THE OR SOCIETY

OR ESSENTIALS

SYSTEM DYNAMICS

SOFT AND HARD OPERATIONAL RESEARCH

EDITED BY

MARTIN KUNC

OR Essentials

Series editor
Simon Taylor
Brunel University
Uxbridge, United Kingdom

The OR Essentials series presents a unique cross-section of high quality research work fundamental to understanding contemporary issues and research in across a range of Operational Research (OR) topics. It brings together some of the best research papers from the esteemed Operational Research Society and its associated journals, also published by Palgrave Macmillan. OR deals with the use of advanced analytical methods to support better decision making. As a multidisciplinary field, it has strong links to management science, decision science, computer science and has practical applications in areas such as engineering, manufacturing, commerce, healthcare and defence. OR has longstanding historical roots. However, as a modern discipline it arguably started during World War II when mathematical techniques were being developed to support the war effort. Now it is common place and a key discipline taught in universities across the world, at undergraduate and postgraduate levels. There are several international societies dedicated to the advancement of OR (e.g. the Operational Research Society and INFORMS The Institute for Operations Research and the Management Sciences) and there are many high quality peer-review journals dedicated to the topic. The OR Essentials books are a vital reference tool for students, academics, and industry practitioners, providing easy access to top research papers on the most cutting-edge topics within the field of Operational Research.

More information about this series at
http://www.palgrave.com/series/14725

Martin Kunc
Editor

System Dynamics

Soft and Hard Operational Research

palgrave
macmillan

Editor
Martin Kunc
Warwick Business School
University of Warwick
Warwick, United Kingdom

OR Essentials
ISBN 978-1-349-95256-4 ISBN 978-1-349-95257-1 (eBook)
https://doi.org/10.1057/978-1-349-95257-1

Library of Congress Control Number: 2017953539

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Macmillan Publishers Ltd.
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

Contents

List of Figures

List of Tables

Improving the Cost-Effectiveness of Chlamydia Screening with Targeted Screening Strategies

Competitive Dynamics in Pharmaceutical Markets: A Case Study in the Chronic Cardiac Disease Market

Introduction

M. Kunc

Introduction

System dynamics (SD) was founded by Jay Forrester at the Massachusetts Institute of Technology in 1957 (Forrester 1961). Different from other operational research (OR) tools and methods, SD can adopt two modes of operations: it can involve the use of qualitative tools (e.g. causal loop diagrams) followed by quantitative simulation (e.g. stocks and flows networks), depending on the purpose of analysis (Wolstenholme 1999). Another interesting distinction of SD modelling is that models can be developed in either isolated or participative modes. Depending on the type of enquiry, there are two basic modes of operation. First is an essentially descriptive mode, which can be defined as a soft perspective and operates in a similar fashion to problem structuring methods. Second is a predictive/prescriptive mode, which can be considered a hard perspective and solves problems in the same manner as forecasting and optimization. Before discussing both perspectives in SD, a brief explanation of hard and soft perspectives is offered. The distinc-

M. Kunc (✉)
Warwick Business School, University of Warwick, Coventry, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_1

Table 1 Comparison of hard and soft perspectives

Author	Soft perspective	Hard perspective
Checkland (1985)	The soft tradition involves an appreciation of the world, or a learning perspective about how systems work, that can be explored through models. Models are intellectual constructs to identify issues and achieve accommodation. The soft tradition in systems thinking is the base for the development of soft systems methodology.	The hard tradition in systems thinking is based on goal-seeking behaviour. This view suggests that social systems can be designed to achieve the objectives for the system. Thus the focus is on efficient means to achieve the objectives. Models are representations of the world that has problems which need solutions. The solution is obtained through quantitative analysis.
Paucar-Caceres (2011)	There is a process of learning from the intervention through understanding the purposes of the actors. The aim is to explore and generate learning while bringing consensus or accommodation between the actors in the system. Quantitative analysis is not useful in a socially constructed reality, so efforts are aimed at maintaining relations through negotiations. Methodologies/methods associated with this paradigm: soft system methodologies, strategic assumption surfacing and testing, strategic choice, cognitive mapping, SODA journey, problem structuring methods, and viable systems model.	Hard approaches try to discover laws ruling the relationships between variables and leading to deep structures. Therefore the interventions are systematic and intent to improve the viability of the system through optimization and problem-solving using quantitative analysis. Systems are sociotechnical with elements combined in a certain structure to achieve goals. Decision-makers follow a rational process but they are bounded rational. Methodologies/methods associated with this paradigm: programming methods, simulation, forecasting, decision trees, queuing theory, Markov analysis, system dynamics and complexity theory.

(*continued*)

Table 1 (continued)

Author	Soft perspective	Hard perspective
Daellenbach, McNickle, Dye (2012)	Soft approaches deal with complex problems which are messy, ill structured, ill defined and dependent on the stakeholders' perspectives. The stakeholders' perspectives are conflicting and without agreement. The approaches intend to structure the problems rather than focusing on a solution through facilitating dialogue so that there is a common understanding. Facilitation is fundamental to working together with stakeholders in order to find a resolution to the problem while achieving accommodation for their implementation. The role is facilitation not an expert analyst/consultant providing a solution. Human aspects in the definition of the problem are important.	Modelling under a hard OR paradigm implies that • the behaviour observed can be captured in mathematical models; • the problem has been clearly defined in terms of objectives, structure, constraints, input data and criteria to evaluate the achievement of objectives, including trade-offs; • then alternative courses of actions are defined in terms of options or decision variables, and optimization of the objectives is the focus; • decision-maker with authority to implement the solution or enforce implementation through the hierarchical chain of command.

tion originates from a long-term debate in the OR community, more precisely the systems community, in the UK. A brief synthesis is described in Table 1, which clearly shows a soft perspective attempt to describe systems. The reason for this approach is that systems are perceived differently by diverse stakeholders. Dissimilar points of view hinder the solution to the issues observed, so the most important step is to achieve consensus on what the system is and then what the problem relating to the system is. On the other hand, a hard perspective aims to solve the problem by starting with the assumption that systems can be described and engineered using insights generated by quantitative models.

Considering the distinctions between hard and soft perspectives obtained from Table 1, the next section offers an approximation to both perspectives from SD scholars. The discussion is a particular and limited perspective on this debate, which has been in the literature for many years. I don't intend either to provide a comprehensive and systematic review or to take a specific position. (To learn the position in detail, see, e.g., Lane 1994, Lane and Oliva 1998, Wolstenholme 1999, Lane 2000 and Homer and Oliva 2001.)

The Case for the Soft Perspective in System Dynamics

Forrester (1987) suggests that the influence and usefulness of SD models rests on the insights and generalizations that can be obtained from the modelling process, the model vs. the modelling process. Models are made to organize, clarify and unify knowledge while providing a valuable perspective on a system having perplexing behaviour for a target audience (Forrester 1987).

Modelling starts by identifying the mental models that are going to be improved, not an efficiency goal (Forrester 1987). One of the reasons for this assertion is that knowledge about structures and policies responsible for system behaviour are necessary to represent a particular problem. SD is close to a problem structuring method (Eden and Ackermann 2006). When this is the case, the objective is to represent the system structure and communicate it in the most transparent way to relevant stakeholders. In group model building literature, quantitative SD models are seen as boundary objects, which are socially constructed artefacts, to make sense of the problems in a similar fashion to problem structuring methods (Andersen et al. 2007). However, the main difference between SD and problem structuring methods is the existence of an organizing framework for the description of the system that led to a quantitative model (Forrester 1994).

SD considers the context of the decisions such as the impact of goals, limited information and bounded rationality (Morecroft 1985). Decisions are captured in a model based on how they are made, not how they should be made. Thus SD is a modelling tool that does not aim to represent rational, optimizing decision-makers who can calculate all options. It recognizes that decision-makers are bounded rational so there is a need to use a soft perspective—for example, optimism drives investment in assets (Morecroft 1999) rather than optimal capital allocation. Homer and Oliva (2001) suggest that soft variables, which are difficult to measure and subject to multiple causes, have been at the core of SD models since the beginning of the field. SD has principles and guidelines for representing decision-making, human behaviour and nonlinear relationships underpinning soft variables.

Forrester (1994) and Wolstenholme (1999) propose that a soft perspective is based only on qualitative analysis and not on quantitative modelling. Qualitative analysis involves the use of causal loop diagrams, or word-and-arrow diagrams, to describe a system in more detail and lead to standalone policy analysis. The intention in qualitative analysis is to improve the thinking about a structure behind a problem. Therefore there is no development of dynamic hypothesis to explain the reference mode.

Homer and Oliva (2001) propose that qualitative analysis describes structure and not dynamics. The richer the description of the system structure, the less the relationship with the root causes of the reference mode, which is the basis for dynamic hypotheses.

Wolstenholme (1999) also suggests that qualitative system dynamics can be based on stocks and flows diagrams. This idea was later adapted to the field of strategic management by many researchers from the London Business School, such as Warren (2002), Morecroft (2002), and Kunc and Morecroft (2009), to represent firms as systems of asset stock accumulation.

The Case for the Hard Perspective in System Dynamics

Forrester (1987) suggests that problems in complex systems are difficult to solve based on intuition because they are high-order nonlinear dynamic systems. Once the knowledge held by decision-makers is captured through the structure of the SD model, then the model is constructed to represent the current behaviour of the system and to offer solutions to the puzzling behaviour (Forrester 1987; Wolstenholme 1999). Forrester (1994) suggests that, in addition to understanding, the modelling project has as a goal an improvement of the system. In other words, this approach does not differ from the practice of an OR user who uses linear programming to optimize the performance of a factory or improve the transportation activities of a company. SD researchers use models to project alternative futures defined by actions taken under a set of conditions (Forrester 1987) in a similar fashion as hard OR researchers, such as Paich, Peck and Valant (2004) in pharmaceutical markets. Currently, more SD researchers are adopting established models and developing incremental innovations to convey the different contexts where a model is applied. This behaviour is not different from the process performed by many hard OR researchers. SD employs numerical data to calculate parameter values, characterize system behaviour and compare with the model's output (Forrester 1987; Homer and Oliva 2001). This behaviour does not differ from the use of numerical data by other modellers. Moreover, the quantitative model tests the hypothesis that the structure of the model is a good representation of the dynamics observed in the reference modes (Homer and Oliva 2001).

Perhaps one of the key differences between traditional hard OR and SD is the emphasis in SD on defining model boundaries where most causal mechanisms lie inside the model. Other hard OR methods may use stochastic variables to reflect the impact of causal mechanisms not accounted for within the model. SD is a deterministic modelling approach owing to the important effort by modellers to account for all causal mechanisms responsible for the behaviour observed. Formulation implies theory-building in SD (Forrester 1994). Another difference with hard OR is considering additional information sources aside from numerical databases for the model, such as users' mental repositories. It is in this area (the use of mental repositories) where SD acknowledges the importance of the description of the problem as an influencing factor in the design of a quantitative model. This is still an area that colleagues using hard OR tools do not appreciate fully but it is considered important by soft OR practitioners.

SD modellers intend to quantify variables that are mostly difficult to represent and where there is huge uncertainty about them. Some SD researchers doubt the usefulness of a quantitative model with so many uncertainties that the results of the model can be misleading (Coyle 2000). The enthusiasm for quantifying relationships originated from the need to test empirically the hypothetical structure. In many cases, the data for hard-to-quantify—or soft—variables is tacit. Tacit data can only be obtained from the elicitation of mental models existing either in decision-makers or hidden in the verbal description of a theoretical framework (see e.g. Sterman and Wittenberg 1999). Afterwards the values of hard-to-quantify variables are evaluated through sensitivity testing. Thus modelling aims to comply with empirical traditions in social science (Homer and Oliva 2001). Table 2 summarizes the soft and hard perspectives in SD.

A Historical Perspective of the System Dynamics Field in Terms of Hard and Soft Contributions from the Publications in *Journal of the Operational Research Society*

I performed a search of *Journal of the Operational Research Society* (*JORS*) using the term 'system dynamics' in the title, abstract or keywords until 2016 and obtained 122 papers. The information obtained shows that *JORS* published one paper per year on average between 1978 until 1996. Then the number of papers increased to five per year on average without including a special issue on SD, which was edited by John Morecroft and Geoff Coyle, published in 1999. The special issue was the only issue dedicated to the field of SD in the history of *JORS*. The list generated from the search (see Appendix) also shows growing use of SD in the field of healthcare in recent years. For example, papers with applications in healthcare account for more than 25 % of those published since 2000.

Next I categorized each paper according to soft and hard perspectives considering the concepts outlined in Table 2. After the papers were categorized, the results were 65 % hard and 35 % soft, without significant

Table 2 Characteristics of hard and soft system dynamics perspectives

Modelling aspect	Characteristics of soft perspective	Characteristics of hard perspective
Objective of modelling	Understanding the system in terms of feedback structures followed by qualitative policy design.	Hypothesis testing of puzzling dynamic behaviour followed by quantified improvement of the behaviour.
Inputs	Mental models are the only input which obtained through mostly facilitated face-to-face meetings. Another input for theoretical models is the researcher's interpretation of existing theories or frameworks.	Mental models/theoretical frameworks are inputs to define the model structure. Numerical data can be sourced from three sources. First, judgemental data is the origin of unknown or hard-to-measure parameters. Second, numerical data is used for parameters when there is available data. Third, facilitation or judgement is usually the source of nonlinear (table) functions.
Process	The system is described through causal loop diagrams These aggregate individual interpretations of causal links between concepts existing in the system. Facilitation is critical to uncovering most of the causal links existing in the group of actors embedded in the system. Quantitative models are used to facilitate learning and discussion during the process.	There are initial descriptions of a hypothesized structure responsible for the dynamic behaviour. The hypothesized structure is transformed into stocks (level), flows (rates), auxiliaries and causal relationships between the elements of the structure. Equations are formulated and parameters populated. Extensive testing of the structure, parameters and outputs is performed to confirm that the structure replicates the dynamic behaviour.
Outputs	Learning about the structure of the system and potential policies. Changes in participants' perspectives on the system. Accommodation and agreement on future policies.	Stylized graphs showing performance over time of relevant variables. Policies to improve dynamic behaviour are tested numerically. Learning about dynamics through the use of models with enhanced interfaces.

difference between 1978–1999 and 2000–2016. There seems to be a strong presence of hard SD compared with soft SD. If the results are compared for the same period with respect to the percentage of papers obtained with the keywords 'problem structuring methods' and 'soft OR' (64 papers) with respect to 'optimization' (725 papers) in *JORS*, they show a significant participation of soft models in SD publications. However, this comparison is not robust and further analysis of papers using soft perspectives is needed, especially to check if they lead to quantitative models.

