Her email address is anastasia.anagnostou@brunel.ac.uk.

SIMON J. E. TAYLOR is the Founder and Chair of the COTS Simulation Package Interoperability Standards Group under SISO. He is the Editor-in-Chief of the UK Operational Research Society’s (ORS) Journal of Simulation and the Simulation Workshop Series. He is Reader in the Department of Computer Science at Brunel University and leads the M&S Group. His email address is simon.taylor@brunel.ac.uk.
ARGUMENTS FOR AND AGAINST THE USE OF MULTIPLE COMPARISON CONTROL IN STOCHASTIC SIMULATION STUDIES

Dr. Thomas Monks
NIHR CLAHRC Wessex
University of Southampton
thomas.monks@soton.ac.uk

Dr. Christine Currie
University of Southampton
Christine.Currie@soton.ac.uk

Dr Kathryn Hoad
University of Warwick
Coventry
Kathryn.Hoad@wbs.ac.uk

ABSTRACT

Pick up any of the standard discrete-event simulation textbooks and you will find that the output analysis section includes a note on multiple comparison control (MCC). These procedures aim to mitigate the problem of inflating the probability of making a single type I error when comparing many simulated scenarios simultaneously. We consider the use of MCC in stochastic simulation studies and present an argument discouraging its use in the classical sense. In particular, we focus on the impracticality of procedures, the benefits of common random numbers and that simulation is very different from empirical studies where MCC has its roots. We then consider in what instances would abandoning MCC altogether be problematic and what alternatives are available. We present an argument for medium to large exploratory studies to move their attention away from classical Type I errors and instead control a subtlety different quantity: the rate of false positives amongst all ‘discoveries’.

Keywords: Output Analysis; Comparison Procedures; Multiple Comparison Control

1 INTRODUCTION

This paper is about making mistakes in stochastic simulation studies. The type of mistake here refers to incorrect interpretation of the stochastic output of a discrete-event simulation (DES) model, as opposed to a mistake in the coding of a model. We focus on mistakes in correctly identifying differences between performance measures in competing configurations (scenarios) of a model when there are several or many alternatives to compare or more formally making a Type I or II error in inference. This is the multiple comparison problem (MCP) that is well known and studied in the empirical sciences. The MCP applies to the plethora of comparison procedures found in the DES literature including more advanced selection procedures. However, most evidence from simulation practice illustrates that basic comparison procedures are dominant (Hoad and Monks 2011; Hoad et al. 2014) and when the MCP is tackled it is done so using multiple comparison control (MCC) advocated in standard DES text books (Hoad and Monks 2011). We consider the use of MCC in stochastic simulation studies in practice and present an argument discouraging its use in the classical sense. In particular, we focus on the impracticality of procedures, the benefits of common random numbers and that simulation is very different from empirical studies where MCC has its roots. We then consider in what instances would abandoning MCC altogether be problematic and what
alternatives are available. We present an argument for medium to large exploratory studies to move their attention away from classical Type I errors and instead control a subtlety different quantity: the rate of false positives amongst all ‘discoveries’.

We present our argument as follows. First we formally define the MCP and how MCC has been approached in DES in both standard comparison and more advanced selection procedures. Second we detail an arguments for and against the use of MCC in practice with final arguments presenting an alternative formulation of the MCP.

2 THE MULTIPLE COMPARISON PROBLEM

At the heart of the classical multiple comparison problem (MCP) lies the Familywise Error Rate (FWER). The FWER represents the probability of making a single type I error (incorrectly rejecting the null hypothesis). For example, when simulating the weekly throughput of two competing configurations (scenarios) of a manufacturing line a Type I error is equivalent to incorrectly concluding that two configurations have different throughputs when in fact they are the same. In stochastic simulation studies an often, but not exclusively, used method of comparison is to construct a 95% confidence interval of the difference between mean throughputs of the two scenarios. In this instance the probability of making a type I error is 5%.

Now consider a manufacturing line where we have five competing scenarios. If we are aiming to choose the best system a simple procedure is to and compare all scenarios against each other in a pairwise fashion. This requires us to construct ten simultaneous confidence intervals. The MCP is the inflation of the FWER when making multiple simultaneous comparisons. The probability of a making a single type I error is given by [1].

\[
P(\text{Making at least one type I error in } m \text{ tests}) = 1 - (1 - \alpha)^m \quad (1)\]

The consequence of [1] is that if we perform ten pairwise comparisons using 95% confidence intervals then the chance of at least one false positive result is 40%. If we perform 100 comparisons the chance of a false positive result is almost certain at 99.4%. Figure 1 illustrates the inflation of the FWER as the number of comparisons ranges between one and 100.

![Figure 1 FWER inflation as the number of comparisons increases. \( \alpha = 0.05 \)](image-url)
3 MULTIPLE COMPARISON CONTROL IN STOCHASTIC SIMULATION STUDIES

We now provide an overview of multiple comparison procedures and control in discrete-event simulation studies. It is worth noting that the use of MCC should depend on the objectives of the simulation study. If the objectives are to optimise a simulation model then selection procedures are more appropriate than MCC. We therefore detail a ‘textbook’ procedure - the Bonferroni Correction; other classical MCC procedures and an overview of selection procedures.

3.1 The Bonferroni Correction

Given the MCP described our standard discrete-event simulation (DES) texts advise on the use of multiple comparison control (MCC) using the Bonferroni Correction to adjust the $\alpha$ level used (see Banks 2005 p449; Law 2006 p537; Robinson 2004 p180). A Bonferroni Correction ensures that the overall probability of making a Type I error remains at level $\alpha$. This is achieved by adjusting the individual $\alpha$ level of the confidence intervals to $\alpha/m$. So in the case of six comparisons an individual $\alpha$ level of $0.05/6$ would be used in each comparison to maintain the FWER. The Bonferroni Correction has the nice property that it makes no assumptions about the independence of scenarios.

3.2 Selection Procedures

The aim of most simulation projects in which the multiple comparison problem arises is to identify the optimal scenario with respect to a particular outcome measure. This can be thought of as optimization via simulation, where the number of alternatives is finite, and the associated methods are often described as selection procedures. When employing simulation optimization it is then necessary to consider the particular algorithm to use and it is also necessary to decide how to decide whether one system is significantly better than the other. As discussed Bonferroni is one way of doing. There are other ways for calculating the probability of correct selection. For example, these can be approximated using the Slepian inequality (Branke et al., 2007) Branke et al. (2007) provide a comparison of different selection procedures for simulation and although they only discuss the situation when the outputs are independent and normally distributed, it still provides an excellent description of the best-known selection procedures: indifference zone (IZ); expected value of information; and optimal computing budget allocation (OCBA).

Of relevance to this article is the discussion of how to measure the quality of a selection procedure. Three measures are worth considering and are discussed in more detail by Branke et al. (2007):

1. Zero-one loss function: $L_{0-1}(D, w) = 1\{w_D \neq w_{[k]}\}$, i.e. the loss-function equals 1 if the wrong selection is made and 0 if the correct selection is made.
2. The opportunity cost: $L_{OC}(D, w) = w_{[k]} - w_D$, which equals 0 if the correct selection is made and is the difference between the optimal value and the selected value if the incorrect selection is made. OCBA procedures are set up to terminate when a selection $D$ has been made and the expected opportunity cost (EOC) associated with this selection $D$, is less than some pre-defined tolerance.
3. In IZ procedures, the measure of performance is the probability of correct selection (PCS) subject to the constraint that the mean of the best solution is at least $\delta > 0$ better than the others. Here, $\delta$ is known as the indifference zone parameter and is defined as the smallest difference in the outputs that the decision-maker believes to be significant.

The focus of research into selection procedures tends to be on improving the efficiency of the sampling whilst ensuring that the relevant quality measure (EOC or PCS) is maintained. Much of the focus of academic research into selection procedures assumes that simulation outputs from different options are independent, which is not the case when common random numbers (CRN) are used; although there are some notable exceptions (e.g. Chick and Inoue 2001; Nakayama 2007; Nelson and Matejcik 1995). The use of CRN induces positive correlation in the outputs of samples coming from different systems. By exploiting this positive correlation, it is possible to reduce the number of simulation...
iterations needed to obtain a set level of confidence, hence the recommendation to use them in all of the standard simulation textbooks. However, it is more difficult to exploit this positive correlation structure in complex selection problems. Nakayama (2007) uses an indifference zone method to show how this can be done and he demonstrates through an example how calculating and taking account of the correlation structure improves the efficiency of the sampling over and above methods in which this correlation structure is ignored, e.g. when using the Bonferroni inequalities.

3.3 Other classical MCC procedures

The Bonferroni Correction detailed in DES text books is not the only MCC procedure available. For a review of alternative procedures see Hoad and Monks (2011). Here we note that the Bonferroni Correction we detail in Section 3.1 is the most conservative approach to controlling the FWER when constructing multiple simultaneous confidence intervals. Hoad and Monks (2011) found that the sequential step-down Bonferroni procedure (Holm, 1979) was less conservative. See Serlin (1993) and Ludbrook (2000) for the procedure to adapt Holm’s original method, based on p-values, to construct simultaneous confidence intervals.

4 AN ARGUMENT AGAINST USE OF MCC

Here we list three arguments against the use of MCC in simulation studies: the practical difficulties in applying the Bonferroni Correction and less conservative FWER MCC to even small scale studies; the use of common random numbers in scenarios; and the differences between empirical and simulation studies.

4.1 The practical difficulties with controlling FWER

A particular practical difficulty arises in medium to large scale studies when attempting to control the FWER. We illustrate this with the Bonferroni Correction procedure outlined in Section 3. Table 1 illustrates the individual confidence intervals needed to maintain a FWER of 5% when comparing up to ten scenarios in a full pairwise fashion. Note how quickly the Bonferroni Correction effects the practical usefulness of results. For instance, beyond five scenarios the individual confidence intervals are above 99.5%. Not only is this incredibly strict, but any useful information of the mean difference between scenarios is lost. Note that even if an overall 10% level of significance was set the individual confidence intervals become extremely strict and difficult to interpret (i.e. approximately 99.5% and above) beyond seven scenarios. It is noteworthy that Hoad and Monks (2011) surveyed 25 simulation practitioners and reported that 36% had conducted one simulation study where the number of scenarios exceeded 100. In these instances using the Bonferroni Correction can lead to many Type II errors.

A further practical difficulty focusing on FWERs is that there is no single statistical definition for what constitutes a family. For example, reconsider our simulation model of a manufacturing line. In addition to throughput we are also interested in comparing the utilization of an expensive resource across the six scenarios. If we are very concerned about FWER then utilization and throughput could be considered as part of the same family. This leads to an individual α level of 0.05/20 = 0.0025. Moreover, what happens if at a later date a further study using the manufacturing line in conducting where new alternatives are tested? Should a modeler consider these new comparisons as part of the same family and hence apply an even stricter per comparison CI? We might even find that previous results that were ‘significant’ are no longer so as we must backwards apply our stricter Bonferroni Correction to the original study.

| Table 1: Feasible and Infeasible use of a Bonferroni Correction in full pairwise comparisons (α=0.05) | 205 |
4.2 Common Random Numbers

In simple terms common (pseudo) random number (CRN) streams assigned to each simulation activity provide a way to synchronize scenarios. The use of the same pseudo random number (PRN) streams induces a positive correlation between the scenarios to compare (see Law, 2006 and Pidd, 2004 for further details). The general purpose of synchronization is as a variance reduction technique (see equation 2) where the induced correlation reduces results in the variance of the difference between scenarios. In addition to variance reduction we argue that CRN greatly reduce the probability of a Type I error.

\[ \text{Var}(s_1 - s_2) = \text{Var}(s_1) + \text{Var}(s_2) - 2\text{Cov}(s_1, s_2) \]  

When CRN work the observed difference between scenarios is extremely likely to be due to the differences in input parameters and not a Type I error due to random sampling ‘noise’. As a simple example, consider a DES model of a single server queue with \( s \) servers illustrated in Figure 2. If we use separate dedicated PRN streams for the inter-arrival and service times and vary \( s \) then we know that each entity has the same arrival time and same service time and can deduce that it is only queuing time that varies between scenarios. Now consider the FWER and Figure 1 with respect to CRNs. If we vary \( s \) so that we conduct 100 comparisons between scenarios with a positive dependency is it valid to conclude that the probability of making a single Type 1 error is inflated to near certainty? If the FWER is a function of random noise between scenarios then it follows that if this noise is removed (or greatly reduced) via CRN then the FWER cannot inflate as is commonly stated. Application of MCC procedures such as the Bonferroni Correction is therefore highly conservative and unnecessary leading to increased Type II errors. This, is however, a simple example where no changes have been made to service or arrival distributions. In section 6 we consider situations where scenarios may not synchronize.

4.3 Simulation is not a classical empirical study

The elephant in the room is, of course, that simulation is quite different from a classical empirical study such as those found in the natural, biological or social sciences. In these cases an experiment is run that simulation modelers might consider a ‘black box’ where the natural, biological or psychological mechanism might not be accessible. In contrast a DES model is a ‘white box’: the internal mechanism – the model logic - is completely accessible and understood to the simulation modeler who created it. As such any unexpected findings can be explored in more detail at the mechanism level i.e. the model can be rerun under the same conditions and additional explanatory information can be exported as part of verification and validation of results.
The ability to rerun a model raises another important difference between classic empirical studies and simulation and in fact a main reason to choose simulation over a real world study (Pidd 2004). That is, that replication, even of models that take several days to run, is relatively cheap compared to real world studies.

Figure 2 Example use of CRN to synchronize scenarios and remove the chance of a type I error

PRN = Pseudo Random Number; scenarios have the same IAT and service distributions

5 AN ARGUMENT FOR THE USE OF MCC

We now consider in what instances would abandoning MCC altogether be problematic and what alternatives are available. We first illustrate an instance where CRN are unsuccessful. We then present an argument for medium to large exploratory studies to move their attention away from classical Type I errors and instead control a subtlety different quantity: the rate of false positives amongst all ‘discoveries’.