Like any search process and further categorization, there are always limitations because authors may use different keywords while employing some of the methods discussed here.

The Book Content

It has been difficult to make a selection of only 10 % of the articles published in *JORS*. There are many excellent contributions that could have been part of this edited volume. However, I believe that the papers selected offer an initial base for more promising work in the field. The list of papers included is given in Table 3. The distribution of papers selected resembles the soft/hard distribution with 29 %/71 %, respectively. The main aim was to select papers after the *JORS* 1999 special issue, except in cases where there were none to reflect a particular issue, such as theory-building and methodologies.

In Part I there are four chapters showing the use of SD in the field of management offering a broad spectrum of the type of contributions typical of SD. Larsen and Lomi (1999) show the use of SD for theory-building and testing in the field of organizational inertia and change. This is a good example of testing a dynamic hypothesis through modelling. Morecrot (1999) uses a case study to examine a theoretical puzzle in corporate strategy literature combining managerial and behavioural decision-making in a quantitative model. This paper reflects another example of dynamic hypothesis-testing. Akkermans and van Oorschot (2005) present the implementation of a performance-measurement system using a causal loop diagram. Then, additional learning was achieved using a quantitative model to demonstrate the future performance of

Table 3 Papers included in this volume

Chapter	Author(s)	Year	Title	Soft/hard perspective
2	Larsen, E.R. and Lomi, A.	1999	Resetting the clock: A feedback approach to the dynamics of organisational inertia, survival and change	This paper use empirical theory building and testing
3	Morecroft, J.D.W.	1999	Management attitudes, learning and scale in successful diversification: A dynamic and behavioural resource system view.	This paper uses an empirical model to replicate company performance
4	Akkermans, H.A. and van Oorschot, K.E.	2005	Relevance assumed: A case study of balanced scorecard development using system dynamics	A soft approach was used to enhance the understanding of causal links in the design of performance measures
5	Scott, R.J., Cavana, R.Y. and Cameron, D.	2015	Interpersonal success factors for strategy implementation: A case study using group model building	The evaluation of the soft approach is performed through a survey of participants
6	Dangerfield B. and Roberts C.	1998	An overview of strategy and tactics in system dynamics optimization	This explores model optimization so it corresponds to a hard perspective
7	Coyle, R.G.	1999	Simulation by repeated optimisation	This explores the perils of a hard approach, model optimization, in SD
8	Syntetos, A.A., Georgantzas, N.C., Boylan, J.E. and Dangerfield, B.C.	2011	Judgement and supply chain dynamics	This uses SD to experimentally test issues with human judgement

(*continued*)

Table 3 (continued)

Chapter	Author(s)	Year	Title	Soft/hard perspective
9	Tako, A.A. and Robinson, S.	2009	Comparing discrete-event simulation and system dynamics: Users' perceptions	This addresses an important aspect of a soft perspective: learning about the system using models.
10	Xing, Y. and Dangerfield, B.	2011	Modelling the sustainability of mass tourism in island tourist economies	A full-scale model of the economy of an island is developed and calibrated with real data and tested using Monte Carlo simulation
11	Olaya, Y. and Dyner, I.	2005	Modelling for policy assessment in the natural gas industry	A model based on economic and optimization concepts calibrated historical data is used to assess policies
12	Howick, S. and Whalley, J.	2008	Understanding the drivers of broadband adoption: The case of rural and remote Scotland	The model starts with a causal loop diagram to represent the dynamic hypothesis. Then a full-scale model is developed using different data sources, sensitivity analysis and policy experimentation
13	Lane, D.C. and Husemann, E.	2008	System dynamics mapping of acute patient flows	This is an excellent example of soft approaches using stocks and flows diagrams

(*continued*)

Table 3 (continued)

Chapter	Author(s)	Year	Title	Soft/hard perspective
14	Evenden, D., Harper, P.R., Brailsford, S.C. and Harindra, V.	2006	Improving the cost-effectiveness of Chlamydia screening with targeted screening strategies	The extensive use of data and other hard OR techniques together with SD led these authors to propose strategies to optimize the performance of the healthcare system
15	Kunc, M. and Kazakov, R.	2013	Competitive dynamics in pharmaceutical markets: A case study in the chronic cardiac disease market	Initially, a causal loop diagram helped to define the dynamic hypothesis. This was validated and policies, such as sensitivity analysis, were suggested using uncertainty ranges

the indicators. Scott, Cavan and Cameron (2015) discuss the effectiveness of soft perspectives to facilitate strategy implementation. In this case they surveyed participants of group model-building using a questionnaire extensively used in group model-building literature (Vennix and Rouwette 2000; Rouwette 2011).

Part II illustrates contributions to methodology and the use of SD models. Initially, two chapters related to model optimization are presented: Dangerfield and Roberts (1996) cover optimization to fit data to the model and policy optimization to improve system performance, and Coyle (1999) addresses issues with optimization applied to SD models. Model optimization received very little attention for many years but it is now becoming more relevant, such as in Rahmandad et al.'s (2015) work using multimethod simulation software. The next chapter, by Syntetos et al. (2011), presents one of the key strengths of SD: experimental studies in supply chain dynamics. Since the seminal paper by Sterman (1989), the work in this area has grown substantially. Finally, Tako and Robinson (2009) present another piece of experimental work on an important recent trend: the use of SD in comparison with other methods. Their chapter shows the results of a study on the perceptions of users with respect to the insights generated by SD compared with traditional discrete-event simulation. While there are no significant differences, SD provides a better understanding of the system structure so it helps conceptual learning.

Part III involves chapters covering industry-level models. SD has a strong tradition in this area owing to its versatility in representing global aggregates and long-term feedback processes. Both aspects are critical characteristics of industry dynamics. The key contribution in this area is the possibility of identifying policies to shape the behaviour of industries, evaluate the potential evolution of industries, and discover the factors affecting their growth and sustainability. For example, Xing and Dangerfield (2011) evaluate the sustainability of mass tourism in islands; Olaya and Dyner (2005) assess policies for the natural gas industry in Colombia; and Howick and Whalley (2008) discuss policies to promote the adoption of broadband in Scotland.

Part IV is devoted to the most promising area of SD: healthcare. SD has demonstrated that is a widely accepted methodology to represent patient and clinical pathways using stocks and flows networks. Then, multiple interventions can be rehearsed on healthcare systems: increasing resources, improving screening strategies and observing the dynamics of medicine costs as a result of the behaviour of patients, government and pharmaceutical companies. Lane and Husemann (2008) demonstrat the use of qualitative systems dynamics using stocks and flows diagrams to structure patient pathways. The conceptual stocks and flows diagrams were later used in workshops to generate ideas for improving the system. This is an excellent example of a soft perspective using the unique characteristics of SD: stocks and flows diagrams. Evenden et al. (2006) won the Goodeve Medal 2006 for the best paper in *JORS* for that year. They employed mixed methodologies (clustering techniques, geomapping techniques and SD to calculate the infection dynamics) to improve the cost-effectiveness of screening strategies in a transmitted infectious disease. Finally, Kunc and Kazakov (2013) presented an analysis of the interventions that governments can make in the pharmaceutical industry to reduce healthcare costs for chronic diseases. The model is strongly based on data and sensitivity analysis of policies, which led to a set of best policies within uncertainty.

Conclusion

After reflecting on the links between SD and soft system methodologies, Morecroft (2015) asserts that SD 'is hard system modelling dressed in soft clothing'. I agree with this assertion. SD can be safely included in the hard OR toolkit like any other tool use for hard system modelling. Basically, objectives for a SD model are defined similarly to a linear programming model but SD modellers try to maximize multiple objectives instead of maximizing just one criterion. The SD model does not pretend to represent the whole system but to be a useful quantitative model to test options to improve performance. In a similar fashion, hard OR models test technical solutions to improve performance. Other evidence for hard modelling is the qualitative aspects of SD defined by formal tools and methods (e.g. causal loop diagrams and feedback loops). However, many problem structuring methods also have formal tools, methods and frameworks (e.g. soft system methodology; Checkland 1985), so they share similarities with SD. The formalism in interventions is necessary to be able to advance the field since there is no possibility of differentiating good from bad practices even in complex and confused worlds without formalism.

Then the next question is about the 'soft clothing', so I discuss a few aspects that dress SD with soft clothes. First, SD modellers explicitly acknowledge that they do not conceptualize models from explicit tangible data but rather mostly from tacit data (e.g. mental models). Hard OR modellers, who work on applications, also engage with tacit data but there is no explicit account of their model-building process and how they reach the final model. For example, the restrictions/constraints to include in a staff rostering model, a typical linear programming problem, will depend on the client and interactively on the data available. Thus there is a model conceptualization process before the technical solution, even in hard OR. In that sense, any OR model is soft. Second, SD modellers are not only concerned about quantitative models to improve unacceptable reference modes. There is an established line of work considering qualitative and quantitative models as transactional objects that facilitate an understanding of complex and confusing worlds. Often a small model that shows some results helps the understanding of complex and confusing views.

Therefore there are no completely separate soft and hard perspectives in SD but the output of the model has to be contextualized with the situation and the objective of the project. It is important that practitioners perform two processes. First, they must position their SD modelling adequately in the respective stream of the SD literature. Second, they must reflect on how a project/intervention could have been done if the opposite perspective had been used. In this way, practitioners will start to recognize the benefits and limitations of adopting a particular perspective.

Appendix

Table 4 List of papers published in *JORS* obtained from the search

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Liu S., Osgood N., Gao Q., Xue H., Wang Y.	Systems simulation model for assessing the sustainability and synergistic impacts of sugar-sweetened beverages tax and revenue recycling on childhood obesity prevention	2016	67	5	H
Scott R.J., Cavana R.Y., Cameron D.	Interpersonal success factors for strategy implementation: A case study using group model building	2015	66	6	S
Brailsford S., De Silva D.	How many dentists does Sri Lanka need? Modelling to inform policy decisions	2015	66	9	H
Syms R., Solymar L.	A dynamic competition model of regime change	2015	66	11	H
Yang S.-J.S., Emma Liu Y.	Anticipated responses: The positive side of elicited reactions to competitive action	2015	66	2	H
Vanderby S.A., Carter M.W., Latham T., Feindel C.	Modelling the future of the Canadian cardiac surgery workforce using system dynamics	2014	65	9	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Demir E., Lebcir R., Adeyemi S.	Modelling length of stay and patient flows: Methodological case studies from the UK neonatal care services	2014	65	4	H
Ahmed R., Robinson S.	Modelling and simulation in business and industry: Insights into the processes and practices of expert modellers	2014	65	5	S
Kunc M., Kazakov R.	Competitive dynamics in pharmaceutical markets: A case study in the chronic cardiac disease market	2013	64	12	H
Xue C.G., Liu J.J., Cao H.W.	Research on competition diffusion of the multiple-advanced manufacturing mode in a cluster environment	2013	64	6	H
Busby J.S., Onggo S.	Managing the social amplification of risk: A simulation of interacting actors	2013	64	5	H
Zarracina M.L.	Heavy fuel oil analysis at Iraq's Bayji refinery	2013	64	4	H
Cannella S., Barbosa-Póvoa A.P., Framinan J.M., Relvas S.	Metrics for bullwhip effect analysis	2013	64	1	H
Harrop N., Gillies A., Wood-Harper A.T.	Actors and clients: Why systems dynamics needs help from soft systems methodology and unbounded systems thinking	2012	63	12	S
Atkinson M.P., Gutfraind A., Kress M.	When do armed revolts succeed: Lessons from Lanchester theory	2012	63	10	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Duran-Encalada J.A., Paucar-Caceres A.	A system dynamics sustainable business model for Petroleos Mexicanos (Pemex): Case based on the Global Reporting Initiative	2012	63	8	S
Wong H.J., Morra D., Wu R.C., Caesar M., Abrams H.	Using system dynamics principles for conceptual modelling of publicly funded hospitals	2012	63	1	S
Paucar-Caceres A., Espinosa A.	Management science methodologies in environmental management and sustainability: Discourses and applications	2011	62	9	S
Xing Y., Dangerfield B.	Modelling the sustainability of mass tourism in island tourist economies	2011	62	9	H
Syntetos A.A., Georgantzas N.C., Boylan J.E., Dangerfield B.C.	Judgement and supply chain dynamics	2011	62	6	H
Howick S., Eden C.	Supporting strategic conversations: The significance of a quantitative model building process	2011	62	5	S
Rouwette E.A.J.A.	Facilitated modelling in strategy development: Measuring the impact on communication, consensus and commitment	2011	62	5	S
Barlas Y., Gunduz B.	Demand forecasting and sharing strategies to reduce fluctuations and the bullwhip effect in supply chains	2011	62	3	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Vanderby S., Carter M.W.	An evaluation of the applicability of system dynamics to patient flow modelling	2010	61	11	S
Smits M.	Impact of policy and process design on the performance of intake and treatment processes in mental health care: A system dynamics case study	2010	61	10	H
Lebcir R.M., Atun R.A., Coker R.J.	System dynamic simulation of treatment policies to address colliding epidemics of tuberculosis, drug resistant tuberculosis and injecting drug users driven HIV in Russia	2010	61	8	H
Higgins A.J., Miller C.J., Archer A.A., Ton T., Fletcher C.S., McAllister R.R.J.	Challenges of operations research practice in agricultural value chains	2010	61	6	S
Maliapen M., Dangerfield B.C.	A system dynamics-based simulation study for managing clinical governance and pathways in a hospital	2010	61	2	H
Adamides E.D., Mitropoulos P., Giannikos I., Mitropoulos I.	A multi-methodological approach to the development of a regional solid waste management system	2009	60	6	H
Wyburn J., Hayward J.	OR and language planning: Modelling the interaction between unilingual and bilingual populations	2009	60	5	H
Syntetos A.A., Boylan J.E., Disney S.M.	Forecasting for inventory planning: A 50-year review	2009	60	S.1	N/A

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Tako A.A., Robinson S.	Comparing discrete-event simulation and system dynamics: Users' perceptions	2009	60	3	S
Kunc M.H., Morecroft J.D.W.	Resource-based strategies and problem structuring: Using resource maps to manage resource systems	2009	60	2	S
Saeed K., Pavlov O.V.	Dynastic cycle: A generic structure describing resource allocation in political economies, markets and firms	2008	59	10	H
Howick S., Whalley J.	Understanding the drivers of broadband adoption: The case of rural and remote Scotland	2008	59	10	H
Santos S.P., Belton V., Howick S.	Enhanced performance measurement using OR: A case study	2008	59	6	S
Newsome I.M.	Using system dynamics to model the impact of policing activity on performance	2008	59	2	S
Lane D.C., Husemann E.	System dynamics mapping of acute patient flows	2008	59	2	S
Kunc M.H., Morecroft J.D.W.	Competitive dynamics and gaming simulation: Lessons from a fishing industry simulator	2007	58	9	S
Paucar-Caceres A., Rodriguez-Ulloa R.	An application of Soft Systems Dynamics Methodology (SSDM)	2007	58	6	S
Evenden D., Harper P.R., Brailsford S.C., Harindra V.	Improving the cost-effectiveness of Chlamydia screening with targeted screening strategies	2006	57	12	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Swart J., Powell J.H.	Men and measures: Capturing knowledge requirements in firms through qualitative system modelling	2006	57	1	H
Chen J.H., Jan T.S.	A system dynamics model of the semiconductor industry development in Taiwan	2005	56	10	H
Olaya Y., Dyner I.	Modelling for policy assessment in the natural gas industry	2005	56	10	H
Akkermans H.A., Van Oorschot K.E.	Relevance assumed: A case study of balanced scorecard development using system dynamics	2005	56	8	S
Powell J.H., Coyle R.G.	Identifying strategic action in highly politicized contexts using agent-based qualitative system dynamics	2005	56	7	S
Taylor K., Dangerfield B.	Modelling the feedback effects of reconfiguring health services	2005	56	6	H
Bennett P., Hare A., Townshend J.	Assessing the risk of vCJD transmission via surgery: Models for uncertainty and complexity	2005	56	2	H
Jan T.-S., Hsiao C.-T.	A four-role model of the automotive industry development in developing countries: A case in Taiwan	2004	55	11	H
Adamides E.D., Stamboulis Y.A., Varelis A.G.	Model-based assessment of military aircraft engine maintenance systems	2004	55	9	H
Hung W.Y., Kucherenko S., Samsatli N.J., Shah N.	A flexible and generic approach to dynamic modelling of supply chains	2004	55	8	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Howick S., Eden C.	On the nature of discontinuities in system dynamics modelling of disrupted projects	2004	55	6	S
Brailsford S.C., Lattimer V.A., Tarnaras P., Turnbull J.C.	Emergency and on-demand health care: Modelling a large complex system	2004	55	1	H
Kleijnen J.P.C., Smits M.T.	Performance metrics in supply chain management	2003	54	5	H
Howick S.	Using system dynamics to analyse disruption and delay in complex projects for litigation: Can the modelling purposes be met?	2003	54	3	S
Hafeez K., Abdelmeguid H.	Dynamics of human resource and knowledge management	2003	54	2	H
Howick S., Eden C.	The impact of disruption and delay when compressing large projects: Going for incentives?	2001	52	1	S
Jan T.-S., Jan C.-G.	Designing simulation software to facilitate learning of quantitative system dynamics skills: A case in Taiwan	2000	51	12	S
Lane D.C.	Diagramming conventions in system dynamics	2000	51	2	S
Roberts C.A., Dangerfield B.C.	A strategic evaluation of capacity retirements in the steel industry	2000	51	1	H
Lane D.C., Monefeldt C., Rosenhead J.V.	Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department	2000	51	5	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Dyner I.	Energy modelling platforms for policy and strategy support	2000	51	2	H
Townshend J.R.P., Turner H.S.	Analysing the effectiveness of Chlamydia screening	2000	51	7	H
Jan T.-S., Jan C.-G.	Development of weapon systems in developing countries: A case study of long range strategies in Taiwan	2000	51	9	H
Barton P.M., Tobias A.M.	Discrete quantity approach to continuous simulation modelling	2000	51	4	H
Mashayekhi A.N.	Project cost dynamics for development policy-making	2000	51	3	H
Morecroft J.D.W.	System dynamics in Europe today: A review of professional infrastructure and academic programmes	1999	50	4	N/A
Coyle G., Morecroft J.	Part 1: System dynamics in the UK and continental Europe	1999	50	4	N/A
Davidsen P.I.	Graduate programmes in system dynamics at the University of Bergen, Norway	1999	50	4	N/A
Milling P.	System dynamics at Mannheim University	1999	50	4	N/A
Morecroft J.D.W.	System dynamics in MBA education at London Business School	1999	50	4	N/A
Anon.	Part 4: Methodological developments in the field	1999	50	4	N/A
Coyle R.G.	System dynamics at Bradford University: A Silver Jubilee Review	1999	50	4	N/A
Coyle J.M., Exelby D., Holt J.	System dynamics in defence analysis: Some case studies	1999	50	4	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Dangerfield B.C.	System dynamics applications to European health care issues	1999	50	4	H
Wolstenholme E.F.	Qualitative vs quantitative modelling: The evolving balance	1999	50	4	S
Coyle G., Morecroft J.	Part 3: Influencing people, policy and management education	1999	50	4	S
Delauzun F., Mollona E.	Introducing system dynamics to the BBC World Service: An insider perspective	1999	50	4	S
Richardson G.P.	Reflections for the future of system dynamics	1999	50	4	N/A
Warren K., Langley P.	The effective communication of system dynamics to improve insight and learning in management education	1999	50	4	S
Anon.	Part 2: Reaching into the broad policy arena	1999	50	4	N/A
Corben D., Stevenson R., Wolstenholme E.F.	Holistic oil field value management: Using system dynamics for 'intermediate level' and 'value-based' modelling in the oil industry	1999	50	4	H
Warren K.	Designing your growth path: An interview with Charles Farquharson	1999	50	4	N/A
Coyle R.G.	Simulation by repeated optimisation	1999	50	4	H
Williams T.M.	Seeking optimum project duration extensions	1999	50	5	H
Winch G.	Dynamic visioning for dynamic environments	1999	50	4	S
Larsen E.R., Lomi A.	Resetting the clock: A feedback approach to the dynamics of organisational inertia, survival and change	1999	50	4	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Buzacott J.A.	Dynamic inventory targets revisited	1999	50	7	H
Morecroft J.D.W.	Management attitudes, learning and scale in successful diversification: A dynamic and behavioural resource system view	1999	50	4	H
Larsen E.R., Bunn D.	Deregulation in electricity: Understanding strategic and regulatory risk	1999	50	4	H
Rodrigues A.G., Williams T.M.	System dynamics in project management: Assessing the impacts of client behaviour on project performance	1998	49	1	H
Lane D.C.	Can we have confidence in generic structures?	1998	49	9	S
Calinescu A., Efstathiou J., Schirn J., Bermejo J.	Applying and assessing two methods for measuring complexity in manufacturing	1998	49	7	H
Goodman M.R.	Study notes in system dynamics	1997	48	11	N/A
Lane D.C.	Invited reviews on system dynamics	1997	48	12	N/A
Richardson G.P., Pugh A.L., III	Introduction to system dynamics modeling with dynamo	1997	48	11	H
Randers J.	Elements of the system dynamics method	1997	48	11	H
Roberts N., Anderson D., Deal R., Garet M., Shaffer W.	Introduction to computer simulation: A system dynamics modeling approach	1997	48	11	H
Dangerfield B., Roberts C.	An overview of strategy and tactics in system dynamics optimization	1996	47	3	H
Wang S.	A dynamic perspective of differences between cognitive maps	1996	47	4	S

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Lane D.C.	On a resurgence of management simulations and games	1995	46	5	S
Williams T., Eden C., Ackermann F., Tait A.	The effects of design changes and delays on project costs	1995	46	7	H
Dyner I., Smith R.A., Peña G.E.	System dynamics modelling for residential energy efficiency analysis and management	1995	46	10	H
Lane D.C.	System dynamics practice: A comment on 'a case study in community care using systems thinking'	1994	45	3	N/A
Wolstenholme E.F., Corben D.A.	A hypermedia-based Delphi tool for knowledge acquisition in model building	1994	45	6	
Wolstenholme E.F.	A case study in community care using systems thinking	1993	44	9	S
Bunn D.W., Larsen E.R., Vlahos K.	Complementary modelling approaches for analysing several effects of privatization on electricity investment	1993	44	10	H
Coyle R.G., Gardiner P.A.	A system dynamics model of submarine operations and maintenance schedules	1991	42	6	H
Worthington D.	Hospital waiting list management models	1991	42	10	H
Roberts C., Dangerfield B.	Modelling the epidemiological consequences of HIV infection and aids: A contribution from operational research	1990	41	4	H
Keloharju R., Wolstenholme E.F.	A case study in system dynamics optimization	1989	40	3	H

(*continued*)

Table 4 (continued)

Authors	Title	Year	Vol.	Issue	Soft/ Hard
Smith A.R., Bartholomew D.J.	Manpower planning in the United Kingdom: An historical review	1988	39	3	H
Doyle Peter, Saunders John	Measuring the true profitability of sales promotions	1986	37	10	H
Coyle R.G.	Representing discrete events in system dynamics models: A theoretical application to modelling coal production	1985	36	4	H
Wolstenholme E.F., Coyle R.G.	The development of system dynamics as a methodology for system description and qualitative analysis	1983	34	7	S
Coyle R.G.	Who rules the waves? A case study in system description	1983	34	9	S
Wolstenholme E.F.	Modelling national development programmes: An exercise in system description and qualitative analysis using system dynamics	1983	34	12	S
Wolstenholme E.F.	System dynamics in perspective	1982	33	6	S
Jones J.W.	Application of a performance based approach to economic valuation of information	1981	32	11	H
Coyle G.	A model of the dynamics of the Third World War	1981	32	9	H
Bates T.	Some comments on system dynamics and J. A. Sharp's Paper 'System dynamic applications'	1978	29	5	H

References

Andersen, D.F., Vennix, J.A., Richardson, G.P. and Rouwette, E.A., 2007. Group model building: Problem structing, policy simulation and decision support. *Journal of the Operational Research Society*, 58, 691–694.

Akkermans, H.A. and Van Oorschot, K.E., 2005. Relevance assumed: A case study of balanced scorecard development using system dynamics. *Journal of the Operational Research Society*, 56(8), 931–941.

Checkland, P., 1985. Achieving 'desirable and feasible' change: An application of soft systems methodology. *Journal of the Operational Research Society*, 36(9), 821–831.

Coyle, R.G., 1999. Simulation by repeated optimisation. *Journal of the Operational Research Society*, 50(4), 429–438.

Coyle, G., 2000. Qualitative and quantitative modelling in system dynamics: Some research questions. *System Dynamics Review*, 16(3), 225.

Daellenbach, H.G., McNickle, D.C. and Dye, S., 2012. *Management Science. Decision Making Through Systems Thinking*, 2nd edition.

Dangerfield, B. and Roberts, C., 1996. An overview of strategy and tactics in system dynamics optimization. *Journal of the Operational Research Society*, 47(3), 405–423.

Evenden, D., Harper, P.R., Brailsford, S.C. and Harindra, V., 2006. Improving the cost-effectiveness of chlamydia screening with targeted screening strategies. *Journal of the Operational Research Society*, 57(12), 1400–1412.

Eden, C. and Ackermann, F., 2006. Where next for problem structuring methods. *The Journal of the Operational Research Society*, 57(7), 766–768.

Forrester, J.W., 1961. *Industrial Dynamics*. Cambridge: MIT Press.

Forrester, J.W., 1994. System dynamics, systems thinking, and soft OR. *System Dynamics Review*, 10(2–3), 245–256.

Forrester, J.W., 1987. Lessons from system dynamics modeling. *System Dynamics Review*, 3(2), 136–149.

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

Howick, S. and Whalley, J., 2008. Understanding the drivers of broadband adoption: The case of rural and remote Scotland. *Journal of the Operational Research Society*, 59(10), 1299–1311.

Kunc, M. and Kazakov, R., 2013. Competitive dynamics in pharmaceutical markets: A case study in the chronic cardiac disease market. *Journal of the Operational Research Society*, 64(12), 1790–1799.

Kunc, M.H. and Morecroft, J.D., 2009. Resource-based strategies and problem structuring: Using resource maps to manage resource systems. *Journal of the Operational Research Society*, 60(2), 191–199.

Lane, D.C., 1994. With a little help from our friends: How system dynamics and soft OR can learn from each other. *System Dynamics Review*, 10(2–3), 101–134.

Lane, D.C., 2000. Should system dynamics be described as a 'hard' or 'deterministic' systems approach?. *Systems Research and Behavioral Science*, 17(1), 3–22.

Lane, D.C. and Husemann, E., 2008. System dynamics mapping of acute patient flows. *Journal of the Operational Research Society*, 59(2), 213–224.

Lane, D.C. and Oliva, R., 1998. The greater whole: Towards a synthesis of system dynamics and soft systems methodology. *European Journal of Operational Research*, 107(1), 214–235.