5.1 Common random numbers do not guarantee synchronization

In Section 5 we outlined the role that CRNs and synchronized scenarios play in removing the need to control the FWER in stochastic simulation studies. Unfortunately there is no way of guaranteeing that CRN will work. We now consider the case where scenarios may not perfectly synchronize, as there may be differences in service time distributions, inter-arrival rates and entity routing logic; and as such not all difference in a point estimate may be explained by the different input parameter sets. Moreover there may be input parameter sets where CRN fail to work at all, may not be used, or in some cases ‘backfire’ (Law 2006) and increase the variance between scenarios.

As an illustration of what can happen without synchronization we recreate and adapt a simple M/M/S example from Law (2006). We have two competing designs. The first design has a single server and a traffic intensity $\rho = 0.9$ ($\lambda = 1 \text{ min}^{-1}; \mu = 0.9$). The second design has two servers and also a traffic intensity of $\rho = 0.9$ ($\lambda = 1 \text{ min}^{-1}; \mu = 1.8$). Solving this analytically we know that design two has the lower average queuing time. Using CRN and 100 replications leads to only three individual runs where the incorrect decision would have been made, i.e. design one appears superior. On average these three runs had a difference, the wrong direction, of 0.2 minutes. If we force the pseudo random number streams out of sync then the number of individual runs where an incorrect decision would be made rises to 40. On average these 40 runs had a difference of 2.2 minutes. Such an example is often used to illustrate variance reduction i.e. CRN reduce the number of runs required to accurately estimate a mean difference. Of equal importance is that rates of Type I errors will increase without synchronization and an inflation of the FWER can be expected; particularly in cases where multiple simultaneous comparisons are made.
5.2 Reframing the MCP

Thus far we have defined the MCP in terms of controlling the FWER: the probability of making a single Type I error in \( m \) comparisons. The importance of avoiding Type I errors is built on the assumption that the role of the simulation study is to assist a decision maker to pick the best system scenario out of \( n \) scenarios. However, if the role of the simulation study is more exploratory or if the decision maker is equally concerned about Type II errors then the MCP can be reframed in terms of Benjamini and Hochberg (1995) false discovery rate (FDR).

Firstly, let us clarify the difference between FDR and FWER. Consider carrying out \( m \) statistical hypothesis tests. Of these \( m \) individual tests, there are \( m_0 \) true null hypotheses and \( m_1 (= m - m_0) \) false null hypotheses. All the possible outcomes of these \( m \) tests are shown in Table 2 (Benjamini and Hochberg, 1995). Obviously, in reality, only \( m \) and \( r \) (the total number of significant results) are known. The FDR is defined by Benjamini and Hochberg (1995) as the expected proportion of type I errors (\( V \)) among all the significant results (\( r \)), i.e. E\[V/r\]. The FWER, however, is defined as the probability that the number of type I errors (\( V \)) is greater than or equal to one, i.e. P(\( V \geq 1 \)). That is, the FWER refers to the probability of making a single type I error in \( m \) comparisons.

<table>
<thead>
<tr>
<th>Actual result</th>
<th>Decision</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) true</td>
<td>( V )</td>
<td>( m_0 )</td>
</tr>
<tr>
<td>( H_0 ) false</td>
<td>( S )</td>
<td>( m_1 )</td>
</tr>
<tr>
<td>Total</td>
<td>( m - r )</td>
<td>( r )</td>
</tr>
</tbody>
</table>

\( V = \) the number of type I errors; \( T = \) the number of type II errors. Only \( m, r \) and \( m - r \) are observable. \( U, V, T, S \) and \( m_1 \) are unknown.

A useful property of FDR is that it also controls the FWER in the weak sense (Benjamini and Hochberg, 1995). Consider an experiment where \( M \) comparisons between scenarios are made. If all of the \( M \) null hypotheses are true (i.e. there is no evidence to suggest differences between scenarios) then FDR is equivalent to Bonferroni. These claims are backed up by several simulation studies (e.g. Benjamini and Hochberg 1995; Benjamini and Hochberg 2000). It has also been shown that the FDR approach can be a more powerful method than the Bonferroni method (Benjamini and Hochberg, 1995; Hoad and Monks, 2011). Since that first paper many ‘improved’ or otherwise connected methods have been published (e.g. Benjamini and Hochberg 2000; Benjamini and Yekutieli 2001; Genovese and Wasserman 2002; Storey 2002; Storey and Tibshirani 2003; Verhoeven et al. 2005).

This concept of FDR control has gained popularity in a number of disciplines e.g. evolution, ecology, biology, genetics (Benjamini and Yekutieli 2001; Garcia 2004; Garcia 2003; Verhoeven et al. 2005), especially those that are more exploratory in nature, where large numbers of hypotheses are required to be tested, but where the strict control of the FWER can be relaxed (Black 2004). It is this idea of exploration that is particularly compelling. In (DES) experimentation it can be argued that two main objectives exist: to explore the solution space in order to learn more about the ‘important’ factors and solution possibilities and/or to find the optimal solution or ‘best of a subset’ of possible solutions. It is important when your aim is to find the ‘best’ scenario, that the probability of making a type I error, an erroneous discovery, is kept to a minimum. The Bonferroni type methods strictly control the probably of making one or more type I errors, but at a cost of power, hence increasing the risk of overlooking ‘real’ differences between scenarios. It can also be argued that when the main aim is to explore the solution space by comparing large numbers of scenarios, it is important to try to reduce the number of type II errors, while still keeping some control of type I errors. The family of FDR controlling methods achieve this aim. Appendix 1 details the procedure for the basic FDR.
6 CONCLUSION

Pick up any of the standard DES textbooks and you will find that the output analysis section includes a note on MCC. The procedures detailed - most typically the Bonferroni Correction - aim to mitigate the problem of inflating the probability of making a single type I error when comparing many simulated scenarios simultaneously. Use of conservative procedures such as Bonferroni fail to take account of the difference between classical empirical research and simulation. DES studies provide enhanced experimental control, for example through the use of common random numbers, compared to a classical experiment. Moreover, the mechanism studied in the experiment - the model - is accessible and understood by its coder; therefore greatly reducing the chances of accepting an erroneous result at face value. In practice it would appear that such procedures are largely ignored (Hoad and Monks, 2011); however, we argue that it is possible to make a Type I error in a stochastic simulation study albeit at a lower rate than seen in empirical studies. In practice then we propose a reframing of the MCP with the use of FDR based procedures that provide a better trade-off between Type I and II errors.

We acknowledge that we have primarily focused on simple comparison procedures. Further work will consider complex problems that may involve ranking and selection procedures or large complex models.

ACKNOWLEDGMENTS

TM is funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) Wessex. The views expressed in this publication are those of the author and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

APPENDICES

A. FALSE DISCOVERY RATE PROCEDURE

1. Carry out m hypothesis tests (i.e. scenario comparisons) and calculate m corresponding p-values.
2. Rank the p-values in ascending order: p(1) ≤ p(2) ≤ p(3) ≤ … ≤ p(m)
3. Define H(i) as the null hypothesis associated with the p-value p(i)
4. For i = m, m-1, m-2, …, 1, let n be the largest i for which p(i) ≤ iα/m.
5. Reject all null hypotheses from H(1) up to and including H(n) (i.e. reject the null hypothesis corresponding to p(n) as well as all those having smaller p-values).

REFERENCES

Monks, Currie, and Hoad


**AUTHOR BIOGRAPHIES**

**THOMAS MONKS** is a Senior Research Fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. He holds a BSc in Computer Science and Mathematics, MSc in Operational Research and PhD in Simulation Modelling. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His specialty is simulation of healthcare systems of unscheduled and emergency care. He is co-chair of the UKs Simulation Workshop 2016 and co-chair of the UKs simulation special interest group. [http://www.southampton.ac.uk/healthsciences/about/staff/tm3y13.page](http://www.southampton.ac.uk/healthsciences/about/staff/tm3y13.page)

**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 mathematical modelling of epidemics, Bayesian statistics, revenue management, variance reduction methods and optimization of simulation models. See http://www.southampton.ac.uk/maths/about/staff/ccurrie.page for more details.

KATHRYN HOAD is a Senior Teaching Fellow at Warwick Business School, University of Warwick. Katy has research and teaching experience in forecasting, statistical analysis and discrete event simulation. She spent many years as co-organiser of the UK Operational Research Society Special Interest Group in Simulation, and over many years has taken on various roles in the running of conferences such as the UK OR Society Annual Conferences, Winter Simulation Conference and Simulation Workshop 16. http://www.wbs.ac.uk/about/person/kathryn-hoad
ABSTRACT

Stopping rules are widely used in simulation to determine when a sufficient number of samples have been drawn to make a conclusive result. Similar experiments are also carried out in the “real world”, often in a web-setting. In both cases, we may wish to set the number of observations dynamically, rather than in advance of a simulation run or real world experiment. In this article, we describe a technique from the machine learning literature known as Hoeffding Races and use a simple example to compare its effectiveness with a standard simulation algorithm coming from Ranking and Selection. We aim to determine whether there are any advantages that can be gained for simulation practitioners by further investigating the field of machine learning, in addition to being able to say that they “do analytics”.

Keywords: Selection procedures, simulation optimization, machine learning, analytics

1 INTRODUCTION

Simulation optimization or optimization by simulation as it is also described, encompasses a wide range of algorithms. The majority of these have their roots in the statistics literature but have been adapted over the years within the simulation community, to make them more specific to simulation problems. There is always a danger of over-specialisation in academia, not moving outside of the small community of similar researchers and simply refining the standard algorithms. In this paper, we aim to look at work from the machine-learning community to determine if there is the potential to learn something from this field when experimenting with simulation models. In particular, we focus on the situation in which we have a small number of alternative systems to compare and one output on which to make that comparison. Similar experiments are carried out in a web-setting, e.g. when trialling different website designs by directing traffic to two or more different options and measuring their success via hit rates or revenues.

In both simulation and web-based experiments, we may wish to set the number of observations dynamically, rather than in advance of a simulation run or real world experiment. In this article, we describe a technique from the machine learning literature known as Hoeffding Races and compare this with one of the standard ranking and selection algorithms.

In the next section, we will introduce Hoeffding Races and describe Ranking and Selection algorithms. In order to provide some comparison of the different stopping rules, we apply these methods to a simple test, in which we are trying to decide which of five options is optimal. We acknowledge that our tests are very basic at this stage and we use just two different scenarios: distinct difference and vague outputs. Finally, we conclude and suggest some possible areas for future work.
2 Hoeffding Races

Hoeffding races are used in the machine learning literature where much of the research has focused on their use in determining the optimal model structure to use to fit a given data set (Maron and Moore, 1993). The algorithms are based on a result from Hoeffding (1963) who showed that if $X_1, X_2, \ldots, X_n$ are independent random variables and $a_i \leq X_i \leq b_i$ ($i = 1, 2, \ldots, n$) then for $\epsilon > 0$, \[ P\{\overline{X} - \mu \geq \epsilon\} \leq \exp\left[-2n^2\epsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2\right]. \tag{1} \]

We would like to be able to state how confident we are in our estimate of the mean. For example, that the probability that our prediction of the mean value is not within epsilon of the true mean is less than or equal to some pre-specified $\delta$, \[ P\{|\overline{X} - \mu| \geq \epsilon\} \leq \delta. \]

Using Eqn. (1), this can be written as \[ \epsilon = \sqrt{\frac{R^2 \ln\left(\frac{2}{\delta}\right)}{2n}}, \tag{2} \]

where $R$ is the range of the data. When comparing two samples, $Y_1, Y_2, \ldots, Y_m$ and $Z_1, Z_2, \ldots, Z_n$ the inequality can be rewritten as \[ P\{\overline{Y} - \overline{Z} - (\mu_Y - \mu_Z) \geq \epsilon\} \leq \exp\left[-2\epsilon^2 / \left(\frac{1}{m} + \frac{1}{n}\right) R_{YZ}^2\right], \]

where $R_{YZ}$ is the range of the difference between random variables $Y$ and $Z$. For samples of equal length, i.e. when $n = m$, this results in the same expression for $\epsilon$ as given in Eqn. (2). When used to determine the number of samples that should be drawn, such strategies are described as $(\epsilon, \delta)$-stopping rules and show some similarities to the indifference-zones commonly used in ranking and selection algorithms in simulation.

Maron and Moore (1994) used Hoeffding’s result as part of an algorithm for finding a good model for a set of data when searching through a finite set of possible models, so-called Hoeffding Races. The algorithm works by discarding models that have a poor fit to the data early in the process to enable the computational effort to focus on differentiating between the better ones. In this paper, we apply their algorithm to finding the best option, where this is defined by the value of its output rather than the quality of its fit. However, the principle remains the same.

When used to determine the best a set of $p$ options, where the best option is that with the largest output, the Hoeffding Races Algorithm can be written as follows (see Mnih et al. (2008) for a clear description of the algorithm and a comparison with other racing algorithms).

1. Do until $n = N$, the maximum number of iterations or one option remains.
2. Set $n = n + 1$
3. Randomly draw a sample from each of the $p$ options and record the outputs $X_{n1}, X_{n2}, \ldots, X_{np}$ and the new estimates of the expected outputs $\overline{X}_{np} = \frac{1}{n} \sum_{i=1}^{n} X_{ip}$
4. Compute the estimated $\delta/MN$ error, $\epsilon_n$ using Eqn. (2) with $\delta/MN$ substituted for $\delta$, assuming each of the options have the same range $R$.
5. Compute the lower bound and upper bound of each of the $p$ options $LB_{np} = \overline{X}_{np} - \epsilon_n$ and $UB_{np} = \overline{X}_{np} + \epsilon_n$.
6. Determine the optimal option $p^*$, such that $X_{np^*} = \text{Max}\{X_{n1}, X_{n2}, \ldots, X_{np}\}$.
7. For each of the $p$ options, $p$ not equal to $p^*$, eliminate option $p$ if $UB_{np} < LB_{np^*}$.
8. Loop
3 RANKING AND SELECTION METHODS

The equivalent simulation optimization algorithms to the Hoeffding Races are ranking and selection algorithms, as described in (Hong and Nelson, 2009) and (Hong et al., 2015) as the number of options under consideration is small and we make no presumptions about the relationships between them.