Larsen, E.R. and Lomi, A., 1999. Resetting the clock: A feedback approach to trie dynamics of organisational inertia, survival and change. *Journal of the Operational Research Society*, 50(4), 406–421.

Morecroft, J.D., 1985. Rationality in the analysis of behavioral simulation models. *Management Science*, 31(7), 900–916.

Morecroft, J.D., 2015. *Strategic Modelling and Business Dynamics: A Feedback Systems Approach*. John Wiley & Sons.

Morecroft, J.D.W., 1995. Management attitudes, learning and scale in successful diversification: A dynamic and behavioural resource system view. *Journal of the Operational Research Society*, 50(4), 315–336.

Morecroft, J.D.W., 2002. Resource management under dynamic complexity. Chapter 2 in Systems Perspectives on Resources, Capabilities and Management Processes (editors Morecroft, Sanchez and Heene), Advanced Series in Management.

Olaya, Y.R.I.S. and Dyner, I.S.A.A.C., 2005. Modelling for policy assessment in the natural gas industry. *Journal of the Operational Research Society*, 56(10), 1122–1131.

Paich, M., Peck, C. and Valant, J.J., 2004. *Pharmaceutical Product Strategy: Using Dynamic Modeling for Effective Brand Planning*. CRC Press.

Paucar-Caceres, A., 2011. The development of management sciences/operational research discourses: Surveying the trends in the US and the UK. *Journal of the Operational Research Society*, 62(8), 1452–1470.

Rahmandad, H., Oliva, R., Osgood, N.D. and Richardson, G., 2015. *Analytical Methods for Dynamic Modelers*. MIT Press.

Rouwette E.A.J.A., 2011. Facilitated modelling in strategy development: Measuring the impact on communication, consensus and commitment. *Journal of the Operational Research Society* 62(5): 879–887.

Scott, R.J., Cavana, R.Y. and Cameron, D., 2015. Interpersonal success factors for strategy implementation: A case study using group model building. *Journal of the Operational Research Society*, 66(6), 1023–1034.

Sterman, J.D., 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3), 321–339.

Sterman, J.D. and Wittenberg, J., 1999. Path dependence, competition, and succession in the dynamics of scientific revolution. *Organization Science*, 10(3), 322–341.

Syntetos, A.A., Georgantzas, N.C., Boylan, J.E. and Dangerfield, B.C., 2011. Judgement and supply chain dynamics. *Journal of the Operational Research Society*, 62(6), 1138–1158.

Tako, A.A. and Robinson, S., 2009. Comparing discrete-event simulation and system dynamics: Users' perceptions. *Journal of the Operational Research Society*, 60(3), 296–312.

Vennix, J.A.M. and Rouwette E.A.J.A., 2000. Group model building. What does the client think of it now? *Proceedings of 2000 International System Dynamics Conference*. System Dynamics Society, Chestnut Hill.

Wolstenholme, E.F., 1999. Qualitative vs quantitative modelling: The evolving balance. *Journal of the Operational Research Society*, 50(4), 422–428.

Warren, K., 2002. *Competitive Strategy Dynamics*. Chichester: Wiley.

Xing, Y. and Dangerfield, B., 2011. Modelling the sustainability of mass tourism in island tourist economies. *Journal of the Operational Research Society*, 62(9), 1742–1752.

Part I

Applications of System Dynamics in Management

Resetting the Clock: A Feedback Approach to the Dynamics of Organisational Inertia, Survival and Change

E.R. Larsen and A. Lomi

Introduction

Studying organisational change requires—almost by definition—a commitment to the analysis of dynamic processes and disequilibrium states. Yet, with few exceptions, our understanding of organisational change processes has have not progressed much beyond static (or comparative static) frameworks in which strategic change is seen as an almost instantaneous transition from one equilibrium configuration to another, with surprisingly little attention given to the multiple adjustment paths that may connect the two states, and to the disequilibrium states likely to be encountered along the transition process [1–3].

Journal of the Operational Research Society (1999) **50**(4)

E.R. Larsen (✉)
Department of Management, Aarhus University, Aarhus, Denmark

A. Lomi
University of Bologna, Bologna, Italy

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_2

The emphasis on equilibrium at the organisational level, and on its individual-level counterpart—optimisation—rests on what March and Olsen [4] have termed the assumption of 'historical efficiency,' or the belief that observed organisational configurations are the result of some (possibly optimal) adaptation processes [5]. One of the consequences of the ubiquity of the historical efficiency assumption in the study of organizations is that: 'While there has been considerable progress in developing frameworks that explain differing competitive success at any point in time, our understanding of the dynamic processes by which firms perceive and ultimately attain superior market positions is far less developed' [6].

These considerations are at the core of a dynamic theory of organisations because it is well documented that organisational structures respond with significant delays to managerial attempts to modify their core features [7], and because competition strengthens both the focal firm and its rivals, resulting in a race where competitors consume their resources just to maintain their relative position [8]. Jointly considered, the presence of time delays in managerial responses, and the self-reinforcing quality of many competitive processes, can be taken as points of departure to explore a wide range of long standing theoretical problems related to the actual degree of responsiveness of organisational structures to managerial action [9]. In this process oriented perspective, the possibility of influencing the dynamics of strategic and organisational change hinges on the understanding of how organisational structures operate over time to defeat—or catalyse—the efforts of policy makers, managers, and planners aimed at reforming organisations. But how should the effect of managerial change attempts on actual organisational change and survival be conceptualised, given the tendency of organisational structures to absorb and dissipate part of the energies and resources devoted to change?

Research in organisational ecology instructs us that change is hazardous because failure may result both from the misperception of the need for change—and hence inaction—as well as the disruptions and uncertainties introduced by the process of change itself [10, 11]. The problematic relationship between strategy conception and execution on the one hand [12], and between strategy execution and its consequences on the other [13], is rooted in the observation that business organisations exhibit many of the characteristics of policy-resistant dynamical systems [3, 14, 15]. However, resistance to change does not necessarily imply that organisations never change, or even that change is infrequent. Organisations do change considerably over time [16] and at times they are able to do so rather creatively [17]. Organisational inertia—like performance—is a relative, rather than an absolute concept and questions arise about how fast established organizations can change to address current needs and capture—or build—new resources.

Against this general background, in this paper we take the theory of structural inertia proposed by Hannan and Freeman [10] as the starting point to develop a dynamic feedback model of organisational inertia and change. We use system dynamics (SD) to simulate the model, test its internal consistency, and explore the full dynamic implications of structural inertia theory. While the application of system dynamics to specific policy and management problems is not new [18–21], its potential as a method for building and testing organizational theories remains largely unexplored [22]. Relatively few examples are available of SD as an aid to theory building and theory testing. One such example is Sterman's [23] attempt at formalising and test Kuhn's theory of scientific revolutions. Another example is Hanneman et al. [22] recent development of a model of state legitimacy and imperialist capitalism. A third example is Sastry's [24] reconstruction of earlier conceptual work on convergence and upheaval in processes of organisational change. Other examples of SD concepts applied to theory building and model conceptualisation include Masuch's [19] work on vicious circles in organisations as particular instances of positive feedback processes, and Hall's early work [18] on the dynamics of organisational pathologies. However, many other cases can be identified in which explanations for particular institutional and competitive phenomena in the organisational world are proposed that hinge implicitly on SD arguments [25, 26]. We find this relatively infrequent application of SD methods to theory building in organisational research surprising mainly because SD methods provide excellent opportunities to: (i) formalise propositions expressed in natural language within more articulated theoretical frameworks, while maintaining the richness and ambiguity of social theories and testing their dynamic consistency; (ii) explore the implications of alternative ways in which theoretical propositions might be linked, and (iii) go beyond the unconvincing image of theory testing as the examination of a series of sequential single-proposition statements about complex social and organisational processes.

Our general goal of this paper is to illustrate the value of system dynamics as a method for theory building and testing in the context of a central debate in current organisational research. A second—and somewhat narrower—objective of our paper is to present a modeling framework that may help organisational theorists and analysts to overcome some of the specification problems typical of empirical research on organisational survival and change inspired by ecological theories of organisations. In this line of empirical work on the causes and consequences of organisational change, it is not always easy to separate dependent from independent variables, and estimation of complete models is often problematic [1]. As, a consequence, many of the complexities arising from the dynamic nature of the theory of structural inertia need to be greatly simplified in order to arrive at estimable statistical models. With this work we hope to be able to establish a structured framework that will help to improve our understanding of the dynamic organisation-level processes that regulate the vital dynamics of individual organisations and that shape the evolution of organisational populations over long periods of time.

A Feedback View of Organisational Inertia and Change

Structural Inertia Theory

Starting from the notion of organisations as change-resistant complex systems for which structural change is at least as risky as stasis, the theory of structural inertia originally proposed by Hannan and Freeman [10] provides a structured framework for thinking about how processes of organisational change unfolds. The theory identifies reliability and accountability as the primary sources of survival advantage for modern complex organisations. Reliability means that organisations are rewarded for reducing the variability of the product or services supplied, and for fulfilling customers' expectations in terms of quality, timing and prices of products. Accountability means that organisations are rewarded for their ability to document how their resources are allocated, and for convincing members, investors and clients of the procedural rationality of the decisions behind specific outcomes. Reliability and accountability are high when organisational goals are institutionalised and activities routinised, but institutionalisation and routinisation also generate inertial pressures because they encourage replication and exploitation of existing competencies [27].

According to Hannan and Freeman [10] structural inertia is not constant over the organisational life-course, but varies systematically with age and size. Specifically, organisational reliability and accountability are assumed to increase monotonically with size and age. Given that resistance to change also moves in the same direction of reliability and accountability over time, it follows that the probability of change decreases as organisations grow older and presumably bigger. Figure 1, taken from Kelley and Amburgey [28] illustrates the basic logic behind the theory of structural inertia as a series of dyadic connections among the core theoretical constructs. The empirical specification of dynamic models of organisational survival and change typically conforms to this sequential linear structure which allows—at least in principle—each individual causal links between the independent and the dependent variable to be empirically assessed given data on a suitable number of organisational life histories.

If we accept it as plausible, the theory of structural inertia has two main counterintuitive implications for our understanding of organisational change. The first is that the same characteristics that give organisations a

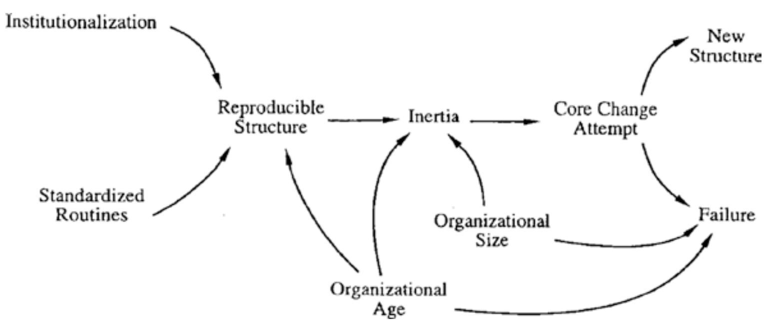


Fig. 1 Structural inertia theory as a series of individual propositions [28]

survival advantage also make them more resistant to change. It follows that selection processes tend to favour organisations that are relatively inert. This conclusion is clearly at odds with the suggestions offered by textbook views of organisational change that tend to see flexibility—rather than inertia—as the key to organisational performance and long-term survival. The second implication is that organisational change is risky in and for itself because it disrupts the routines in which organisational memory and competencies are stored [29], and calls into question the (internal and external) bases of institutionalisation and legitimation [30, 31]. As a consequence, organisations in the process of fundamental transformation are 'between a rock and a hard place' in the sense that change processes themselves may increase organisational failure rates independent of their content, that is, of whatever organisational characteristics are being changed [11, 32]. To the extent to which young organisations are exposed to a 'liability of newness', the tendency of organisational failure rates be higher in the early stages of organisational life and to decline with age [33], major structural changes imply that established organisations may once again be exposed to the causes of failure typical of young organisations, like, for example, the need to establish a framework of trust within which strangers can cooperate and agree on the appropriate sanctions for opportunistic behaviour, and the lack of consistent solutions to routine problems [31]. In this sense, major organisational changes can be said to 'reset the clock' that regulates the vital dynamics of individual organisations and increase, at least temporarily, the hazards of failure [11, 30].

To date, only few empirical studies are available that explored the effects of change and 'resetting the clock' on organisational survival. This situation reflects both the relative novelty of the framework, as well as the problems related to the identification and estimation of possible underlying statistical models that can disentangle the individual effects of change contents and processes [1]. According to the theory, process effects of change on organisational mortality are positive, but content effects can be negative *or* positive, hence it is difficult to estimate the individual effects due to the former (process of change) while holding constant those induced by the latter (content of change). As it could perhaps be expected under these circumstances, received empirical evidence is mixed: while one study supports selected aspects of the theory like, for example, age dependence in failure rates [11], other studies report change effects that are strongly contrary to the theoretical predictions [28, 34].

The theory of structural inertia can be seen as a dynamic theory with multiple feed back loops that are both explicit as well as implicit in the original formulation. On the one hand, the presence of multiple feedbacks is fully consistent with the process view or organisational change underlying the theory, but on the other hand, the sequential version of the model used in empirical studies (and summarised in Fig. 1) does not adequately capture the complexity of the relationship between organisational inertia, change and survival implied by the theory. In the following section, we present a fully dynamic version of the structural inertia model in order to (i) explore the internal consistency of the underlying theory; (ii) understand the dynamic feedback structure behind processes of organisational change; and (iii) explore the relation between organisational change, experience and survival in order to clarify some fundamental organisational level process that may be consistent with what we know about the ecological dynamics of organisational populations.