A common approach to such problems is the indifference-zone formulation described in Hong et al. (2015). The indifference-zone is defined by two parameters $\alpha$ and $\delta$ such that the best solution will be selected with probability greater than or equal to $1 - \alpha$ providing the difference between the best and next-best solutions is greater than $\delta$.

Branke et al. (2007) carried out a study to compare the effectiveness of ranking and selection procedures when outputs are independent and normally distributed. While the assumption of independence might not always hold in a simulation setting, it is definitely valid for the work considered here. Their findings suggest that OCBA, Linear Loss (LL) and the indifference zone procedure KN++ (developed by Goldsman et al. (2002) based on the original KN algorithm of Kim and Nelson (2001)) all work well to solve the problem of selection, with OCBA and LL perhaps performing the best. Due to time limitations we only consider the KN++ algorithm in this particular paper but acknowledge that future work could extend to providing a full comparison with OCBA and LL.

3.1 Procedure KN++

The particular version of KN++ that we use comes from Branke et al. (2007), which is set up for independent samples. We reproduce their algorithm here with a minor modification to allow for the fact that we are finding the option with the maximum return, rather than that with the minimum return.

1. Do until only one option remains in the set of valid options $I$
2. Set $\gamma = \frac{1}{2} \left[ \left( \frac{2\alpha}{k-1} \right)^{-\frac{1}{2}} - 1 \right]$ and $h^2 = 2\gamma(n-1)$. Update the sample statistics for all of the remaining options in $I$, $\bar{x}_i, \bar{\sigma}_i^2$.
3. For all $i, j \in I$, and $i > j$, set $d_{ij} = \bar{x}_i - \bar{x}_j$ and $\epsilon_{ij} = \max \left\{ 0, \frac{\delta^2}{2n} \left( \frac{h^2}{\bar{\sigma}_i^2 + \bar{\sigma}_j^2} - n \right) \right\}$
4. If $d_{ij} < \epsilon_{ij}$ then stop sampling option $i$. If $d_{ij} < - \epsilon_{ij}$ then stop sampling option $j$.
5. Return remaining system as best.

4 EXPERIMENTATION

In this article, we consider a simple test of the two algorithms where the options are drawn independently from five different normal distributions. The means and variances of these normal distributions are given in Table 1.

<table>
<thead>
<tr>
<th>Option</th>
<th>Mean in Test 1</th>
<th>Variance in Test 1</th>
<th>Mean in Test 2</th>
<th>Variance in Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.25</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.25</td>
<td>3</td>
<td>25</td>
</tr>
</tbody>
</table>
As can be seen from the table, we have two tests, the first being relatively straightforward as each of the distributions is distinct and the second being somewhat harder as the variance associated with each of the options is larger.

5 RESULTS

5.1 Test 1: Distinct Output
Table 2 shows the iteration number at which the two algorithms recommend stopping sampling of each of the first four options. Neither algorithm suggests discarding option 5 before any of the other options.

**Table 2:** Results for Test 1 showing the mean number of iterations needed before the algorithm recommends the option can be discarded based on 10 simulations.

<table>
<thead>
<tr>
<th></th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoeffding Races</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R = 2</td>
<td>6.4</td>
<td>12.2</td>
<td>28</td>
<td>109</td>
</tr>
<tr>
<td>R = 5</td>
<td>42.3</td>
<td>75.9</td>
<td>169</td>
<td>688</td>
</tr>
<tr>
<td>R = 10</td>
<td>174</td>
<td>308</td>
<td>689</td>
<td>2750</td>
</tr>
<tr>
<td>KN++</td>
<td>1.8</td>
<td>1.9</td>
<td>2</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 2 shows that the number of iterations the Hoeffding Races algorithm has a strong dependence on R, the parameter describing the range of the data but even for small values of R, the number of iterations needed before the algorithm finds the best option (option 5) is still significantly higher than for KN++. The large number of iterations needed to draw a conclusion about the outputs is something that Mnih et al. (2006) also mentions, blaming this on the inclusion of the data range in the expression for the stopping criterion.

5.2 Test 2: Vague Output
Table 3 shows the iteration number at which the two algorithms recommend stopping sampling of each of the first four options. In this example, the Hoeffding Races algorithm has a probability of correct selection of 0.8 when the range parameter R = 2. It also has much more difficulty in correctly identifying the right option to discard first. This is understandable because for this example, the standard deviation is equal to 5 and consequently a reasonable range for the generated data would definitely be greater than 2.

**Table 3:** Results for Test 2 showing the number of iterations needed before the algorithm recommends the option can be discarded.

<table>
<thead>
<tr>
<th></th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoeffding Races</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R = 2</td>
<td>2.6</td>
<td>14.7</td>
<td>11.7</td>
<td>114</td>
</tr>
<tr>
<td>R = 5</td>
<td>35.2</td>
<td>71.9</td>
<td>154</td>
<td>615</td>
</tr>
<tr>
<td>R = 10</td>
<td>157</td>
<td>289</td>
<td>666</td>
<td>2890</td>
</tr>
<tr>
<td>KN++</td>
<td>46.3</td>
<td>64.2</td>
<td>104</td>
<td>184</td>
</tr>
</tbody>
</table>

Table 3 shows that the Hoeffding Races algorithm again suggests that more iterations are needed to confirm the optimal solution than KN++ for R = 5 or R = 10. When R = 2, the Hoeffding Races algorithm does suggest a smaller stopping time, but this is at the expense of accuracy.

For a given value of R, the Hoeffding Races algorithm suggests that fewer iterations are needed when the variance of the simulation output is higher. This may seem counter-intuitive but, if implemented correctly, the Hoeffding Races algorithm should use a different value of R dependent on the variance of the
simulation output. Where data are expected to have a higher variance, and hence a greater range, \( R \) should be set higher than in situations where the data have a lower variance.

6 CONCLUSION

The results suggest that the KN++ algorithm coming from the simulation literature performs better on the two sample problems that we have considered than the Hoeffding Races algorithm, producing a more reliable result and generally, less conservative suggestions for the number of iterations that need to be made before discounting an option. Mnih et al. (2008) suggest adjustments to Hoeffding Races in order to take account of the variance of the output rather than the range, which would help make the algorithm less conservative. Whether they can reduce the number of iterations below the number recommended by KN++ should be tested and we leave this work to a subsequent article.

The most obvious reason for the better performance of the KN++ algorithm is that it recalculates the variances of the observations for each of the options in each iteration, allowing it to use a more accurate estimate of the variability. In contrast, the Hoeffding Races algorithm uses the range parameter \( R \) to dictate the width of the bounds on each of the option means and consequently it is unable to react if the variability is different from that predicted by the user.

Given the toy nature of this problem, it is hard to draw general conclusions but it does appear to suggest that the selection procedures used by the simulation community are very efficient and perhaps deserve to be publicized more widely, particularly among data scientists. However, there is still a great deal to learn about the details of their procedures and the particular problems that they encounter, as well as a need to carry out a much more substantial test than that described here.

ACKNOWLEDGMENTS

We are grateful to Weiran Ding for introducing us to this very interesting subject area.

REFERENCES

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 mathematical modelling of epidemics, Bayesian statistics, revenue management, variance reduction methods and optimization of simulation models. See http://www.southampton.ac.uk/maths/about/staff/ccurrie.page for more details.

ANDRIA ELLINA is a PhD researcher in Mathematical Sciences at the University of Southampton, where she is working with Christine Currie and Christophe Mues.
ABSTRACT

This paper focuses on a system dynamics simulation approach modelling the flow of frail elderly patients in health and social care systems in the South of England, UK. A system dynamics model of this pathway is designed that enables testing and informs redesign of the pathway to enhance integrated working between health and social care agencies leading to reduced number of acute hospital admissions and fewer delayed transfers to other health and social care providers. Given the complexity of the model, a number of different scenarios are considered.

Keywords: Simulation, System Dynamics, Healthcare Systems

1 INTRODUCTION

This paper describes a system dynamics simulation study focusing on the flow of frail elderly patients in a health and social care system in the South of England, UK. The model was designed to help stakeholders analyse a number of options to redesign the pathway in order to enhance integrated working between health and social care agencies leading to a reduced number of acute hospital admissions and fewer delayed transfers to other health and social care providers. The project was large and complex, and raises several issues in managing the expectations and understanding of simulation clients in healthcare simulation projects and the study feasibility.

The paper is structured as follows. First, we provide some context to the problem by reviewing the aging population and traditional healthcare improvement approaches. Second, we review previous system dynamics projects in related areas of healthcare. We then provide an overview of the case study, including the model and scenarios investigated. The concluding discussion reflects on the difficulties of conducting a simulation study focused on such a large part of a health and social care system.
2 BACKGROUND

Spending on health and social care has slowed dramatically in the past few years as governments have adjusted public spending in the wake of the impacts of the global financial crisis and the following recession. In the UK, the planned National Health Service (NHS) expenditure for 2015/16 is £116bn in cash terms, which is £3.1bn more than expenditure in 2014/15. Simultaneously, councils have cut £4.6bn from the adult social care budget since 2009-10. Furthermore, they are required to make £500m of service reductions over the next year while facing £600m in extra service demand and price inflation- equivalent to an overall cut in funding of £1.1bn (ADASS 2015). It is expected that the coming half of the century will see an increase in the elderly population due to “the greying of the baby boomers.” (Lanzieri 2011). In the UK, the number of people aged 65 and over is projected to increase from 10.4m in 2011 to 17.79m by 2037 (Eurostat 2011).

There is increasing concern in healthcare organisations throughout the world about the consequences of an aging population and how to manage the increasing complexity of patient needs. Increases in healthcare costs are strongly associated with increasing age and it is widely projected that the healthcare of the elderly will account for an increasing proportion of health budgets in the future. In order to manage future demands and costs, commissioning organisations have started to try and estimate the effect that an aging population will have on future costs and how budgetary priorities will have to change. An important aspect of this is to gain a better understanding of how various health and social care services interact.

Hospitals have a responsibility to ensure that patients are discharged from hospital care in a safe and efficient manner (Preen 2005). This is becoming increasingly important, considering the trend towards shorter hospital stays and more care in the community (Mistiaen et al. 2000). It is evident that cost containment is a major feature of all healthcare systems, especially for acute hospital services (Schwartz and Mendelson 1991). Recent trends include decreasing the length of stay for inpatient care; reducing the number of long stay beds; moving care into the community; an increased use of day surgery; providing increased levels of acute care at home (“hospital at home”); and policies such as discharge planning (Shepperd et al. 2013). It has been shown that discharge planning strategies can reduce unplanned readmission to hospital for some patients (Phillips et al. 2004). The aim of discharge planning is to reduce hospital length of stay and unplanned readmission to hospital, and improve the co-ordination of services following discharge from hospital. Over the past decade, rates of emergency hospital readmission (defined as an unplanned readmission within 28 days of leaving) have risen; the increase has been particularly steep for those over the age of 75. Consequently, reducing the occurrence of emergency hospital readmission for older people is a key issue for healthcare providers including the NHS.

Simulation models create virtual worlds that offer decision makers the capability of conducting experiments and evaluating system interventions (Sterman 2000). They provide low-risk and low-cost platforms to learn and gain understanding about healthcare systems. Such modelling enables decision makers to push the virtual system to extreme conditions, extend the time of observation, and strengthen and relax assumptions, which is often impossible or infeasible in the real world (Sterman 2000). They provide immediate feedback to decision makers, allowing them to gain years of virtual experience and knowledge about the system and interventions by revealing dynamics and mechanisms that are otherwise not obvious (Sterman 2000, Sterman 2006).

3 MODELLING OF LARGE SCALE HEALTHCARE SYSTEMS

Simulation is a modelling tool widely used in operational research (OR), where computer models are deployed to understand and experiment with a system. Two of the most established simulation approaches are discrete-event simulation (DES) and system dynamics (SD) (Pitt et al. 2015). Both DES and SD models are simplified representations of a system developed with a view to understanding its performance over time and to identifying potential means of improvement.

Recently, Young (Young 2005) advanced SD as a method for improving healthcare management. SD combines qualitative and quantitative aspects, and aims to enhance understanding of a system and the
relationships between different system components. Recently, SD has been applied to number of large scale, complex health care systems. For example, Brailsford et al. (Brailsford et al. 2004) report a study of the use of SD to model emergency and on-demand health care in Nottingham, UK. The paper includes a representation of patient flows through different departments in a hospital. The model covers the whole health system from NHS Direct to outpatient clinics and A&E departments.

Wolstenholme et al (Wolstenholme et al. 2004c) utilised SD at a national level to inform government policy on reimbursement for delayed hospital discharges. Their models were parameterised with the available health system data to estimate the effects of various policies such as implementing intermediate care on flow bottlenecks. They additionally assisted local health and social care communities in the UK to interpret and apply national policy frameworks for older people (Wolstenholme et al. 2004a, Wolstenholme et al. 2004b).

SD has successfully been used as a tool in improving patient experience focusing on answering questions about different patient pathways and what might speed or slow patient flow in an acute hospital setting (Lane and Husemann 2007). More recently, Esensoy and Carter (Esensoy and Carter 2015) used SD to model the feedback structure around patient flows between the health system sectors. They specifically focused on patient handovers between health system sectors. The model was kept broad and general as it was not defined to address any specific policy question.

This paper complements this literature by presenting an SD model specifically focusing on the patient pathway of those 65 and over and their handovers between health and social care system sectors.

4 CASE STUDY

In this paper we have focussed on the development of a high level model of the patient discharge pathway from an acute hospital to community based services, focusing on patient flow between various health care systems and social care.

A simulation model focusing on the provision of health and social care for over 65’s across a UK city in the south of England was developed, focusing on understanding provision for the frail elderly and how this could be developed in future.

The initial phase of research was qualitative. This involved the development of a detailed “system map” showing the interconnectedness of provision for this age group. This map was shared with a variety of stakeholders, a process which not only improved the accuracy of the map but also raised their awareness and understanding of care beyond their own areas.

![Simplified Overview of the System](image-url)
Following this, a stock-flow modelling approach was used, where stocks represented accumulations of patients (e.g., waiting for a bed in the hospital; occupying a bed in an acute admission ward, waiting for discharge) and the flows were the admission, transfer, treatment and discharge rates. The model was created using the software package Vensim DSS version 6.3 (Vensim 2015). Stakeholders from the local Clinical Commissioning Group and healthcare provider Trusts were involved throughout the study to design scenarios, so that the model could be used to assist with thinking through the consequences of potential changes.

A simplified overview of the model can be seen in Figure 1, while Figure 2 (in Appendix) depicts the model created in Vensim. There are 12 discharge destinations built-in in the model. These could be grouped into four distinct groups shown in Figure 1. The model is built in multiple layers, these layers hold the influence diagrams for all the flows and stocks represented in Figure 2.

We did not attempt to model intra-hospital flow in detail (blue boxes), including instead the high level stocks of patients waiting for a bed, treatment and, on completion of treatment, waiting for discharge. We have also modelled a series of discharge destinations out of the hospital (green boxes). Patients can transfer between all of these community stocks as well as being admitted to the hospital.

The patient population is dynamic and includes background mortality rates and the arrival of new potential patients as people reach the age of 65. The model includes multiple layers that define the relationship between flows in and out of each stock.

As in all stock-flow systems, the contents of each stock are updated at regular intervals by solving a set of difference equations representing the inflows and outflows from that stock. Vensim presents results in the form of graphs and tables, but most of our output was exported to Excel for analysis and presentation purposes. The outputs include the waiting times; and occupancy rates for each of the stocks in the model.

### 4.1 Overview of the Model

In the model represented in Figure 2 we have focused on patients aged over 65 admitted to hospital and their flow through the healthcare system and social care in the Southampton area. As can be seen, the flows between each stock is controlled by the rate of flow to that stock, as well as its capacity and flows out of that stock. The blue boxes labelled “Waiting –“represent the patients occupying beds in the hospital while waiting for a place at the required discharge destination.

Below we give a detailed explanation of this model:

**Acute Hospital Care:** All the blue boxes are stocks within the hospital. With the exception of the first box (Waiting for Admission), all the boxes use resources such as hospital beds.

**Home and Community Care:** Through consultation with stakeholders, 12 discharge destinations were identified (green boxes) ranging from own home without any care, to community hospital beds and out of area services.

**Admission Avoidance:** Admission avoidance is defined here as a reduction in the proportion of patients admitted to hospital and not necessarily stopping patients from attending A&E.

### 4.2 Data

In general, data requirements are high – 130 individual data items are required in the model. This includes flow of patients to and from healthcare providers, available capacity, along with length of stay metrics. A number of different data sources have been utilised including Secondary User Services (SUS), along with nationally reported data and expert opinion, where data was not available.

As outputs, we are interested in the effect of various changes to the admission rate into the acute treatment unit along with the waiting time post treatment and pre discharge for each of the 12 discharge destinations, and the effect of capacity variations on the above.

To simplify data entry and model testing, all input variables are linked to an external Excel workbook. The outputs are also exported to Excel for analysis purposes.
As the project is still in progress, the data requirements have not fully been met and effort is being made into collection of the remaining data items.

4.3 Validation & Experimentation

The model is created using the “system map” developed in the initial phase of this project which was shared and verified by the stakeholders. Additionally, availability of data would allow for as-is scenario modelling that would further verify accuracy of the model.

Throughout the development of the model various meetings were held with the stakeholders where the model was presented and various possible what-if scenarios were discussed. The following three scenarios were proposed by stakeholders:

1. **Merging of the community crisis and reablement teams.** This scenario analyses the impact on admission and discharge rates from the acute hospital. After a stay in hospital, reablement teams assist patients in maintaining or regaining their ability to live at home. This is done through providing support in developing new ways to manage and to do things differently. Support could be for as little as a few days or up to a maximum of 6 weeks. On the other hand, the aim of the community crisis team is to avoid hospital admissions where possible and assist with hospital discharges from the acute Trust, as well as responding to adults experiencing either health-related or a social care crisis. The potential synergy and the consequent impact on the acute hospital admission and discharge rate was investigated.

2. **Introduction of Integrated Living Services and reduced bed stock in the community.** This includes the extension of rehabilitation and reablement services at home, as well a reduction in the number of beds for these purposes in the community.

3. **A redesigned discharge pathway.** This scenario reduces the number of options on discharge and introduces an integrated discharge team that would handle follow-on care of patients at home rather than in hospital.

4.4 Illustrative Results

As an illustration we have provided below an example where the number of reablement beds is reduced. Let’s assume there are in total 40 beds available and approximately 15% of patients are discharged from the acute hospital unit to such beds in the community. We also assume that the initial occupancy rate is 75%. It can be seen in Figure 3 that when the capacity is at 40 beds, the number of patients waiting for a reablement bed is close to 1 on a daily basis. In this scenario the system reaches an equilibrium at 35 beds, which is equivalent to 87.5% utilisation. Reduction of beds to 30 causes disruption in the system, resulting in an ever increasing list of patients waiting for a reablement bed in the acute treatment unit. To further illustrate the sensitivity of this model to input data, especially the discharge rates from the acute hospital, assume the actual rate of discharge to a bed based reablement facility is 10% instead of 15% in the above

![Illustrative results for reduction of number of reablement beds, given discharge rate of 15%](image)

**Figure 3** Illustrative results for reduction of number of reablement beds, given discharge rate of 15%.
example. Figure 4 shows the corresponding results. As can be seen, unlike the above example, reduction of beds from 40 to 30 does not cause any disruption to the system, as even in the case of 40 beds capacity, the system reaches an equilibrium at around 22 beds, which well below 30.

The purpose of the above example is to illustrate the type of “what-if” scenario analysis that could be performed based on the model, as well as the importance of reliable input data in such analysis.

5 CONCLUSION & DISCUSSION

One of the core assumptions in SD is that the behaviour of the system is due to its structure and not due to external forces or factors. Although SD models can be formulated at many different levels of detail, such models in health care are most traditionally aggregated as in the model above, in the sense that they characterise the population in terms of sizes of subpopulations, rather than at an individual level. In general terms, SD can produce patterns and trends, as well as mean values as outputs from the model. The patterns and trends resulting from simulation experimentation with different policies or strategies (“what-if” questions) can be analysed to inform decision making.

Nonetheless, there are challenges to using and implementing such large and complex dynamic simulation models. Stakeholder involvement, beginning with the project scoping and continuing through the model design and evaluations, is critical to a successful project. However, identifying a representative group of stakeholders to support the broad scope of the modelling, and managing such a large group with different agendas and expectations, can be a challenging task. The same is true when it came to selecting scenarios. Stakeholders lack of experience in SD and macro-level modelling and its capability, along with the natural tendency to focus on low level operational measures instead of higher level overviews of the whole system, proved to be the main obstacles in managing the project as well as the stakeholders.

Furthermore, data requirements to parameterise the model were difficult to fulfil due to lack of access to certain data, costs associated with data acquisition, and data availability. For example, routinely collected data on outflow of patients from the acute hospital is used for financial purposes as opposed to understanding the flow of patients to community providers, and data are not necessarily in a suitable format for modelling. Nevertheless, SD models provide an advantage because their structure will not be limited by the available data and they can be used to do exploratory analyses until the additional data can be incorporated (Pruyt et al. 2014). It may be difficult at times to communicate how these models are built and the details of their mathematical structure. This can sometimes be interpreted by users (i.e., policymakers) as lack of transparency. These structures and sophisticated calculations, however, are necessary to adequately represent the problem and to obtain accurate results.

![Figure 4](image_url) **Illustrative results for reduction of number of reablement beds given discharge rate of 10%.**
ACKNOWLEDGMENTS

This article presents independent research funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Wessex. The views expressed in this publication are those of the author(s) and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

NHS Southampton City CCG funded the initial phase of this research that developed the qualitative map. We are thankful for the input and cooperation of the wide range of professionals who assisted with the development of the System Map and subsequently with the provision of data.

REFERENCES


AUTHOR BIOGRAPHIES

RUDABEH MESKARIAN is a research fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. She holds a BSc in Investment and Financial Risk Management, MSc in Operational Research and PhD in Operational Research concentrating on stochastic optimisation.

MARION PENN is a research fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. She received a BSc (Hons) Mathematics, an MSc in Operational Research and a PhD in Operational Research, with a focus on hospital scheduling.

THOMAS MONKS is a Senior Research Fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. He holds a BSc in Computer Science and Mathematics, MSc in Operational Research and PhD in Simulation Modelling. His specialty is simulation of healthcare systems of unscheduled and emergency care.

MELINDA TAYLOR is a research fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. She holds a BSc in Psychology, MBA focusing on health service management and a PhD in Information Systems. She is also a registered nurse and has clinical and management experience within the NHS and private sector.

JOHANATHAN KLEIN is an Associate Professor of Management Science at Southampton Business School, University of Southampton, UK. Trained as a physicist, his current research interests include problem structuring and qualitative modelling, the nature and use of knowledge in problem-solving practice, and the nature and value of evidence in decision-making.

SALLY C. BRAILSFORD is Professor of Management Science at the University of Southampton, UK. She received a BSc in Mathematics from the University of London, and MSc and PhD in Operational Research from the University of Southampton. Her research interests include simulation modeling methodologies, system dynamics, health service research and disease modeling, and the modeling of human behavior in healthcare systems. She is chair of the European Working Group on OR Applied to Health Services (ORAHS) and is an Editor-in-Chief of the journal Health Systems. She is on the editorial boards of Health Care Management Science, the Journal of Simulation, and Operations Research for Health Care.

PAUL BENSON is a Senior Commissioning Manager at NHS Southampton City Clinical Commissioning Group. He holds a BSc in Biological Sciences and the Diploma of the Institute of Health Services Management. He has held a wide range of roles across the public, private and voluntary healthcare sectors.
Meskarian, Penn, Monks, Taylor, Klein, Brailsford and Benson

in the UK and has had a long-standing interest in “whole-system” approaches to healthcare planning and delivery. He holds an unpaid role as a Governor of Western Sussex Hospitals NHS Foundation Trust.
A APPENDICES

Figure 2 Overview of the model developed in Vensim.
ABSTRACT

This paper develops and explores the simulated responses of the public and interested parties, to the government's proposal to explore the siting of a Geological Disposal Facility (GDF) in their area over a 12 month period. Our objective is to enhance understanding of group and world interactions, and to explore if the formal structuring of public deliberations with different stakeholders could bias the result. This is done using a system dynamics approach to model the flows between different opinions over time for each stakeholder group. The meaning of influence in this context is discussed. Particular interest is paid to the influence of a large environmental concern group on the flow rates between each of these states for the general public. We conclude by discussing some planned short term extensions which will refine and validate the model's structure and defining values.

Keywords: System Dynamics, Influence, Public Participation

1 INTRODUCTION

Over the last few decades the government has devoted considerable attention to creating a more long-term solution for the disposal of nuclear waste. They have decided that it would be most suitable to use an underground geological disposal facility (GDF), planned to be sited somewhere within the UK\(^1\). Their recent attempt at siting a GDF in Cumbria became unsuccessful when the Cumbrian County Council withdrew their interest in 2013. This left no remaining sites to host the project\(^2\). To try to avoid this impasse in future they will be attempting to repeat the siting process with a more significant focus on public participation, a strategy that has had success in other countries when siting a GDF (e.g. Canada and Sweden\(^3\)).

---


\(^2\) http://www.bbc.co.uk/news/uk-england-cumbria-21161367

\(^3\) http://www.nda.gov.uk/publication/geological-disposal-overview-of-international-siting-processes/
The Government is openly committed to volunteerism and public engagement during the siting process for the geological disposal of nuclear waste. When the process moves to identification of potential sites for a GDF, the government intends an open debate to occur between all parties about facility siting concerns and to reach conclusions that reflect the true preferences of those involved. This debate may be conducted in many ways and involve many participatory instruments: e.g. interactive websites and forums, social media, stakeholder workshops, citizen juries and town hall meetings. Some of these, such as town hall meetings, will be formal and established by Radioactive Waste Management (RWM); others, such as social media discussions will be informal and created by the participants themselves. The government would not want the portfolio of the formal instruments and their scheduling in the debate to affect the outcome. One main intention behind our modelling is to investigate whether any unanticipated biases might inadvertently enter this deliberation process.

With the focus heavily shifting to public participation there is more concern with how to structure the process to assess and reflect public opinion. In particular we might ask if different formal structures would alter the strengths of opinions of stakeholder groups. Also we might ask how significant groups within the community might affect public support. These are questions we hope to be able to explore through our modelling and analysis. By experimenting with different structures of the process, changing interactions between parties and different world conditions we can explore this space. In particular, we can investigate the influence of a particular system structure or variable has on the overall opinion of the GDF proposal. For more specific information on the problem, see the 2014 White Paper ‘Implementing Geological Disposal’

4 Specifically, the purpose of our model is to enhance understanding of the important interactions underlying the described process whilst allowing the identification of the most influential groups or sections of the model.