Resetting the Clock: A Feedback Model

The causal loop diagram in Fig. 2 connects three of the central concepts in the theory of structural inertia: inertia, performance reliability and change attempts. Inertia affects change attempts negatively, and change attempts decrease reliability because they disrupt internal and external networks in which organisations are embedded [11]. Finally, due to the

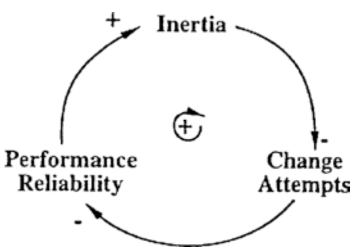


Fig. 2 A positive feedback loop which increases (or decreases) inertia

high degree of replicability and routinisation needed to stabilise performance over time [10], reliability increases organisational inertia. This first loop implies that a positive feedback process is at work to increase the level of organisational inertia over time. According to the theory, structural inertia reduces the number of change attempts, and this will result in higher reliability, that is, improved ability to reproduce past behaviour. But reproducibility induces further increases in inertia. The dynamic behaviour generated by the positive feedback process implied by this first causal loop is exponential growth.

But what are the limiting factors that prevent organisational inertia from increasing indefinitely? It is not easy to find an explicit answer to this question in the ecological literature on organisational change. The causal loop diagram reported in Fig. 3 illustrates a negative feedback process that may possibly limit the accumulation of structural inertia over time. According to the diagram, as inertia increases the likelihood of successful change becomes smaller. In turn, prolonged periods of stasis will increase the pressure for change in the organisation. As pressure for change increases, it is reasonable to expect that at least some new change attempts will be made. According to the theory, repeated attempts at changing organisational structures and processes decrease reliability and reset the internal organisational 'age clock.'

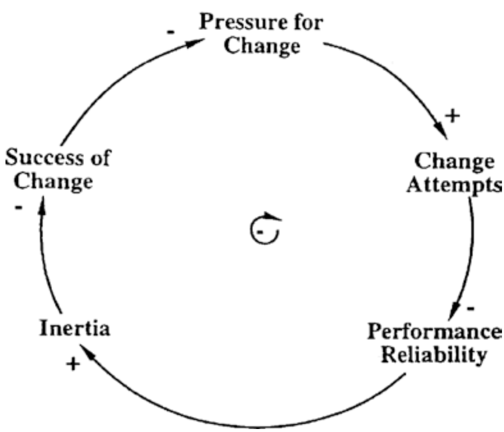


Fig. 3 A negative feedback loop which controls the growth of inertia

Complexity in dynamic feed-back models is introduced by the number of loops, and the way in which the loops are coupled. In fact, while individual loops may give rise to predictable dynamic behaviour (exponential growth or decline), it becomes almost impossible to predict the behaviour of a system with four or more feedback loops. In this sense, the theory of structural inertia can be viewed as a complex statement about the processes responsible for organisational change, resistance to change, and failure. The empirical studies available have tested individual propositions derived from the theory, but in order to understand its full dynamic implications, we have to represent the theory of structural inertia as a system of interdependent statements, that is as a system of equations rather than a sequence of separate 'hypotheses.' Figure 4 contains the complete feedback representation of the theory of structural inertia that we develop in this paper. In the figure, the letters and numbers reported on the directed lines connecting the variables refer to the assumptions (*A*) and the theorems (*T*) as they are, respectively, imposed and derived in the original formulation of the theory [10].

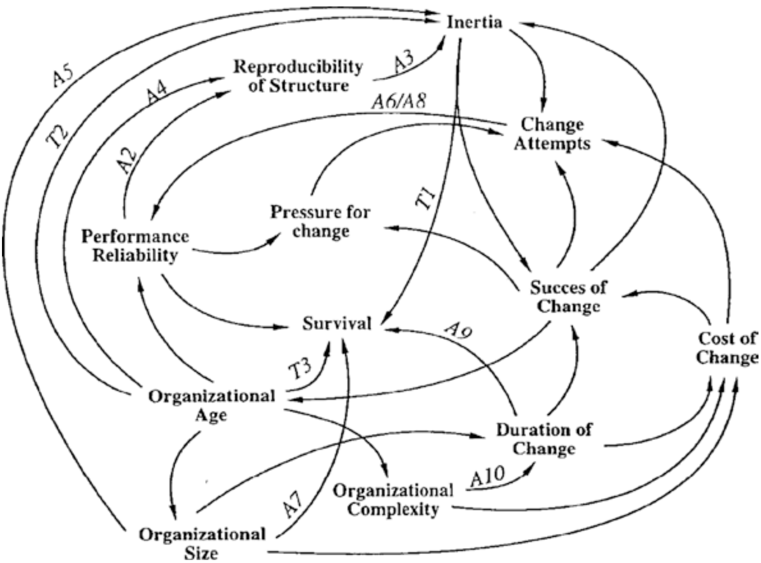


Fig. 4 Complete conceptual feedback diagram of structural inertia

For example, *A5* placed along the line connecting *Organisational size* and *Inertia* indicates that according to *Assumption 5* in the theory: 'The level of structural inertia increases with size for each class of organisation.' (Hannan and Freeman [10], p. 158). Similarly, *T*3 placed on the line going from *Organisational age* to *Survival* indicates that, according to *Theorem 3* in the theory: 'Organisational death rates decrease with age' (Hannan and Freeman [10], p. 157). In Fig. 4 we note that all of the ten assumptions underlying the theory are represented, with the only exception of assumption 1 according to which 'Selection in populations of organisations in modern societies favours forms with high reliability of performance and high levels of accountability' ([10], p. 154). This assumption that in a fundamental way, motivates the whole ecological theory of structural inertia and change, is not directly representable because it is based on the population-level concept of 'selection,' while in this paper we concentrate on organisational-level processes. In other words, in our model we try to specify possible firm-level processes that are consistent with the macrolevel relationship between accountability, reliability and selection observed in the study of the dynamics of organisational populations.

Finally, we note that all of the 5 theorems in the theory are represented, even if theorems 4 and 5 can be derived only indirectly. Theorem 4 (according to which 'Attempts at reorganisation increase death rates' [10], p. 159) is represented indirectly because we portrayed the effects of *Change attempts* on *Survival* as mediated by the possible deterioration in *Performance reliability* that may be induced by reorganisation. In other words, our representation does not rule out *a priori* the possibility of beneficial content effects of change. Similarly, theorem 5 (according to which 'Complexity increases the risk of death due to reorganization' [10], p. 162) is also represented indirectly in Fig. 4 because we were reluctant to specify *Complexity* as a *direct* cause of mortality for organisations that are undergoing change, and because the theoretical literature offers conflicting suggestions on the relation between organisational complexity and performance [35–37]. Rather, we saw *Complexity* operating on organisational mortality through intermediary factors such as the *Duration* or *Cost of change.*

Obviously, at this level of generality nothing is being said about exactly how the different concepts in the theory are related, that is, about the functional form of the relationship among variables. This will be done in the next section in which we formalize the feed-back structure of the theory of structural inertia, and translate it into a system of difference equations represented as a series of interlinked stocks and flows diagrams. This will be the last step needed before the actual simulation of the dynamic behaviour of the system.

From Feedback Loops to Dynamic Models

Before moving on to the detailed description of model specification, it is important to emphasise that our goal is *not* to provide a realistic model of inertia and change in a specific (or even 'representative') organisation, but rather to provide a system dynamics model of a *theory* of organisational inertia and change. For this reason we are searching for a minimal model specification that may allow us to explore the dynamic implications of the theory, and test its internal consistency. With this goal in mind, in this section we will discuss the critical parts in the formulation of a relatively small system dynamics model of inertia and change. The final model will contain 31 variables expressed as a system of differential equations. The reduced-form of the model includes 5 differential equations, but in this form, each individual equation would not lend itself easily to interpretation.

As Fig. 5 illustrates, structural *Inertia* is formulated as a stock (or 'accumulator') variable. In practice, this means that inertia can be accumulated

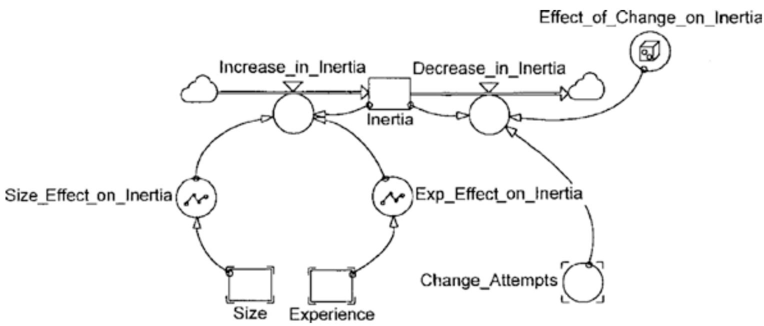


Fig. 5 The formulation of inertia in the model

over time; it can both increase, as well as decrease depending on the dynamics of the two corresponding flow variables (*Increase in Inertia* and *Decrease in Inertia*) indicate. *Inertia* is measured in dimensionless units through an index function. In the version of the model that we present, we assume that change attempts are intendedly adaptive and have the basic objective of decreasing structural inertia, that is making the organisation more responsive to changes in whatever contingencies management considers relevant. In practice, this goal may or may not be achieved depending on the effects of a number of other factors. These other factors are captured by the variable called *Effect of Change on Inertia*. This is modelled as a normally distributed stochastic term that determines the actual magnitude and direction of the impact of change attempts on structural inertia, which may range from almost null (change attempts have no implications for inertia), to strongly negative (change attempts decrease inertia, and reset the organizational age clock).

In keeping with the original formulation of the theory, structural inertia is affected by organisational age and size. To model these effects we use what is referred to as 'graphic converters' which specify the functional relationship between age, size and organisational inertia as a graph function. These qualitative relationships are represented in Fig. 6. Graphic converters make it exceedingly simple to test the modelling implications

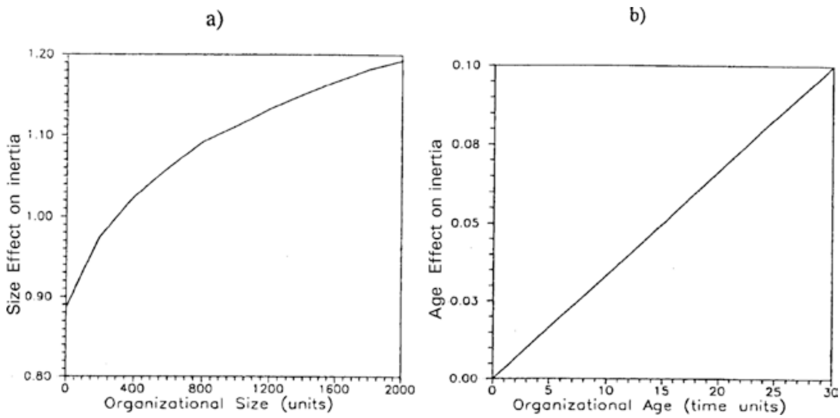


Fig. 6 Two examples of graph functions in the model, (**a**) the relationship between size and inertia, and (**b**) the relation between age and inertia

of a variety of functional links between age, size and organizational inertia—a problem that invariably comes up in empirical research, but one that does not have a direct empirical solution. The consequences of non-proportional effects of size are shown in Fig. 6a which implies some sort of diminishing effect of size on inertia. Simply put, an increase in organisational size from, say 10–20 has a larger effect on inertia than an increase from 1200–1210. The implications of a simple linear relationship are reported in Fig. 6b.

In Fig. 7 we illustrate how the pressure for change—represented as a stock variable—builds up and how it eventually generates change attempts. We assume that *Pressure for Change* increases when there is a gap between the *Expected* and *Actual Reliability*. Any difference between the expected and actual reliability will cumulate into additional units of *Pressure for change*. *Expected Reliability* is modeled as a combination of the expected *Trend in Reliability* (a terms which implies that past accomplishments provide at least some information about future accomplishments) and a 'stretch' parameter. The 'stretch' parameter indicates how much the organisation is expected to improve its *Reliability*, independent of past performance.

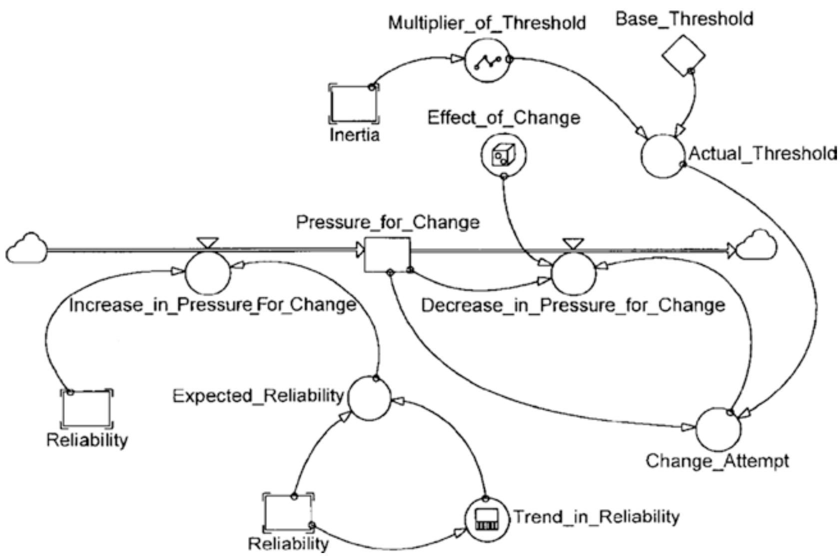


Fig. 7 The formulation of pressure for change and change attempts in the model

In Fig. 7, *Change Attempts* are modelled as a threshold function according to which change attempts will be made whenever the pressure for change reaches a given threshold value. However, the value of the threshold is itself dynamic, and fluctuates over time around a given 'base threshold.' The actual value of the base threshold changes as inertia increases so that the pressure for change needs to increase further to trigger new change attempts. *Change Attempts* work to release some of the accumulated *Pressure for Change* through the outflow *Decrease in Pressure for Change* How effective a change attempt is depends on a number of external and internal circumstances. As before, we use a stochastic variable called *Effect of Change* to reflect this basic indeterminacy in processes of organisational change. Some change attempts will be very successful and pressure for change will drop dramatically, while other attempts might do very little for releasing pressures for change.