We have chosen a system dynamics approach (Sterman J, 2000) to model public participation because there are many feedback loops and dependencies, which particularly fit this modelling paradigm. For example, such methods can capture simple hypotheses such as the more support the GDF has, the more pressure is put on those against the proposal to change their minds. Another example is the hypothesis that celebrity opinions are able to change public opinion substantially, e.g. if a celebrity scientist such as Professor Brian Cox states an opinion on Twitter then many may follow him. We will be modelling public opinion by using stocks to count the number of people in support of the GDF proposal, and the flows to define how quickly people change their opinions over time.

There are many stakeholder groups that we could consider including in the model. However for our more basic preliminary model we will only include 4; the local government, general public, local industry and a large environmental concern group (ECG). Our model will then be used to analyse the shifting opinions of the parties which are being modelled. However the main focus of the model will be to provide a measure of influence of one group over the behaviour of another group, and to assess how that influence could be affected by different public deliberation structures. In particular we focus on the influence of the ECG on the general public. We will do this by applying a one-out method that stemming from Cook’s distance in regression (Cook R D and Weisberg S (1982)) to evaluate changes in flow rates as we remove the ECG stakeholder group from the process.

2 A SYSTEM DYNAMICS APPROACH

2.1 Model Details

Our System Dynamics model will be produced using the iThink software package, which was selected after comparing several other modelling packages compatible with System Dynamics. Within iThink, we will
split the model into various ‘sub-models’ called modules, which are depicted as squares with rounded edges in diagrams of the process.

**Figure 1** Top-Level display of the model from iThink. Directed red lines indicate interactions between modules.

For simplicity our initial model will assume there are 4 main groups of interest (local government, local industry, general public and the ECG). These are each described by two modules, for example; ‘General Public’ which holds the stocks representing current public opinion on the proposal and ‘Public Variables’ which holds the information used to calculate the flows between stocks in the ‘General Public’ module. The model is currently split into 9 modules. The simplest modules are ‘World Conditions’, which stores most general constants used by the model and certain time-dependent variables, such as ‘Radiation Incidents’, and the ‘Results’ module, which collects information about the current state of the stakeholder groups.

The next module to consider is the ‘General Variables’ module, which is where general feedback and interactions are included. This module is used by all groups in a similar way (although slight group specific changes are added in the group variable modules). The converters contained here are used, alongside any group specific variables or interactions, to calculate the flow rates from the stocks in the group variable modules (e.g. ‘Public Variables’). Figure 2 shows the structure of the stocks and flows of the ‘General Public’ module below.
2.2 Assumptions

We need to make certain simplifying assumptions of the real world scenario. Here we go through the general assumptions that have been made when constructing the model. Other group or variable specific assumptions have been made for the interactions of the model which we do not have the space to discuss fully in this paper.

- **Assumption 1:** The views of all people in reference to the proposal can be described by exactly one of the five following states; ‘strong positive’, ‘positive’, ‘neutral’, ‘negative’ and ‘strong negative’.
- **Assumption 2:** When individuals change their views, they must do so between the adjacent states. For example an individual in the ‘Positive’ state can only change states to ‘Strong Positive’ or ‘Neutral’, however they cannot go straight to ‘Negative’ or ‘Strong Negative’.

Figure 2

‘General Public’ module. This uses information from the ‘Public Variables’ module to calculate the flows.
Gilbert, French, and Smith

- Assumption 3: All individuals are contained in one of the defined groups of the model. For example ‘Public’, ‘ECG’ or ‘Local Government’ for the basic model.
- Assumption 4: An individual cannot change between defined groups mid-simulation (for example a retiring government member).
- Assumption 5: All individuals in the same defined group carry the same values and preferences about the proposal.
- Assumption 6: There is no random error in reporting the support for or against the proposal (although it may be time-lagged).

It is clear that some of these assumptions are substantial simplifications made to allow for the preliminary model to be developed; however we expect to relax these over time (in particular Assumptions 3, 4 and 6). For example Assumption 5 at the moment seems quite unlikely, however when we start to include more groups it will become closer to reality. Assumption 2 on the other hand is included to allow for slightly a more simplified model. In particular it enables us to only consider flow rates between adjacent states. This greatly reduces the number of flows we must define.

2.3 Model Construction

Due to the size and complexity of this model, we are unable to go into much detail of individual interactions for each of the stocks, flows and converters. However we will give an overview of the general trends that the model has been built around. Keep in mind that the values used are not based on real data (although we are currently looking at modeling another scenario where we have data to get a better idea of the scale of the flows) but are instead designed around a set of relative importances.

2.3.1 Flow Rates

Each group that has been modeled has 4 primary flow rates associated with them; ‘Strong Positive’, ‘Positive’, ‘Negative’ and ‘Strong Negative’. These flow rates are scaled according to the total population size. These 4 primary flow rates are used in the main group module, for example ‘General Public’ seen in Figure 2. In most cases the base values have been used, with the exception of arriving into the ‘Neutral' stock, where the flow rates have been scaled by a converter representing the trust that the group has in Government.

The most interesting point about these primary flow rates is that they have been defined analogously to the way a linear utility function might be defined. Each group is given the relative importances of 4 main attributes; ‘Perceived Benefit’, ‘Perceived Risk’, ‘Ideological and Political Attitudes' and ‘Familiarity with Nuclear Industry’. These attributes have then been scaled according to their relative importance and combined into the 4 primary flow rates. The only exception to this is the two strong flow rates which have a slightly different relative importance set-up. This tends to favour evidence rather than popular opinion.

2.3.2 Attributes

Given how the primary flow rates have been calculated, the definition and weighting of our 4 main attributes is clearly important to how the groups will respond to different scenarios. These are, in the majority of cases, defined as a linear combination of other variables of interest, for example through word of mouth, media, government trust and predicted environmental damage. The relative weightings of these variables, and of the attributes themselves, were defined according to the relative importance relationships derived from the literature. (e.g. Kim et al, 2013; Tokushige K et al, 2007; Kunreuther H et al, 1990; Jenkins-Smith H C et al, 2010; Sjöberg L and Drottz-Sjöberg B-M, 2001)

The most consistent of these relationships was that ‘Perceived Risk’ is more important than ‘Perceived Benefits’ (e.g. Tanaka Y, 2004). Alongside this, trust in the government and experts clearly played a key part, particularly in the assessment of benefit perception (e.g. Siegrist M, 2008). The group's familiarity with the nuclear industry and their ideological attitudes tended to play a much smaller role, but one of the
key components to reducing the perceived risk of a group was strong risk mitigation plans that the group can trust (e.g. Kunreuther H et al, 1990; Sjöberg L and Drottz-Sjöberg B-M, 2001).

2.3.3 Stocks

Each stakeholder group its opinions represented by 5 stocks. ‘Strong Positive' through to ‘Strong Negative' as seen in Figure 2. These represent the different states of mind that individuals of the group can have about the proposal (based on the categorisation used by the MRWS Partnership). The initial population of the group is divided between the ‘Positive' and ‘Negative' state according to an adjustable variable representing the initial support for the proposal. The difference between a negative and positive state should be relatively clear, while the difference between, for example, ‘Positive' and ‘Strong Positive' is that an individual in the ‘Strong Positive' state believes they have sufficient evidence to back up their support of the proposal. The flow rates to these stronger states have been developed to reflect this, placing a stronger emphasis on evidence and facts to get into or out of these states.

While the model is currently too complex to assess all feedback loops associated with the stocks (we are currently working on solving this complexity issue), the dominant feedback loops are all self-reinforcement loops of increasing (decreasing) support resulting in more positive (negative) media reporting and word of mouth effects. The other dominant feedback loop is between the positive (negative) states of the general public and the local government, which are tied together particularly on the government's side.

2.4 Influence in System Dynamics

While we have a functional model that can be used to predict acceptance rates of different groups over the 12 month period, we are also interested in assessing the influence of the ECG over the general public's behaviour during the simulation. But to move forward we must first attempt to explicitly define what we might mean by influence. Since such measures of influence in group negotiations have so far been neglected in the literature.

**Definition:** When group 1 is influential at level $\delta$ over another group 2, then the behaviour (as measured by the flow rates of the group) of group 2 changes by at least $\delta$ when group 1 has been removed from the discussion. This behavioural change will be measured by some distance measure $D(.,.)$ applied to the flows of group 2 before and after the removal of group 1.

We also need to know how we will measure a change in behaviour to assess against our level $\delta$. This will be done by using the Euclidean Distance to measure the size of the change of the flows associated with group 2 when group 1 is removed. We have chosen the Euclidean Distance here to serve as a starting point. Therefore, when we say group 1 is influential over group 2, we mean that group 1 causes noticeable changes in the flows of group 2. There will be several types of influence associated with this which we will introduce shortly.

As we have a time series for each of the flows, we should also be able to go into more detail on the influence on specific flows. For example being able to state that the removal of the ECG group has reduced the negative flows, but had relatively little impact on the positive flows on the ‘General Public' group, which may be useful for users of the system to know. To explore this, we will consider 3 influence measures on different levels of the process.

2.4.1 Influence over a flow of a group

The first influence measure will be used for most of our analysis. We use a distance measure (Euclidean Distance as a preliminary metric) to compute how much the flow rate (of a specific flow) changes between the two scenarios we are considering (the process including or excluding the ECG). This can be considered

---

http://www.westcumbriamrws.org.uk/documents/Awareness_Tracking_Survey_1.pdf
the main building block for use in the following two influence measures. This will give us one value for each time point of the simulation, showing how much the flow rate has changed at that particular time. These values will then be aggregated and scaled to give us a single value which describes how the removal of the ECG has affected that particular flow of that particular group.

2.4.2 Influence over a group

The next two influence measures are very similar, the difference between them is the area of the model feeling the influence. For the first case, we consider the influence of the removal of the ECG over a specific group. This involves averaging the influence over the 4 flow types of that group (strong positive to strong negative), resulting in a single value that describes the influence the removal of the ECG has on that specific group. We could use this to explore which groups could be strongly related to one another.

2.4.3 Influence over a flow

Similarly to the last influence measure, we consider influence over a specific flow. Again, we will average the influence over a specific flow type over each of the groups, resulting in a single value that describes the influence the removal of the ECG has over a particular flow. This could be useful analysing if the group being removed (ECG in this case) tends to change a particular flow type. For example, we may expect the ECG to increase the negative flow rates.

3 PRELIMINARY RESULTS

Our main objective here was to assess the impact that removing the ECG would have on the behaviours of the other parties involved. The influence measures given in the last section was applied to our preliminary model where an ECG was considered with 80 members. In this model, the general public has 500 members and the local government had 50 members. The number of time points considered is 25, to allow for a comparison to be made each half month, for the full 12 months. The influence measures were then applied to the flow rates at each of these time points according to Section 2.3 and the results of this have been given in the table below.

Table 1 Table showing the comparison of the two cases where an ECG has been included and excluded in the model. Each cell contains the influence value given from using the influence measures in Section 2.3.

<table>
<thead>
<tr>
<th></th>
<th>ECG</th>
<th>General Public</th>
<th>Local Government</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong Positive Flow</td>
<td>0.0713</td>
<td>0.3342</td>
<td>0.2028</td>
<td></td>
</tr>
<tr>
<td>Positive Flow</td>
<td>0.1204</td>
<td>0.2998</td>
<td>0.2101</td>
<td></td>
</tr>
<tr>
<td>Negative Flow</td>
<td>0.3136</td>
<td>0.4352</td>
<td>0.3744</td>
<td></td>
</tr>
<tr>
<td>Strong Negative Flow</td>
<td>0.1699</td>
<td>0.1793</td>
<td>0.1746</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.1688</td>
<td>0.3121</td>
<td>0.2405</td>
<td></td>
</tr>
</tbody>
</table>

Keeping in mind that each of the influence values above is an ‘average’ influence over the time points considered, it is clear that the ‘Negative Flow’ has been the most affected by the removal of the ECG, seeing the largest influence value and average distance for both groups. Also, the flow rates for the Local Government has been impacted far more than for the General Public due to having more direct communication with the ECG. It is worth noting that the removal of the ECG decreased the scaled flow rates in all cases meaning that the change of opinion had slowed. However the Negative flow rate had decreased far more than any others, resulting in relatively less people entering the negative states. Another interesting feature is that the removal of the ECG has affected the negative rate noticeably more than the Strong Negative rate, largely due to the evidence required to move an individual from the ‘Neutral’ to
‘Negative’ states being quite different to moving an individual from the ‘Negative’ to ‘Strong Negative’ states.

![Influence of ECG removal on General Public over Time](image)

**Figure 3** Time series showing the influence values of the ECG removal on the flow rates of the general public at each time point considered. Influence values are calculated according to Section 2.3 from the difference in flow rates. Note that values are often small but non-zero.

Figure 3 gives us more insight about the time intervals for which the removal of the ECG has most impacted the flow rates of the General Public. Here we can visualise the time points that have contributed the most to the calculation of influence value shown in Table 1. Immediately it is clear that the majority of the influence of the flow rates comes between months 2 and 5 for all flows. It is between these months that the ECG releases an information pack to the public to reduce support (this is the only information pack released during the simulation), and so the removal of the ECG caused this pack to never be released. Using both Figure 3 and Table 1 allows us to better understand where groups are exerting influence, and what is the driving force behind this influence.

### 3.1 Model Validation

To assess the model's suitability for its intended purpose (to develop understanding of the influential interactions and provide a basis to compare the influence of a change in scenario over each group's behaviour) we applied several validation tests (Forrester J W and Senge P, 1980; Barlas Y, 1994; Khazanchi D, 1996; Martis M S, 2006; Kampmann C E and Oliva R, 2009). These included extreme-conditions tests for the adjustable parameters, time-step sensitivity and structural verification against the known format of the last proposal attempt in Cumbria (MRWS Partnership). We are also in the process of calibrating the model to past data through the construction of a second model (which we also hope to use to simplify the model to better identify feedback loops). Because the model is currently very complex the feedback loops in the system are difficult to analyse, although this issue should be resolved in the near future. Typically
the feedback loops involved are self-reinforcing, with a minor reinforcing effect on other groups of the same state. The main exception to this is that the support shown within the general public has a strong reinforcing effect on itself through influencing the support of the local government.

Our model is still developing: there are some extra features in the structure that are being built into the model (for example additional groups) which will help to address the plausibility of Assumptions 3 and 4 in particular. We will continue to revisit these validation tests to ensure that any new additions do not compromise the integrity of the model structure. Finally, due to the lack of data, we will be performing a structural and behavioural face-validity test once the model has been further developed.

4 EXTENSIONS

What has been shown in this paper has been a basic preliminary model. We are currently building a second model to better understand the interactions between different stakeholder groups that is based on the data gathered by the MRWS Partnership during their last attempt at siting a GDF. Preliminary explorations of this model suggest that it is able to recreate the behaviours observed throughout the MRWS partnership activity and predict the actual voting outcome. We plan to use insights gained from the second model to refine the interactions and values in the model described in Section 2. Ideally we will expand our model to have enough stakeholder groups such that nearly all people that are modeled can be described relatively well by one of the groups, but having few enough groups so that the analysis of group influence is still meaningful and the interactions between them can still be explained relatively easily. We have also started exploring different formal deliberation structures, where we have certain stakeholder groups have consistent meetings.

The longer term extension however is to compare the structure and results of this model to a probabilistic model of the same scenario, for example using agent-based modeling through a queueing system. These two models would be explored individually, and then contrasted on the strengths and weaknesses of each, and which is more useful in practice. For example, it may be easier in some cases to elicit probabilities from experts about changes in state of mind than eliciting some of the more exact numerical values required by the deterministic system, while the deterministic system could be far easier to explain and pitch to those outside of academia. In particular we will be looking at which model is the most useful for measuring the influence of the different stakeholder groups.

5 CONCLUSION

This paper set out to provide a measure of how influential an ECG may be over the public's behaviour when they are removed or included in the discussions about the proposal to build a GDF in the area. We can see from our results that we can identify some more general expected behavioural changes of the groups involved when one has been removed. However more importantly, we can begin to make predictions of specific events or responses which make these groups so influential over others.

While the removal of an entire ECG from the process may not be realistically possible, this method could still be used to compare different scenarios of interest. For example we may be interested in the influence of world events or restrictions on the media over the different stakeholder groups, so we are not limited to removing entire groups. This could be used for two primary purposes, the first would be to explore the scenario to develop better understanding of the stakeholder groups involved, and how they respond to changing world conditions and the various feedback loops they are involved in. Secondly it could help in event planning, to explore any possible bias that could be introduced through different debate structures, and how much they could affect the support of the proposal. The results have been promising overall, and exploring the idea of influence is a novel addition to System Dynamics.

---

6 http://www.westcumbriamrws.org.uk
REFERENCES


AUTHOR BIOGRAPHIES

MATTHEW GILBERT received a MSc of Mathematics and Statistics from the University of Warwick in 2013. He is currently in his third year of his PhD at the same university studying under Prof. French and Prof. Smith. He is funded through an EPSRC Studentship.

SIMON FRENCH is a professor in the Department of Statistics at the University of Warwick and has had a long career in modelling risks and decisions in nuclear energy. He has also worked extensively on risk communication and stakeholder engagement.

JIM Q SMITH has been a Professor of Statistics at the University of Warwick for about 20 years and has published widely in Statistics, Decision Analysis and Artificial Intelligence. He has a particular interest in rational and probabilistic decision making in complex environments including those with multiple agents.
SYSTEM DYNAMICS’ POTENTIAL IN DEVELOPING SAFETY IN COMPLEX SYSTEMS–LITERATURE REVIEW AND A PILOT STUDY FOR MEDICATION SAFETY IN A HOSPITAL PHARMACY

Mohammed Ibrahim-Shire
Loughborough University
Epinal Way, LE11 3TU
Loughborough
m.ibrahim-shire@lboro.ac.uk

Dr. Gyuchan-Thomas Jun
Loughborough University
Epinal Way, LE11 3TU
Loughborough
g.jun@lboro.ac.uk

Professor Stewart Robinson
Loughborough University
Epinal Way, LE11 3TU
Loughborough
S.L.Robinson@lboro.ac.uk

ABSTRACT
System dynamics modelling includes a set of conceptual and numerical methods that are used to understand the structure and behaviour of complex systems. A system dynamics model represents the causal relationships, feedback loops, and delays that are thought to generate the system behaviour. Its full potential as a method to understand and develop safety in complex dynamic systems has not been fully identified. A systematic approach-based literature was conducted to identify studies employing SD applications in safety-critical domains. Thirty-seven studies were included and classified based on a customised human factors safety taxonomy framework. Results identified several gaps with a focus on healthcare. It also provides the rationale for an on-going pilot study applying SD to simulate the impact of staff workload on medication safety in a hospital pharmacy.

Keywords: System Dynamics, Literature Review, Simulation Modelling, Safety

1 INTRODUCTION
System dynamics is an analytical modelling methodology, its origins of which are attributed to Forrester (1961) in his pioneering work on “industrial dynamics” in the 1960s. Today, SD methodology is used beyond the industrial setting and has been applied in many different fields of study including healthcare. SD combines both qualitative and quantitative aspects and aims to enhance understanding of complex systems, to gain insights into system behaviour. The qualitative aspect entails the construction of “causal maps” or “influence diagrams” in which the system structure and the interrelations between the components of a system are explored. The quantitative aspect entails the development of a computer model in which flows of material or information around the system are modelled and bottlenecks identified. Such models can then be used in a “what if” mode to experiment with alternative configurations, flows, and resources.

Whilst SD modelling has gained popularity as a tool in a variety of industries such as engineering, economics, defence, ecology and business (Homer & Hirsch, 2006), its potential has not yet been fully realised as a tool for understanding trade-offs between safety and efficiency and making strategic decisions.
in safety-critical industries. There are a considerable number of SD applications on safety (e.g. patient safety, traffic safety, nuclear safety, etc.) in various fields but no comprehensive systematic approach review of the use of SD modelling in safety-critical domains has been published. The overall aim of our study is to evaluate the extent, quality and value of system dynamics applications in safety-critical domains, classify into a safety taxonomy framework and identify the gaps. To this end, this paper focuses on how SD modelling has been used in safety-critical environments as described in the literature and the gaps revealed by it.

2 METHODS

2.1 Literature Search

Systematic approach methods were employed to gather and evaluate relevant papers for this literature review. A range of databases were searched for published articles up to July 2015 on system dynamics and safety. The databases include PubMed, Web of Science, Science Direct and Google Scholar databases. The grey literature was also searched using Google search with key terms mentioned below but eventually excluded to manage the scope of the review and time constraints. Papers eligible for inclusion were those that described applications of system dynamics modelling to understand/improve system safety. Specific key terms that were searched included 'system dynamics', 'safety', 'accident', 'errors'. The reference lists of each article were reviewed to identify additional resources. Articles were selected on the basis of their thematic and content relevance to the inquiry. Thirty-seven papers were identified that reported findings that applied SD modelling in safety.

2.2 Analysis – Safety Taxonomy

We have looked at the four most popular safety frameworks (FRAM, AcciMap, HFACS and STAMP) to establish a safety taxonomy framework and found HFACS to be the most suitable. We adopted the combination of the Human Factors Analysis and Classification System (HFACS) framework and Rasmussen’s risk management framework (Rasmussen, 1997) to identify and classify the SD applications in the selected articles. The HFACS comes equipped with its own taxonomy to classify and analyse human error and accident causations, but lacks a crucial tier that is equivalent to the government tier in Rasmussen’s six-levels of risk management framework. A new tier was introduced therefore changing the original HFACS framework into an extended HFACS framework with an addition of a new tier called External Factors as shown in the first row of Table 2. This encompasses regulatory, social, political, environmental, and economic influences and will allow the categorisation of studies that cannot be fitted in the standard tiers.

3 RESULTS

The results of the literature search are presented in Tables 1 and 2 which show the 37 papers we identified and classified according to the extended HFACS framework. The 37 papers were concentrated in the fields of aviation, construction, disaster-prevention, industrial systems, drugs and terrorism, government, healthcare, military, nuclear and traffic. Healthcare topped the list with a total of 11 papers that applied SD to improve a safety-critical aspect. Qualitative SD is based on creating casual loop diagrams and using these to explore and analyse the system. Quantitative SD is based on quantitative computer simulation modelling using purpose built software.

In terms of which modelling aspect of SD was used, studies applying only qualitative SD (causal-loop diagrams) were 11, whilst those applying quantitative SD (stock-and-flow diagrams) were 14. Lastly, the total number of papers applying both casual-loop and stock-and-flow diagrams numbered 12.
Authors, for instance, have improved modelling system safety problems through the application of qualitative and quantitative SD. Amongst them are enhancing healthcare safety, through estimating potential outcomes, analysing reasons other than cost on why systems safety is failing, to discussing bottlenecks in critical services. Authors have also improved safety through calling for greater decision-making by basing it on system analysis, analysing past behavioural events in modelling structure to plan effective safety policies as well as looking at a holistic approach to analyse beyond human error in accidents. These examples provide a clear indication of how, through the effective application of SD, safety can be improved in safety-critical industries.

In Table 1, the outcome of each applied SD application is categorised in three common categories, namely: strategic problem solving, consensus building and policy changes. These three categories reveal the most intended outcomes of virtually implemented SD applications. The most repeated theme from the literature was focused on strategic problem solving as the intended outcome.

In Table 2, the thematic content of each paper is classified according to its primary foci (highlighted in dark grey) and its secondary foci (highlighted in light grey). Primary foci are identified as the strong themes of the paper, whilst secondary foci are identified as visible, but not central themes in the papers. The literature review on existing SD applications to system safety indicated that most of the literature concentrated on improving safety in the higher tiers of the hierarchy whilst a few studies have been dedicated at the operator end or the lower tiers. This is true for applied SD applications in the healthcare sector as illustrated in Table 3. Furthermore, existing literature have been predominately modelled using hard variables. There is a clear gap in System Dynamics modelling practice to combine social intangible variables with hard variables in order to produce meaningful and reliable results to aid decision-making. Soft variables have the ability embody the human element in problems and can greatly enhance the total understanding of the holistic output of the simulated model (McLucas, 2003).

4 PILOT STUDY

To contribute to the existing literature in applied SD modelling in safety-critical domains with a focus on healthcare, we are currently exploring a case study in the University Hospitals of Leicester (UHL) pharmacy where we will analyse how SD approach can assist hospital pharmacy staff management by modelling and simulating the impact of staff workload on medication errors in various scenarios. High pressure workload is currently a problem for many hospital pharmacists. For a number of years now, changes to the pharmacist’s role, pressures to meet targets, staff shortages, long working days with no opportunities for rest breaks, and an increasing administrative burden have left pharmacists struggling to cope with rapidly developing workloads and an increase in dispensing errors. This led to concerns that patient safety is being compromised. Berwick’s review into patient safety (Berwick, 2003) crucially highlighted the urgent need for developing methods and guidance for staffing ratios based on dynamic understanding of staff workload and systematic approach.

The study will primarily adopt the participatory group model building methodology which will involve participants directly in model building and analysis. A collective model qualitative model derived from the participants’ mental models will subsequently be converted to a quantitative stock-and-flow diagrams. The model data will provide a platform to look at the feasibility applying SD in terms of data, time and knowledge requirements and the utility which encompasses the decisions and learning outcomes. This in turn will allow existing hospital pharmacy management to view and decide the staff workload management issue by supporting them to better consider the impact of staff workload on cost and safety.

Additionally, the output model will place a focus on combing soft and hard variables to account for the behaviour over time and produce a snapshot of reality.
**Table 1 Outcome of Applied SD Applications in Safety-Critical Domains**

<table>
<thead>
<tr>
<th>Study</th>
<th>Theme</th>
<th>Sector</th>
<th>SD Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson &amp; Anderson, 1994</td>
<td>Policy changes</td>
<td>Healthcare</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Bouloiz, H. et al., 2013</td>
<td>strategic problem solving</td>
<td>Industry</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Carhart, N.J., 2009</td>
<td>strategic problem solving</td>
<td>Nuclear</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Cooke, D.L., 2003</td>
<td>consensus building</td>
<td>Mining</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Ellis, B.Y.R.E., 2004</td>
<td>strategic problem solving</td>
<td>Drug and Terrorism</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Goh, Y.M., Love, P.E.D., et al., 2012a</td>
<td>consensus building</td>
<td>Mining</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Goh, Y.M., Love, P.E.D., et al., 2012b</td>
<td>policy changes</td>
<td>Mining</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Guo, S., Roudsari, A. &amp; Garcez, A., 2013</td>
<td>strategic problem solving</td>
<td>Healthcare</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Han, S. et al., 2014</td>
<td>strategic problem solving</td>
<td>Construction</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Homer, J.B., 1984</td>
<td>strategic problem solving</td>
<td>Healthcare</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Jiang, Z. et al., 2015</td>
<td>policy changes</td>
<td>Construction</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Kontogiannis, T., 2011</td>
<td>strategic problem solving</td>
<td>Healthcare</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Lane, D.C., Monefeldt, C. &amp; Rosenhead, J. V, 2000</td>
<td>policy changes</td>
<td>Healthcare</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Lattimer, V. et al., 2004</td>
<td>strategic problem solving</td>
<td>Healthcare</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Leveson, Couturier &amp; Thomas, 2012</td>
<td>strategic problem solving</td>
<td>Healthcare</td>
<td>Qualitative</td>
</tr>
<tr>
<td>McDonnell, G., 2005</td>
<td>consensus building</td>
<td>Healthcare</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Min, P. &amp; Hong, C., 2011</td>
<td>strategic problem solving</td>
<td>Disaster</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Minami, N. a. &amp; Madnick, S., 2009</td>
<td>policy changes</td>
<td>Military</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Mohamed, S. &amp; Chinda, T., 2011</td>
<td>strategic problem solving</td>
<td>Construction</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Morris, A., Ross, W. &amp; Ulieru, M., 2010</td>
<td>strategic problem solving</td>
<td>Healthcare</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Oliva, R., 2001</td>
<td>strategic problem solving</td>
<td>Industry</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Salge, M. &amp; Milling, P.M., 2006</td>
<td>policy changes</td>
<td>Nuclear</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Shin, M. et al., 2014</td>
<td>strategic problem solving</td>
<td>Construction</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Simonovic, S.P. &amp; Ahmad, S., 2005</td>
<td>policy changes</td>
<td>Disaster</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Tang, Z., 2007</td>
<td>strategic problem solving</td>
<td>Government</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Taylor, K. &amp; Dangerfield, B., 2004</td>
<td>policy changes</td>
<td>Healthcare</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Topošek, D. &amp; Lipičnik, M., 2009</td>
<td>policy changes</td>
<td>Traffic</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Ureyy, M. &amp; Shakarian, A., 2008</td>
<td>strategic problem solving</td>
<td>Aviation</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Wang, J.Y.H. et al., 2013</td>
<td>policy changes</td>
<td>Military</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Wei, Z. et al., 2012</td>
<td>strategic problem solving</td>
<td>Aviation</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Wu, Q. &amp; Xie, K., 2012</td>
<td>strategic problem solving</td>
<td>Aviation</td>
<td>Quantitative</td>
</tr>
</tbody>
</table>
Table 2 Matrix Grid of SD Applications in Safety-Critical Domains