Figure 8 shows the dynamics of reliability. *Reliability* is a complex construct presented as the joint consequence of routinisation, formalisation and institutionalisation [10]. To make the concept of reliability more specific, we simply model it as the inverse of 'variability,' which is itself a function of organisational experience and size, plus an exogenous baseline variability level that is always present in organisations. As the organisation grows older, gains experience and becomes larger the initial variability in production activities and quality decreases due to routinisation and

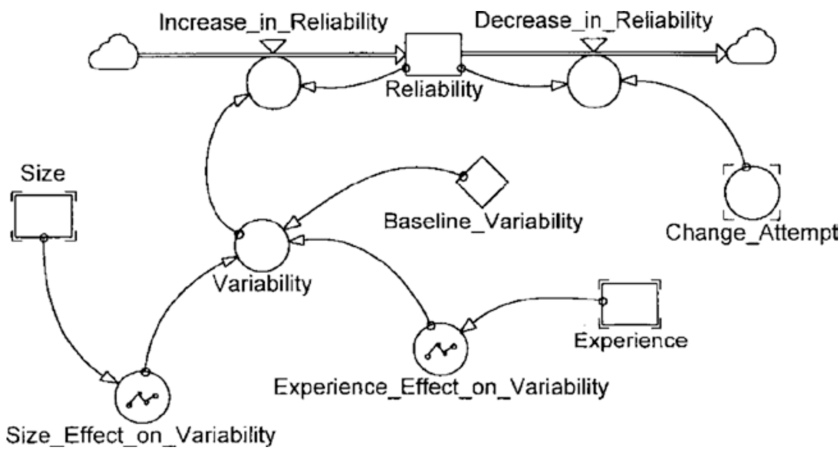


Fig. 8 The formulation of reliability

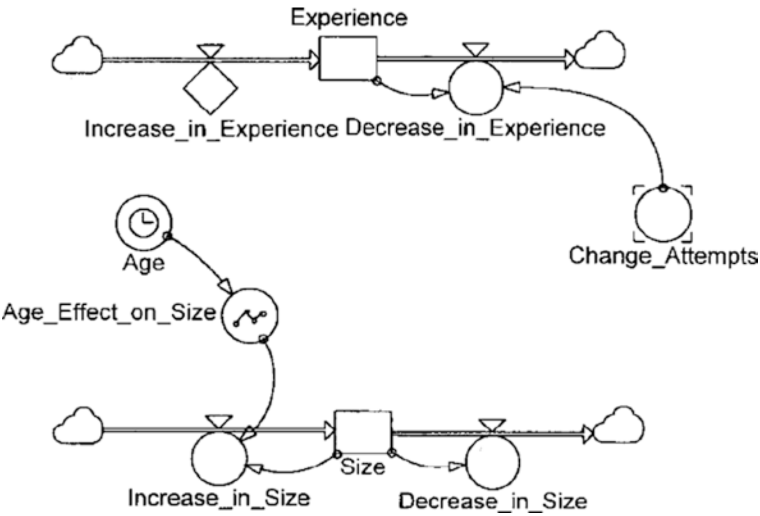


Fig. 9 The formulation of organisational experience and size

learning. As before we use 'graphic converters' to specify the qualitative relationship between the value of experience, size and variability.

In Fig. 9, *Organisational Experience* and *Size* are modelled as stock variables that may increase or decrease over time. We assume a systematic connection between organisational size and age, represented in the model by the graphic converter called *Age Effects on Size*, in which age coincides with simulation time. We chose to define the organisational age clock in terms of value of *Experience* in order to avoid any confusion with the typical meaning of the word 'age.' In this sense, major organisational changes 'reset the clock' to the extent to which they make accumulated experiences, competencies and knowledge obsolete [38]. In keeping with the original formulation of the theory, we use a graphic converter to define organisational size as a monotonically increasing function of age [10], but a wide range of different assumptions about the functional form of this relationship could be formalised to capture specific effects related to processes of organisational learning.

Finally, Fig. 10 illustrates the complete structure of the model that we simulate and analyse below. The complete system of equations that is implied by the diagram contained in Fig. 10, and the specific numerical values used to initialise the system are reported in Appendix.

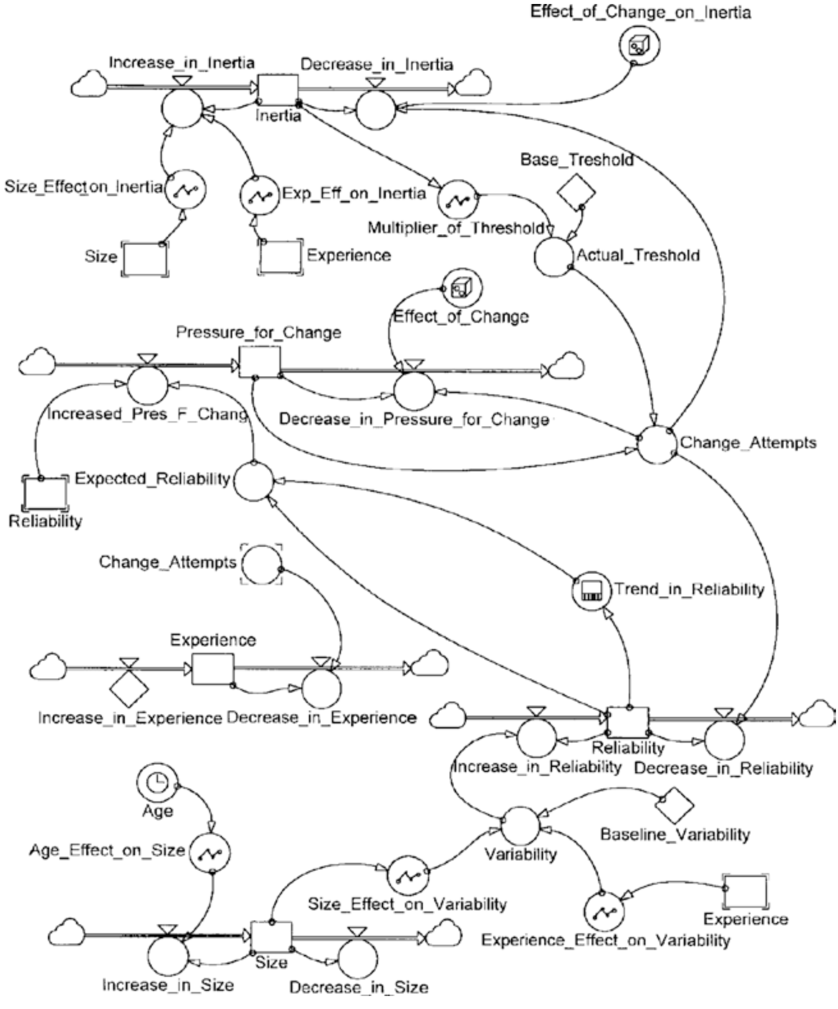


Fig. 10 The complete model

Methods

We rely on system dynamics (SD) to explore the qualitative dynamics of organisational inertia and change implied by ecological theories of organisations because we see three main advantages of SD in the analysis of organisational evolution in the context of the theoretical tradition in which the paper is rooted. Firstly, nonlinearities and disequilibrium assumptions are easily incorporated into SD models and this makes it possible to test for a wide range of possible relations among the factors that underlie change processes. Secondly, SD models allow to analyse the role of time delays explicitly, therefore allowing to explore different ways in which inertia operates on organisational structures. Thirdly, the focus on feedback processes in which individual variables are embedded makes SD particularly useful as a way of representing situations characterised by a systematic interdependence among co-occurring causal factors. While many contemporary theories of organisations suggest that nonlinearities, disequilibrium states, delayed effects, and feedback processes should be at the heart of our understanding of organizations and institutions [4, 5, 39], empirical research is often unable to sustain the analytical complexities implied by these theoretical suggestions.

Simulating Organisational Theories

There are three main motivations for relying on simulation rather than direct data analysis to explore the dynamics of the ecological theory of organisational inertia, survival and transformation. Firstly, the formulation of the hypotheses for the purpose of empirical data analysis encourages a fragmented and (comparative) static view of theoretical systems. As Sutton and Staw put it: '[H]ypotheses can be part of a well crafted theoretical argument […] but hypotheses do not (and should not) contain logical arguments about why empirical relationships are expected to occur' [40]. To our mind, this is precisely what makes computer simulation as useful as systematic empirical research for extending and testing organisational theories [41]. In a view of 'theory as narrative' [12], it makes little sense to extract and test individual propositions because what makes a theoretical narrative valuable is the way in which these propositions are interlinked. The statistical machinery used in empirical research is functional to what we can call a single-proposition approach to the study of organisations. This is unfortunate because it forces researchers to ignore what makes organisations an interesting and challenging object of study, such as for example, the lack of a clear *a priori* distinction and sequential dependence between dependent and independent variables, the presence of multiple time delays that characterise economic and social relations, and the complex structure of feedback processes in which the 'dependent' variables are embedded [43].

Secondly, theory discovery and testing by computer simulation is particularly useful when the existence of equilibrium points is less substantively important (and/or theoretically interesting) for the understanding of the social or economic system under study than the trajectories connecting these points and the speed at which the system converges to (or as the case may be, moves away from) specific equilibrium states [22]. Computer simulation becomes practically useful as a tool for theory building when: '[T]he guiding frame is that of a world of processes unfolding in time and flowing back upon each other' [22] that is when attention must be paid to the historical dynamics behind the observable outcomes of institutional and competitive processes. In this sense, simulation holds great promises for going beyond the unhelpful distinction between 'quantitative' and 'qualitative' approaches to the study of organisations, and therefore facilitating the composition of an artificial distinction that is one of the most enduring sources of disagreement about what exactly counts as 'theory' in organisational research [44].

Thirdly and more specific to the current work—a wide range of simulation methods are gaining legitimation as means of improving our understanding on key theoretical issues in organisational ecology research such as the relationship between diversity and competition [45], learning and evolution [46], adaptation and selection [47], and growth rates and organisational size distributions [48]. Computer simulation is also increasingly common as a means of testing the qualitative long-term implications of empirical estimates [5, 49] and as a tool for the rigorous development of theories concerning problems that resist direct empirical investigation like, for example the role of unobserved heterogeneity in organisational mortality rates [50, 51], the evolutionary implications of adaptive learning processes [52], and the role of micro-connectivity in the evolution of organisational populations [53]. Finally and very much in the spirit of the current work, ecological theories have been recently represented as computer models and evaluated by automatic theorem-provers through the development of logical formalism and language [54]. In most of these cases, computer simulation has helped to generate new theoretical insight, increase the coherence and focus of empirical research, and understand the qualitative implications of quantitative estimates. Obviously, these considerations do not imply that alternative ways of thinking about organisations cannot benefit from simulation in the same way, and perhaps even more. Rather, this literature suggests that ecological theories tend to be particularly good candidates for the development of simulation models due to their relatively high degree of formalisation, and the insistence of their proponents on comparability and cumulation of results across different studies as desirable properties of empirical research [55].

In closing, it may be worth mentioning that at no point in the present work we portray simulation as an alternative to well-crafted empirical research. Rather, we think of computer simulation as a way of exploring the dynamic implications of theoretical narratives, and therefore as a way of strengthening the link between organisational theories and history.

Results

Figure 11 shows the behaviour of the model during the first 20 time units periods. The *Variability* in the organisation decreases in an exponential fashion over time. In this specific case, the variability decreases almost by 50% over the first 20 time periods. As a direct consequence, reliability increases monotonically over the same period. *Expected Reliability* is a variable derived by extrapolation from *Reliability*.

Differences between *Reliability* and *Expected Reliability* create tensions in the organisation, this tension slowly is converted into *Pressure for Change* (Fig. 11d). The pressure for change eventually triggers *Change Attempts*. However, as the pressure for change gains momentum, inertia increases the *Threshold for Change*. Hence, change will happen depending on the relative speed at which these two quantities move over time. As Fig. 11e shows, inertia grows exponentially which implies that change is relatively easier to achieve in the early stages of organisational life, but becomes progressively harder as organizations age.