<table>
<thead>
<tr>
<th>External Factors</th>
<th>Organisational Influences</th>
<th>Unsafe Supervisions</th>
<th>Precondition for Unsafe Acts</th>
<th>Unsafe Acts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulatory</td>
<td>Resource Management</td>
<td>Inadequate Supervision</td>
<td>Physical Factors</td>
<td>Environmental Factors</td>
</tr>
<tr>
<td>Others</td>
<td>Organisational Climate</td>
<td>Planned Inappropriate Operations</td>
<td>Technical Factors</td>
<td>Condition of Operators</td>
</tr>
<tr>
<td></td>
<td>Organisational Pressure</td>
<td>Fault-to-Correct Problem</td>
<td>Adverse mental State</td>
<td>Adverse External Environmental Factors</td>
</tr>
<tr>
<td></td>
<td>Inadequate Supervision</td>
<td>Supervisory Violations</td>
<td>Adverse Physiological State</td>
<td>Physical Mental Fatigue</td>
</tr>
<tr>
<td></td>
<td>Physical Factors</td>
<td></td>
<td>Crew Resource Errors</td>
<td>Personnel Factors</td>
</tr>
<tr>
<td></td>
<td>Technical Factors</td>
<td></td>
<td>Personal Redundancy</td>
<td>Human Errors</td>
</tr>
<tr>
<td></td>
<td>Adverse mental State</td>
<td></td>
<td>Skill-Based Errors</td>
<td>Human Violations</td>
</tr>
<tr>
<td></td>
<td>Adverse External</td>
<td></td>
<td>Reciprocal Errors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Environmental Factors</td>
<td></td>
<td>Decision Errors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Physical Mental Fatigue</td>
<td></td>
<td>Routine Violations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exceptional Errors</td>
<td></td>
</tr>
</tbody>
</table>

Papers

- Anderson & Anderson, 1994
- Bouloiz, H. et al., 2013
- Carhart, N.J., 2009
- Ellis, B.Y.RE., 2004
- Goh, Y.M., Love, P.E.D., et al., 2012c
- Guo, S., Roudsari, A. & Garcez, A., 2013
- Han, S. et al., 2014
- Homer, J.B., 1984
- Jiang, Z. et al., 2015
- Kontogiannis, T., 2011
- Lane, D.C., Monefelt, C. & Rosenhead, J. V, 2000
- Lattimer, V. et al., 2004
- Leveson, Coutlier & Thomas, 2012
- McDonnell, G., 2005
- Mo, P. & Hong, C., 2011
- Minar, N. & Madnick, S., 2009
- Mohamed, S. & Chida, T., 2011
- Morris, A., Ross, W. & Uleru, M., 2010
- Oliva, R., 2001
- Sujeo, M. & Mbaro, P.M., 2006
- Shin, M. et al., 2014
- Simonovic, S.P. & Ahmad, S., 2005
- Tang, Z., 2007
- Taylor, K. & Dangerfield, B., 2004
- Topolšek, D. & Lipičnik, M., 2009
- Wang, J.Y.H. et al., 2013
- Wei, Z. et al., 2012
- Wu, Q. & Xie, K., 2012
CONCLUSION

System dynamics has the potential to significantly improve our capabilities and understanding in areas not well addressed by traditional safety approaches. It presents organisations and management as a tool for discerning the dynamic world of today, and offers insights to the potential trajectories they might encounter once faced with critical decisions that will affect safety. It allows every manager to see the wider scheme of things thereby enhancing the mental model and establish greater understanding of the system. The ability of SD to demystify complex problems provides a basis for understanding the current state of the system and for identifying safety improvements.

The output indicates that the majority of implemented SD applications in all sectors are primarily focused to improve the safety of external, organisational and management tiers, not so much in the workplace environment and the operator tiers. Further, most common intended output was strategic problem solving, indicating that modellers applied SD to problems with the intention to solve the perceived problem as opposed to establishing consensus building or change policies. As a result, there is a gap in the literature where applications of SD are grossly underrepresented in the sharp-end of safety. This holds particularly true in healthcare where applied SD is only found in the upper tier of the hierarchy. A future research question would be the utility and feasibility of applying SD to better understand, improve and aid safety amongst operators in the work environment. As evidenced in literature on safety, SD has the potential to contribute to safety in safety-critical domains although it is heavily underutilised. In addition, safety and sustainability in healthcare can be investigated to give decision-makers a better understanding of the complex system dynamics. Our pilot case study will aim to contribute to this ongoing question by answering how SD can improve the decision-making context in hospital pharmacy domain. The study will also fill the perceived gap of employing a system dynamics model that aims to combine soft and hard variables to produce a meaningful models that reflect the reality.
REFERENCES


Ibrahim-Shire, Jun, and Robinson


**AUTHOR BIOGRAPHIES**

**MOHAMMED IBRAHIM SHIRE** received a BSc (Hons) in Computer Science with Management from the University of Leicester in 2008 and a MSc in Internet Computing and Network Security from Loughborough University in 2012. He is currently doing his PhD in Design School at Loughborough University.

**GYUCHAN THOMAS JUN** is a lecturer at Loughborough Design School, Loughborough University. His research interest has been in bringing mapping, modelling and simulation approaches to healthcare service design to improve patient safety as well as process efficiency. He is particularly interested in the utility and usability of those approaches.

**STEWART ROBINSON** is Professor of Management Science and Acting Dean at the School of Business and Economics, Loughborough University, UK. He is also President of the Operational Research Society (2014–15), co-founder of the Journal of Simulation, and co-founder of the OR Society Simulation Workshop conference series.
THE HOSPITAL DISCHARGE GAME: A GAME THEORY-INSPIRED WORKSHOP TO ENCOURAGE COOPERATION BETWEEN HEALTH AND SOCIAL CARE ORGANISATIONS

Dr. Daniel Chalk
NIHR CLAHRC for the South West Peninsula,
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon,
EX6 7ST
d.chalk@exeter.ac.uk

Dr. Sean Manzi
NIHR CLAHRC for the South West Peninsula,
University of Exeter
South Cloisters, St Luke’s Campus, Exeter, Devon,
EX6 7ST
s.s.manzi@exeter.ac.uk

ABSTRACT

Traditionally, health and social care organisations have operated independently, with minimal interaction and poor integration of the various services with which a patient comes into contact. In some cases, ‘perverse’ systemic incentives have discouraged cooperation and led to a ‘silo mentality’. In this paper, we present an interactive workshop game that can be played with those working in and around health and social care. The game places teams in the roles of community and acute hospitals, and asks them to make discharge decisions about patients. However, a ‘perverse’ incentive is present in the system, and opposing teams are not allowed to communicate, which leads to increasingly ‘selfish’ decision making. We outline the details of this game, and show how it can be used as a tool to facilitate understanding of the benefits of cross-organisational communication and cooperation. We also present data from an initial pilot of this workshop.

Keywords: Interactive Workshop, Game Theory, Systems Thinking, Health and Social Care

1 INTRODUCTION

NHS organisations face increasing pressures coping with growing demand and declining resources (NHS England, 2013). Patient journeys through entire health and social care systems have traditionally been neglected as the focus of service improvement, which has instead sought to optimise organisations as standalone entities, rather than components of larger systems (Bell et al, 2006). However, more recently organisations have been increasingly encouraged to adopt whole systems thinking when considering how to improve their services, with an increasing focus on moving towards integrated care models (Ham et al, 2011; Naylor et al, 2015). Despite this, many organisations continue to operate as ‘silos’, perhaps because of a lack of incentive to adopt service improvement measures that benefit the whole system whilst potentially leading to apparent short-term falls in performance for the individual organisation (Naylor et al, 2015; Mannion and Braithwaite, 2012).

In this paper, we present a game that uses principles from Game Theory (von Neumann, 1959), and specifically the Prisoner’s Dilemma (Rapoport and Chammah, 1965), in order to starkly convey the pitfalls of making decisions that only benefit an individual organisation within a larger system, particularly where ‘perverse’ incentives do not encourage cooperation between organisations in a system. We outline the rules of the game, explain how the mechanics of the game lead to certain behaviours being exhibited, and then present results from a pilot of the game, which was run as a workshop for professionals working across health and social care in South West England. We show how the results from the pilot demonstrate the
desired effects of the game, and argue that this game could be used as a tool to promote whole systems thinking within the NHS, amongst both practitioners and policy makers.

2 THE HOSPITAL DISCHARGE GAME

2.1 Rules of the Game

Participants in the workshop are split into two teams – one representing an acute hospital, and one representing a community hospital. Each team should have around 3-5 members. For large workshop groups, multiple pairs of teams can be used. The game is Java-based and installed on one computer for each pair of teams.

The objective of each team is to discharge patients such that they minimise the number of penalty points that they accrue. The winning team is the one at the end of the game with the fewest penalty points. Play is divided into a series of turns, with the acute hospital and community hospital alternating play within each turn. The acute hospital begins the first turn. The team approaches the computer and is presented with a list of patients in their hospital, including an ID for each patient and their state of health (on a scale from 0% to 100%, with 100% representing full health). They are also told how many patients are currently waiting for a bed in their hospital, along with the total number of patients currently in the hospital (both in a bed and waiting for a bed), the total number of beds in the hospital, the number of new patients that have arrived this turn and the number of patients they have discharged this turn. The hospital must then decide which (if any) of their patients they will discharge, and they are encouraged to discuss their decision amongst themselves, taking care to ensure they are not heard by the other team. Once they have made their decision, play passes to the community hospital, who must make the same decision for their hospital. Once they have made their decision, the game moves to the next turn and play passes back to the acute hospital.

For every patient for whom a hospital does not have an available bed, the hospital incurs 100 penalty points. All patients discharged from the acute hospital are immediately sent to the community hospital. All patients discharged from the community hospital are immediately sent home. Patients who have been sent home may be readmitted to the acute hospital. The probability of a patient being readmitted to the acute hospital on any given turn is the inverse of their state of health (so if their state of health is 60%, they have a 40% chance of being readmitted each turn). The acute hospital incurs a penalty of 100 points for every patient who is readmitted to the hospital. In addition, on each turn a random number of new patients will arrive into the acute hospital, with randomly generated states of health.

The state of health of patients gradually improves whilst they are in hospital. Specifically, each patient in the acute hospital improves by 10% each turn (up to a maximum of 100%), whilst patients in the community hospital improve by 5% each turn (up to a maximum of 100%). However, this information is not disclosed to participants ahead of the game, and they are simply told that patients gradually get better whilst they are in hospital. The state of health of patients at home does not improve, and this is disclosed to participants. Patients cannot die in this game.

The number of turns over which the game is played can be decided by the facilitator, but it is important that participants are not told how many turns the game will last to avoid the potential for endgame strategies to be adopted. It is recommended that each hospital be given at least 10 turns to allow for pressures to build up.

Figure 1 shows a screenshot of the Java-based game. The number of turns, maximum number of new patients that can arrive each turn, the maximum state of health of new patients arriving, the total number of beds in each hospital, the number of patients in each hospital at the start of the game, the per-turn state of health increase in each hospital, the penalty points accrued for not having an available bed and the penalty points accrued by the acute hospital for a readmission are specified by the facilitator before the game begins (Figure 2). These details are not disclosed to the participants ahead of playing the game.
2.2 ‘Perverse’ Incentives

Whilst it is never explicitly highlighted to participants, and participants are not allowed to ask questions about the rules prior to playing the game to avoid the risk of highlighting the nature of the system to others, the system has a ‘perverse’ incentive that encourages conflict and selfish decisions to be made. The performance of each hospital is only judged in terms of the amount of penalty points that are accrued, and penalty points are only accrued for not having any available beds. As such, the optimum strategy should be to discharge all patients at all times, regardless of their state of health, in order to minimise the probability of not having an available bed at any time. The team that ‘wins’ the game will be the one that adopts this strategy first. This strategy is equivalent to the Prisoner’s Dilemma in Game Theory, in which the optimum ‘rational’ strategy is to make the most selfish decision, in order to maximise your average expected reward (Rapoport and Chammah, 1965).

The ‘perverse’ incentive is placed in the game in order to encourage teams to behave increasingly ‘selfishly’, by not considering the impact on the other hospital or on the patient, and to emulate real-world performance measures that only assess limited aspects of single organisations (Pollitt, 1986; Mannion and Braithwaite, 2012). In order to increase the probability of ‘selfish’ behaviour occurring, the game should be configured such that the flow of patients into the acute hospital quickly becomes overwhelming. If the acute hospital is seeking to minimise their penalty points, they will quickly start increasing the number of patients they discharge. This, in turn, increases the burden on the community hospital, leading to an
increased quantity of discharges here too. Consequently, more patients will be sent home in poorer states of health, increasing the probability of readmission to the acute hospital. This leads to a ‘snowballing’ effect, in which teams become increasingly burdened and increasingly ‘selfish’ in their decision making. If only one team behaves selfishly, the other team will incur significant penalty points as they will not be able to provide sufficient beds for the patients entering their hospital.

![Game theory in healthcare](image)

**Figure 2** Screenshot of the initial configuration options for the Java-based Hospital Discharge Game

### 2.3 The Second Iteration

Once the game has been played, participants should have an opportunity to reflect on what happened during the game. Specifically, they should have an opportunity to discuss what strategies they adopted when playing the game, and whether these strategies changed over the course of the game. The facilitator should then explain the ‘perverse’ incentive present in the system, the resultant ‘optimum’ strategy to win, and how this encourages conflict between teams.

Once the participants have understood the outcomes of the first iteration of the game, it is played again. However, in this second iteration of the game, there are several key changes to the rules. First, each pair of teams is now working together to compete against other team pairs (if there was only one pair of teams in the first iteration, teams should be split here to form two pairs of teams). Second, the winning pair of teams is the one with both the lowest penalty points accrued across both hospitals, and the highest average patient state of health at discharge from the community hospital. Third, the acute and community teams within a pair should discuss with each other what decision each hospital should take on each turn.

By changing the rules in this way, teams are actively encouraged to cooperate to meet a common goal – to reduce penalties across both hospitals and maximise patient state of health. This removes the ‘perverse’
incentive, as each hospital is now rewarded for improving the health of patients and sharing the burden of admissions across the system. Teams working cooperatively are able to discuss the current bed occupancies in each hospital and future discharge plans, allowing informed and integrated decisions to be made. This removes the snowballing effect observed in the first iteration of the game, and leads to shared responsibility between teams, resulting in lower average penalties across both organisations and higher states of health at discharge.

Once the second iteration has been played, the facilitator should again encourage reflection and discussion of how the game was played, particularly focusing on the differences in team behaviour and strategies between the first and second iterations. This may be followed by a discussion on the benefits of cooperation and whole systems thinking, and an introduction to Game Theory and the Prisoner’s Dilemma.

3 PILOT RUN OF THE GAME

On 16th June 2015, a pilot session of the Hospital Discharge Game was run as part of the Peninsula Collaboration for Health Operational Research and Development (PenCHORD) Seminar, Showcase and Workshop Series. These events are a capacity building initiative of the NIHR CLAHRC for the South West Peninsula (PenCLAHRC), and are targeted primarily at those working in and around health and social care in the South West of England. The key aim of the events is to increase awareness in the NHS of Operational Research, and how such techniques can be applied to problems in health and social care to inform decision making. As part of this, an afternoon workshop session at each event introduces an aspect, technique or idea from Operational Research using an interactive format. The Hospital Discharge Game was piloted during this session at 16th June event.

3.1 Participants and Game Configuration

There were around 30 participants in the pilot session of the game, the majority of whom worked in the NHS (with the remaining participants representing healthcare researchers). The participants were divided into three pairs of teams, such that each team had around five team members. Each acute hospital had a total of ten beds available to them, whilst each community hospital had a total of 15 beds. At the start of the game, the acute hospital had 12 patients in the hospital (and therefore two immediately without beds), and the community hospital had 15 patients. Patients in the acute hospital improved at a rate of 10% per turn, compared to community hospital patients who improved by 5% each turn. Penalties for lack of beds were set to be the same for both hospitals – 100 points for each patient without a bed each turn. The acute hospital received an additional penalty of 100 points for every readmission to the hospital. A minimum of one and a maximum of seven new patients would arrive at the acute hospital each turn, presenting with a state of health between 1% and 80%. The game was played for 20 turns for the first iteration, and 10 turns for the second iteration.

3.2 Results from the First Iteration of the Game

In the first iteration of the game, Team Pair 1 incurred 45,600 penalty points in the acute hospital compared with 63,200 penalty points in the community hospital. Team Pair 2 incurred 78,500 penalty points in the acute hospital compared with 5,100 points in the community hospital. Team Pair 3 incurred 85,300 penalty points in the acute hospital compared with 6,300 points in the community hospital.

In both Team Pair 2 and 3, the community hospital dominated the game by adapting to selfish decision making more quickly than the acute hospital. The community hospitals in Team Pairs 2 and 3 discharged a total of 290 and 262 patients respectively, compared to the acute hospitals who discharged 275 and 247 patients respectively. In contrast, the community hospital in Team Pair 1 lost the game, and only discharged 276 patients compared to the acute hospital’s 333 discharges.

The average state of health of patients discharged from the community hospital was 48.2% for Team Pair 1, 52.2% for Team Pair 2, and 65.3% for Team Pair 3.
Table 1 provides full details of the total discharges, readmissions, penalties and mean state of health at discharge for each hospital in each team pair in the first iteration of the game.

Table 1 Total discharges, readmissions, penalties and mean state of health at discharge across the two hospitals in each Team Pair for the first iteration of the game

<table>
<thead>
<tr>
<th>Team Pair</th>
<th>Acute Total Penalties</th>
<th>Community Total Penalties</th>
<th>Acute Discharges</th>
<th>Community Discharges</th>
<th>Acute Readmission Penalties</th>
<th>Acute Readmission</th>
<th>Mean State of Health at Discharge from Acute Hospital</th>
<th>Mean State of Health at Discharge from Community Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45600</td>
<td>63200</td>
<td>333</td>
<td>276</td>
<td>26400</td>
<td>264</td>
<td>39.5%</td>
<td>48.2%</td>
</tr>
<tr>
<td>2</td>
<td>78500</td>
<td>5100</td>
<td>275</td>
<td>290</td>
<td>26300</td>
<td>263</td>
<td>52.6%</td>
<td>52.2%</td>
</tr>
<tr>
<td>3</td>
<td>85300</td>
<td>6300</td>
<td>247</td>
<td>262</td>
<td>22300</td>
<td>223</td>
<td>65.4%</td>
<td>65.3%</td>
</tr>
</tbody>
</table>

3.3 Results from the Second Iteration of the Game

In the second iteration of the game, Team Pair 1 incurred 9,300 penalty points in the acute hospital and 7,000 penalty points in the community hospital. Team Pair 2 incurred 15,800 penalty points in the acute hospital and 1,900 points in the community hospital. Team Pair 3 incurred 15,600 penalty points in the acute hospital and 1,500 points in the community hospital. Overall, the total penalty points accrued across both hospitals fell by 92,500 for Team Pair 1, 65,900 for Team Pair 2, and 74,500 for Team Pair 3. It should be noted that the second iteration only ran for half the length of the first iteration, but even considering this the differences are significant.

The average state of health at discharge from the community hospital increased significantly in the second iteration of the game. Specifically, there was an increase of 51.8% for Team Pair 1, 25.1% for Team Pair 2, and 10.3% for Team Pair 3.

Table 2 provides details of total discharges, readmissions, penalties and mean state of health at discharge for each hospital in each team pair in the second iteration of the game. Table 3 compares total penalties and community hospital mean state of health at discharge across the two iterations of the game.

Table 2 Total discharges, readmissions, penalties and mean state of health at discharge across the two hospitals in each Team Pair for the second iteration of the game

<table>
<thead>
<tr>
<th>Team Pair</th>
<th>Acute Total Penalties</th>
<th>Community Total Penalties</th>
<th>Acute Discharges</th>
<th>Community Discharges</th>
<th>Acute Readmission Penalties</th>
<th>Acute Readmission</th>
<th>Mean State of Health at Discharge from Acute Hospital</th>
<th>Mean State of Health at Discharge from Community Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9300</td>
<td>7000</td>
<td>28</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>87.7%</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>15800</td>
<td>1900</td>
<td>42</td>
<td>47</td>
<td>3100</td>
<td>31</td>
<td>75.3%</td>
<td>77.3%</td>
</tr>
<tr>
<td>3</td>
<td>15600</td>
<td>1500</td>
<td>45</td>
<td>50</td>
<td>3300</td>
<td>33</td>
<td>75.7%</td>
<td>75.6%</td>
</tr>
</tbody>
</table>

4 DISCUSSION

In this paper, we have outlined an interactive game that can be played with health and social care professionals in order to improve understanding of the importance of whole systems thinking when making decisions that affect other organisations. The game presents an exaggerated example of a ‘perverse’ incentive that could discourage cooperation between organisations. Nevertheless, ‘perverse’ incentives can
be found in the way in which some organisations in the NHS are rewarded or penalized (Naylor et al, 2015; Berry et al, 2015; Kar, 2015).

**Table 3 Comparison of total penalties and mean state of health at discharge between the first and second iterations**

<table>
<thead>
<tr>
<th>Team</th>
<th>Total Penalties Across Both Hospitals (First Iteration)</th>
<th>Total Penalties Across Both Hospitals (Second Iteration)</th>
<th>Difference in Total Penalties</th>
<th>Mean State of Health at Discharge from Community Hospital (First Iteration)</th>
<th>Mean State of Health at Discharge from Community Hospital (Second Iteration)</th>
<th>Difference in Mean State of Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>108800</td>
<td>16300</td>
<td>-92500</td>
<td>48.2%</td>
<td>100%</td>
<td>+51.8%</td>
</tr>
<tr>
<td>2</td>
<td>83600</td>
<td>17700</td>
<td>-65900</td>
<td>52.2%</td>
<td>77.3%</td>
<td>+25.1%</td>
</tr>
<tr>
<td>3</td>
<td>91600</td>
<td>17100</td>
<td>-74500</td>
<td>65.3%</td>
<td>75.6%</td>
<td>+10.3%</td>
</tr>
</tbody>
</table>

The pilot of this workshop demonstrated the intended effects of the game. During the first iteration, decisions were made under increasing pressure from overwhelming demand for beds and a lack of knowledge of the capacity and decisions of the other hospital in the system. Participants visibly became increasingly frustrated with the escalating number of admissions or readmissions into their hospital, leading to a gradual increase in ‘selfish’ decisions that did not consider patient wellness or the impact on the other organization. In turn, this led to a ‘vicious circle’ in which increasing demand was placed on the other organization, causing them to have to discharge more patients early, leading to more readmissions and escalating levels of demand for both services.

In contrast, when playing the second iteration, teams were able to communicate and make joint decisions about their discharges. Both teams were working towards a common goal – low penalties and high levels of wellness for their patients across the system. Consequently, teams made decisions that may not have appeared immediately optimal when considering only their service, but which improved the outcomes in the system as a whole. This led to significant increases in patient state of health at discharge, and significant decreases in penalties from beds being blocked.

This game shares some similarities with the Beer Distribution Game (Sterman, 1984), which is a long-standing Operational Research game designed to demonstrate the dangers of lack of communication and individual-level optimisation within a larger system. However, the Beer Distribution Game uses a production line as its example context, which may not immediately find resonance with health and social care professionals and the way in which they view their systems. In contrast, the Hospital Discharge Game presents a pertinent example of a problem within healthcare systems, and one which is likely to be familiar to many working within the NHS. Therefore, the potential for health professionals to understand the direct relevance of the game to their own organisational thinking is arguably more significant. Another game – Friday Night at the ER (http://fridaynightattheer.com/ accessed 5th November 2015) – uses principles from the Beer Game in the context of managing patient flow in a hospital during a 24 hour period to teach principles of systems thinking. However, this game does not contain ‘perverse’ incentives within the system, meaning that the resultant behavior, outcomes and intended reflection are different to the Hospital Discharge Game.

There has been much effort in recent years to improve collaboration between NHS organisations, and move towards integrated care models that consider the whole system within which patients access care (Ham et al, 2011; Naylor et al, 2015). However, it can often be difficult to move thinking away from the ‘silo mentality’ that has plagued the health service for many years, not helped by performance measures that often still focus on individual organisation performance (Mannion and Braithwaite, 2012). We propose that this workshop game could be an effective means of conveying a simple message to health and social
Care professionals – that whole systems thinking, and increased cooperation between organisations, can lead to improved patient outcomes and improved longer-term performance across the system. In addition, this game could be used with policy makers to demonstrate the risks of employing performance measures that only consider the performance of an individual organization, and the potential significant benefits from incentivising cross-organisational service improvement.

ACKNOWLEDGMENTS

This research was funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

Berry N, Gardner T and Anderson I (2015). On targets: How targets can be most effective in the English NHS. Policy Analysis

AUTHOR BIOGRAPHIES

DANIEL ChALK achieved a BSc Computer Science at Exeter University in 2005. He completed his PhD in Biosciences at the same university in 2010, exploring the use of simulation to model the landscape-scale foraging behaviours of bumble bees to predict the risk of GM to non-GM crop cross-pollination. He is currently a Research Fellow within the Operational Research team (PenCHORD) of the South West Peninsula Collaboration for Leadership in Applied Health Research and Care (PenCLAHRC), where he applies modelling, simulation and analysis techniques to help inform decision making in the local NHS, and has a particular interest in modelling behavioural aspects within health and social care systems. Dr Chalk also leads a number of initiatives that build capacity within the NHS to generate and use operational research evidence, including a training programme that helps those working in the NHS to develop practical modelling skills that they can use in their own organisations.
SEAN MANZI received a BSc (Hons) Psychology from Plymouth University in 2011. He completed his PhD at the same University in 2014. He is currently an Associate Research Fellow with the South West Peninsula Collaboration for Leadership in Applied Health Research and Care. Dr Manzi is using his multidisciplinary research experience to improve healthcare provision and develop new research methods to support this process. http://clahrc-peninsula.nihr.ac.uk/staff/sean-manzi