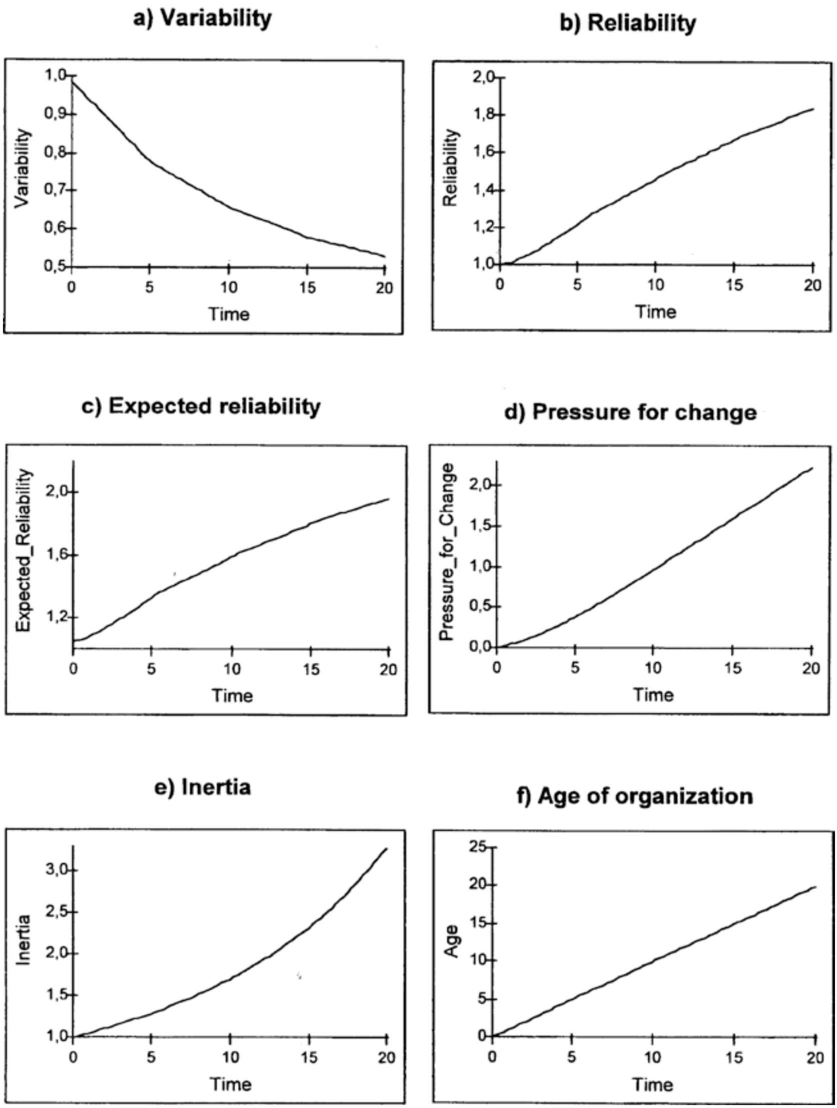


Fig. 11 Results from a 20 period simulation

Figure 12a–f illustrate the simulation results after 100 time periods. After the initial fall in variability and the corresponding increase in reliability in Fig. 12a, b, reliability tends to stabilise at relatively high levels

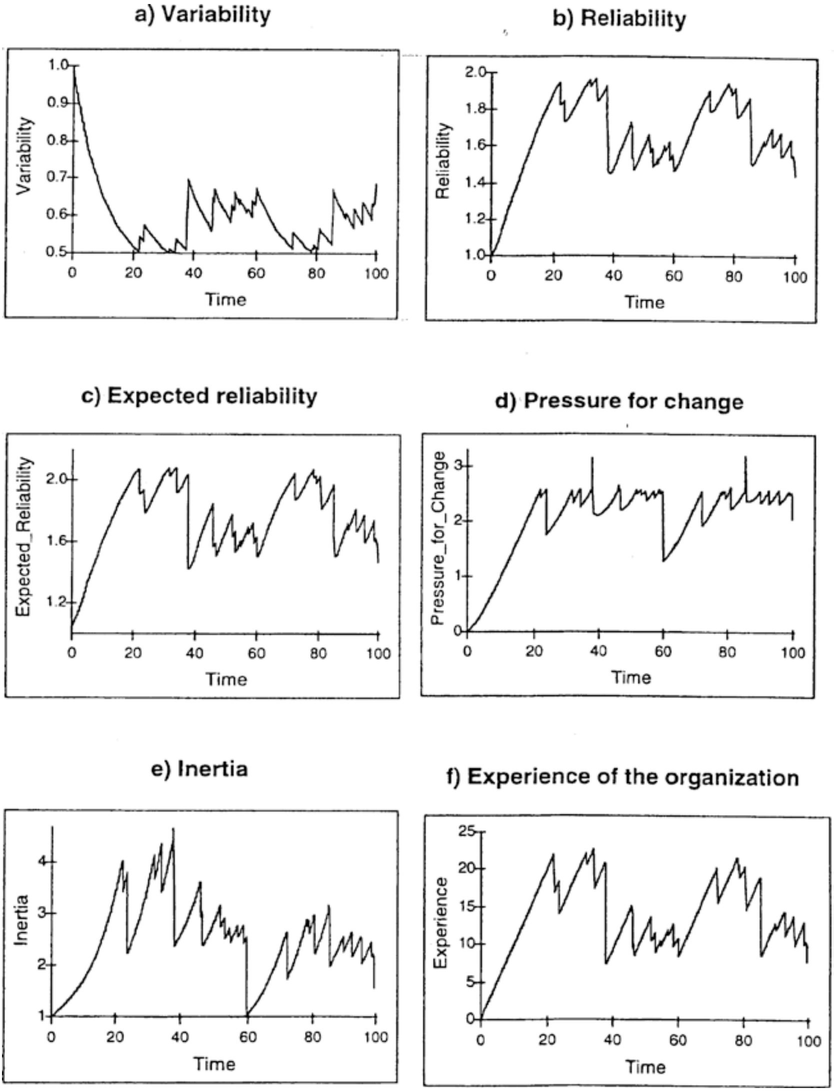


Fig. 12 Results from a 100 period simulation

for the following 20 simulation periods. During the same period inertia fluctuates between 2.5 and 4 with two identifiable long cycles—these figures being interpretable only in a relative sense, given that *Inertia* is a dimensionless index function by construction. As we reach period 40, a major change event drives inertia down, resulting in a sudden increase in variability (and corresponding drop in reliability). Interestingly, this change never had any long-term influence on the pressure for change, which remains roughly at the same level, despite the drop in inertia and accumulated experience in the organization. It is not before 60 time periods that another organisational change attempt reduces the pressure for change, in this case by more than a factor of two. Consequently, inertia drops dramatically from around 2.5 down to 1. At this point the organisation is almost back to its starting point, in the sense that inertia is reset to its initial value. During the process of change, the organisation has managed not to dissipate all the value of its previous experience, that decreased by factor of four between period 20–40.

From period 60–65 the organisation reaches the levels of experience and reliability that it originally had. As we mentioned before, pressure for change grows along with inertia, which means that the actual threshold also increases making change attempts less likely depending on the relative speed of these two quantities. Between period 65–80 we can observe a spell of relative stability (much like the period from 10–40). Around period 80, another major change attempt is taking place in the organisation, and again the value of experience drops to half of its previous value and reliability drops similarly. It is worth noticing that the organisation tends to get locked into quasi cyclical patterns of performance reliability over time. Obviously these cycles may not be observable for organisations whose survival threshold—defined in terms of an unobservable level of reliability above which the organisation is exposed to very high risks of failure—is sufficiently low.

We conclude our analysis by exploring the qualitative implications of alternative ways of representing the relationship between organisational experience and inertia or in other words between the accumulation of organisational competencies, and the tendency of the organisational structures in which these competencies are encoded to become more resistant to change over time. In Fig. 7 the *Trend in reliability (TIR)* operator captures the attitude of the organisation toward its own past performance (defined in terms of reliability), that is, defines the value of experience for the organisation. Formally,

$$TIR = \left[\frac{\mathrm{d}\left(\overline{R}\right)}{\mathrm{d}t} \frac{1}{\overline{\overline{R}}} \right]; \frac{\mathrm{d}\overline{R}}{\mathrm{d}t} = \frac{R - \overline{R}}{T\left(\overline{R}\right)},$$

where $\overline{R}$ is the average (or 'expected') reliability calculated as a first order exponential smoothing of the observed level of reliability (*R*), and $T\left(\overline{R}\right)$ is defined as '*Time to Average*,' a constant term that may vary significantly across organisations and that could be interpreted as the extent to which the organisation is subject to short term pressure on performance. If the value of $T\left(\overline{R}\right)$ is small, management will put a strong emphasis on most recent results, and consider them as a benchmark to evaluate current performance. As a consequence, as *actual* performance starts drifting away from *expected* performance, pressure for change will build up relatively fast forcing management to take immediate action. Figure 13a, b report the results of simulations in which $T\left(\overline{R}\right) = 1,3,$ and 5 respectively. All simulations performed previously assumed $T\left(\overline{R}\right) = 3$ (for details see the equation defining *Trend in Reliability* reported in Appendix).

During the first 20 periods, organisations characterised by different time orientations do not differ significantly. However, organisations with shorter-term orientations exhibit lower levels of inertia after period 40. As inertia decreases, the threshold for change decreases making future change more likely. Short-term pressures to meet performance expectations keep inertia low but tend to lock the organisation into a situation in which competencies are hard to build and preserve because change generates more change. As Fig. 13b illustrates, the level of competencies (or cumulated experience) of the organisation characterised by $T\left(\overline{R}\right) = 1$ is about half the level of competencies of the organisation in the baseline case (for which $T\left(\overline{R}\right) = 3$), and at times it drops to zero which correspond to a complete resetting of the clock that regulates organisational survival according to ecological theories of change. As expected, the opposite result obtain when we set $T\left(\overline{R}\right) = 5$. The main effect of a longer 'time drag' is to decelerate the cumulation of pressure for change. As a consequence, inertia will reach relatively high levels before change attempts become unavoidable. An interesting point to note is that an organisation that evaluates its current performance relative to performance levels

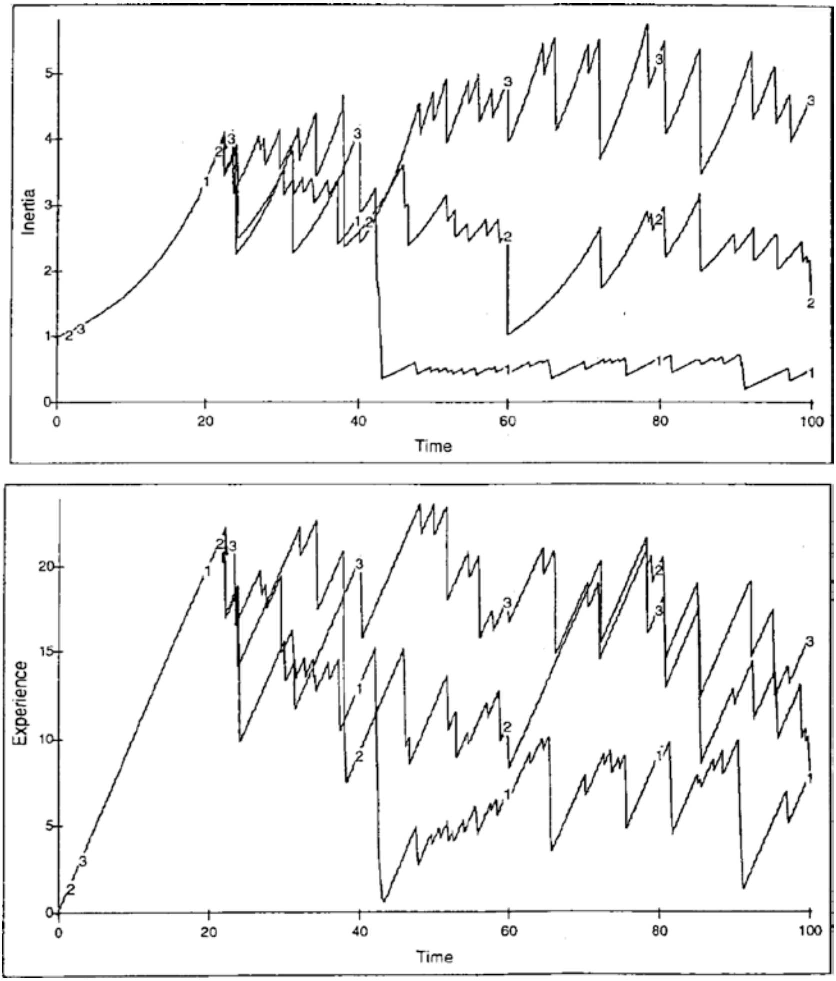


Fig. 13 The effect of time pressure on inertia and experience

reached in a less recent past (because $T(\bar{R}) = 5$) builds up competencies faster than a similar organisation, but with a shorter-term orientation (because $T(\bar{R}) = 3$). However, interorganisational differences in accumulated competencies tend to vanish after time $t = 60$, indicating the existence of an optimal (or 'ideal') level of resistance to organisational change. Below this point organisations change rapidly, but find it very difficult to stabilise their knowledge. Above this point, relatively high levels of organisational inertia are not compensated in the long run by a parallel accumulation of competencies and increase in reliability.

Discussion and Conclusions

According to one view of the organisational world, as organisations grow old and large they accumulate competencies, resources and knowledge that can be deployed to sustain and improve their competitive position. An alternative view suggests that as organisations grow old and large, their structures become progressively more vulnerable to processes of self-reproduction that dissipate resources and decrease their ability to respond adequately to the challenges of innovation and change posed by new rivals [56]. Which of these to views is more realistic depends on assumptions about *organisational inertia*, that is, about the relative speed (and cost) at which (i) organisational structures can be changed to address emergent needs; (ii) established organisations can move to occupy new resource spaces; and (iii) pre-existing corporate actors can generate and retain new resources internally. For these reasons the notion of structural inertia is central to our understanding of the dynamics of organisations and competition.

In this paper we concentrated on the part of population ecology theories of organisations that more directly deals with organisational inertia and change, and reformulated some of the central assumptions and propositions in system dynamics terms. We selected this specific theory of organisations because the clarity of its original formulation makes it particularly suitable to formalisation. One of the main motivations for translating the ecological theory of structural inertia into a system dynamics model was that empirical studies that have attempted to test the theory directly have been forced to ignore the complex feedback structure linking individual propositions for the purpose of specifying estimable statistical models. Perhaps the main motivation for the modelling exercise that we presented was our conviction that this 'single proposition' approach to organisational research greatly reduces the complexity, and intellectual value of theoretical narratives developed to account for relevant features of the organizational world. By using simulation methods we could exploit the rich dynamic feedback structure implicit in the original formulation to test its internal consistency, and explore the link between organisational inertia, the value of organisational experience and change.

The results reported suggest three main dynamic implications of the ecological model of organisational change. Firstly, organisational structures need to be in place before competencies can be created and resources built. In our models this conclusion is supported by the fact that structural inertia (which is linked to organisational size and age) builds up faster than organisational experience (which is dissipated—at least in part—by change attempts). We take this as supporting evidence for the proposition that routinisation of procedures, formalisation and investments in the other factors typically seen as determinants of organisational inertia, are needed *before* the organisation can exploit its knowledge, activate its resources, and build its competencies. In a world in which selection is based—at least in part—on reliability of performance, accountability of decision processes and reproducibility of structures, organisations that manage to reduce the variability of the products supplied or services rendered, stabilise their quality and fulfill customers' expectations in terms of timing and prices, may enjoy a significant competitive advantage over less reliable rivals. To the extent that reliability can be seen as a cumulative property of processes of exploitation of existing competencies [27], this result is broadly consistent with the claim that '[A]daptive processes characteristically improve exploitation *faster* than exploration. *These advantages of exploitation cumulate*. Each increase in competence at an activity increases the likelihood of rewards for engaging in that activity, thereby further increasing the competence and the likelihood' ([27], p. 73. Emphasis added). We could not find a more accurate description of our model of structural inertia as a dynamic positive feedback process, resulting both in the progressive cumulation of organisational competencies, as well as an improved ability to reproduce past behaviour.

Building on this insight, a second conclusion supported by our models is that inertia does not have the exclusive effect of making organisational structures less responsive to *external* stimuli [43]. As the level of structural inertia increases, the internal pressure for change obviously increases making change attempts more likely. But as inertia increases, the actual threshold for change also increases making change attempts less likely to succeed, at least for a given level of managerial effort. Hence inertia is a relative concepts not only because it implies a comparative assessment of the speed of organisational change and the speed of environmental change, but also—and perhaps mainly—because actual change depends on the *internal dynamics* of change attempts, and levels efforts needed to mobilise resources.

Because inertia acts as a multiplier of the threshold for change, as inertia decreases, for example, because the organisation is undergoing major transformations, the threshold for change decreases making future change more likely. This result is consistent with empirical evidence produced by studies of organisational mortality and change that have found that the probability of organisational change increases with the number of prior changes of the same type because processes or repetitive inertia operate both on stasis and change [11, 28]. For example, conditional on their age, Finnish newspaper organisations that changed the content and frequency of their publication at any time during the period 1771–1963 were shown to be more likely to experience similar types of change events again in the future [11].

Thirdly, alternative assumptions about managerial attitudes toward the value of experience have far-reaching implications for the dynamics of structural inertia and competence building in organizations, processes known to play a critical role in the evolution of organisational communities [38]. The results of the simulation experiments that we reported imply the existence of an 'ideal' level of resistance to change that allows organisations to build new resources and develop novel competencies, while simultaneously limiting fluctuations in the level of reliability, and reducing the rate of obsolescence of existing competencies. This conclusion is broadly consistent with current results in the area of organisational learning according to which organisations face an inescapable trade-off between processes of exploitation of old certainties, and processes of exploration of new possibilities [27]. According to our models, organisations in which the pressure for change builds up relatively fast as a consequence of deliberate managerial actions aimed at keeping the level of inertia low, do not find change particularly problematic because as inertia decreases the threshold for change decreases, and future change becomes more likely. However, while short-term pressures to meet performance expectations keep inertia low, they also tend to lock the organisation into a situation in which core competencies are hard to build and preserve because change generates more change. We believe that this result could provide a starting point for improving the quality of the theoretical debate on organisational change because it does not imply or assume that organisations are monolithically 'inert' or infinitely 'plastic.' Rather, this result clearly suggests that an optimal level of resistance to change exists below and above which organisational performance *can* be improved by stabilising existing routines or—as the case may be—disrupt them.

At the current stage our modelling efforts suffer from two main sets of limitations. The first is related to the fact that we presented a 'model of a model,' rather than a model of a specific organisational situation, or an empirically defined organisational problem. As a consequence the model reflects and in a way, accentuates the simplification of the underlying theoretical narrative and at no point we pursue the objective of improving its realism. A common criticism of 'models of models,' that is of more or less rational reconstructions of theoretical narratives, is that they do not so much reproduce the original theory as they reinvent it. This poses delicate problems of model validation [57, 58].

Clearly, many questions remain about the extent to which this problem is specific to system dynamics models, or—as we tend to believe—this represents a rather more general problem of interpretation of complex mental models [23, 59, 60]. In this respect, the main difference between models of theories and models of concrete processes seems to have less to do with the specific validation method that is appropriate, and more to do with the sources of information that trigger the model building exercise (informal—but 'locally informed'—mental models of 'managers' in the case of concrete processes, and formal—typically 'global' or structural but less detailed—mental models of 'theorists' in the case of theories).

A related problem typical of this kind of 'second order models' is that many elements of model specification may look arbitrary because organisational theories tend not to be developed in explicit dynamic terms and rarely specify exact functional forms that conceptual associations among variables ought to assume (i.e. they tend to lack a clear reference mode) [57].

In the analysis of real-world systems this problem is often circumvented by extracting patterns from historical data. In models of theories this solution is not readily available and the analyst is left with the task of extracting information on exactly how the variables are linked by interpreting theoretical narratives often expressed in natural language [57, 59]. The modelling framework that we adopted does not solve this problem, but graphic converters allow to test a wide range of possible functional relationships that may exist among high-level constructs like, for example, 'organisational size' and 'inertia,' which according to the theory are linked by a monotonically increasing function and explore their implications for the robustness of the theory. A detailed analysis of the sensitivity of the structural inertia model to alternative assumptions was far beyond the scope of the present study in which we concentrated on establishing a structured context within which issues of sensitivity and robustness can be addressed in the future. As we mentioned in the paper however, this problem does not have a direct empirical solution, although assumptions about specific functional forms have far reaching model specification implications in empirical research.

The second set of limitations—not entirely independent from the first—concerns issues of model validation, that is the assessment of the extent to which the range of dynamic behaviours produced by the model is consistent with what we know about actual organisations. Obviously, 'what we know' may take a variety of different forms including—but not necessarily limited to—numerical statistics. Accordingly there are several approaches to the validation of simulation models [61]. When modelling empirically observed processes, the problem of model validation can be addressed by analysing the extent to which the dynamic behaviour of the model reproduces history. Leaving aside questions about the adequacy of this intuitive and generally accepted way of validating simulation models, the theoretical nature of the underlying constructs prevented us to validate our models by direct comparison with history. Rather we took a 'link-by-link' approach to model validation by examining the conceptual arguments behind individual connections among variables. Hence, although our model is in broad qualitative agreement with the predictions made by ecological theories of organisational change, we cannot claim that were able to reproduce specific historical processes of organisational survival and transformation. More work and much bigger models are needed before we can extend the basic feed-back representation of the theory of structural inertia presented in this paper to include elements of realism grounded in a detailed understanding of specific organisational situations. This is likely to be the future direction that our research will take as we continue to explore new ways of designing models capable of capturing and representing the full dynamics implied by complex theoretical narratives about processes of organizational change.

Appendix: Model Equations and Documentation

Note that in a number of formulations below a time constant of 1 is assumed, but not made explicit in the model, to avoid 'cluttering' the model unnecessary with variables that have no influence (as they have the value of 1). In these cases the dimension is given as dimensionless/time. The equations below are in Powersim® format.

Due to the discrete nature of some of the rates in the model, the results presented will change when DT changes (DT can be interpreted as the organisational monitoring period). The results in this paper were obtained with DT = 0.125.

	Experience(t) = Experience(*t* – *dt*) – *dt** Decrease_in_Experience + *dt** Increase_in_Experience
	{Dimensionless. The accumulation of experience in the organisation}
init	Experience = 0
	{Dimensionless. Initial experience}
	Inertia *(t)* = Inertia(*t* – *dt*) – *dt** Decrease_in_Inertia + *dt** Increase_in_Inertia
	{Dimensionless. The accumulation of inertia in the organisation}
init	Inertia = 1
	{Dimensionless. Initial inertia in the organisation}
	Pressure_for_Change *(t)* = Pressure_for_Change (*t* – *dt*) – *dt** Decrease_in_Pressure_for_Change + *dt** Increased_Pres_F_Chang
	{Dimensionless. The accumulation of pressure for change in the organisation}

(*continued*)

(continued)

init Pressure_for_Change = 0
{Dimensionless. Initial pressure for change}
Reliability *(t)* = Reliability (*t* – *dt*) – *dt** Decrease_in_Reliability + *dt** Increase_in_Reliability
{Dimensionless. Accumulation of reliability in the organisation}
init Reliability = 1
{Dimensionless. Initial reliability of the organisation}
Size *(t)* = Size (*t* – *dt*) – *dt** Decrease_in_Size + *dt** Increase_in_Size
{Dimensionless. Size of the organisation}
init Size = 10
{Dimensionless. Initial size of the organisation}
Decrease_in_Pressure_for_Change = IF(Change_ Attempt > 0, Effect_of_ Change* Pressure_for_Change, 0)
{Dimensionless/ Time, The decrease in pressure for change in the organisation–Hidden time constant of 1}
Decrease_in_Experience = IF (Change_Attempt > 0.1, 3.8* Experience, 0)
{Dimensionless/Time. Decrease in experience, 3.8 is a scaling parameter which depends on the integration method, DT and how much experience can be lost in one organizational change organisation—Hidden time constant of 1}
Decrease_in_Inertia = IF (Change_Attempt > 0, Inertia * Effect_of_ Change_on_Inertia, 0)
{Dimensionless/Time, 0 is a parameter which determine the size of a change attempt that has to take place before inertia decrease organization—Hidden time constant of 1}
Decrease_in_Reliability = IF (Change_Attempt > 0.5, Reliability, 0)
{Dimensionless/Time. 0.5 is a parameter which determine the size of a change attempt that has to take place before reliability decreases organization—Hidden time constant of 1}
Decrease_in_Size = Size *0
{Dimensionless/Time. The model assumes that there is no direct decrease in size-organisation—Hidden time constant of 1}
Increase_in_Reliability = (1/Variability) – Reliability
{Dimensionless/Time organisation–Hidden time constant of 1}
Increase_in_Size = Size * Age_Effect_on_Size
{Dimensionless/Time organization—Hidden time variable Of 1}
Increase_in_Inertia = Inertia * Exp_Eff_on_Inertia * Size_Eff_on_ Inertia + 0.05
{Dimensionless/Time. 0.05 is assumed to be the steady accumulation of inertia that takes place in organizations—organization—Hidden time constant of 1}
Increased_in_Pres_F_Chang = Expected_Reliability – Reliability
{Dimensionless/Time. Increase in Pressure for Change organization—Hidden time constant of 1}
Actual_Threshold = Multiplier_of_Threshold * Base_Threshold

(*continued*)

(continued)

{Dimensionless}
Age = TIME
{Dimensionless, Age is equal to time in the model, as the organisation was created at time 0}
Age_Effect_on_Size = GRAPH(Age,0,3,[0, 0.171, 0.199, 0.183, 0.151, 0.097, 0.062, 0.041, 0.022, 0.009, 0.001 'Min: –0.1; Max: 0.2'])
{Dimensionless. The relationship between age and size}
Experience_Effect_on_Variability = GRAPH(Experience, 0, 5, [1, 0.79, 0.67, 0.59, 0.54, 0.51, 0.48, 0.46, 0.43, 0.42, 0.41 'Min: 0; Max: 1'])
{Dimensionless. The relationship between experience and variability}
Change_Attempts = IF (Pressure_for_Change > Actual_Threshold, 1, 0)
{Dimensionless}
Effect_of_Change = NORMAL(1.6, 2, 27363)
{Dimensionless. The stochastic effect of change given by a normal distribution}
Effect_of_Change_on_Inertia = NORMAL(3.2, 2, 27363)
{Dimensionless. The stochastic effect of change on inertia given by a normal distribution}
Exp_Eff_on_Inertia = GRAPH(Experience 0, 3, [0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1 'Min: –0.2; Max: 0.2'])
{Dimensionless. The relationship between experience and inertia}
Expected_Reliability = Reliability + Reliability * Trend_in_Reliability + 0.05 * Reliability
{Dimensionless. 0.05 is the baseline improvement to be expected by the management/shareholders per time unit}
Multiplier_of_Threshold = GRAPH(Inertia, 0, 0.3, [0.739, 0.767, 0.809, 0.896, 1.03, 1.125, 1.202, 1.261, 1.286, 1.289, 1.293 'Min: 0.7; Max: 1.5'])
{Dimensionless. The relationship between inertia and threshold}
Size_Eff_on_Inertia = GRAPH(Size, 0, 200, [0.886, 0.991, 1.025, 1.06, 1.093, 1.112, 1.133, 1.147, 1.167, 1.182, 1.193 'Min: 0.8; Max: 1.2'])
{Dimensionless. The relationship between size and inertia}
Size_Effect_on_Variability = GRAPH(Size, 0, 500, [0.99, 0.89, 0.82, 0.76, 0.71, 0.68, 0.66, 0.64, 0.62, 0.61, 0.61 'Min: 0; Max: 1'])
{Dimensionless. The relationship between size and variability}
Trend_in_Reliability = TREND(Reliability, 3, 1)
{Dimensionless. Trend is based on a 3rd order smoothing of reliability}
Variability = Base_Variability * Size_Effect_on_Variability * Age_Effect_on_Variability
{Dimensionless}
Increase_in_Experience = 1
{Dimensionless. In this model we assume that experience accumulates with a constant rate organisation—Hidden time variable of 1}
Base_Threshold = 2
{Dimensionless}

References

1. Barnett WP and Carroll G (1995). Modeling internal organizational change. *Ann Rev Sociol* **21**: 217–236.
2. Baron JN, Burton DM and Hannan MT (1996). The road taken: Origins and Evolution of employment systems in emerging companies. *Indust and Corporate Change* **51**: 239–275.
3. Lomi A, Larsen ER and Ginsberg A (1997). Adaptive learning in organizations: A system dynamics-based exploration. *J Mgmt* **23**: 561–582.
4. March JG and Olsen JP (1984). The new institutionalism: Organizational factors in political life. *Am Polit Sci Rev* **78**: 734–749.
5. Carroll G and Harrison J (1994). On the historical efficiency of competition between organizational populations. *Am J Sociol* **100**: 720–749.
6. Porter M (1994). Toward a dynamic theory of strategy. In R Rumelt, D Schendel and D Teece (Eds.) *Fundamental Issues in Strategy*. Cambridge, MA: Harvard Business School Press, pp 423–461.
7. Chen MJ and Macmillan IC (1992). Non response and delayed response to competitive moves: The roles of competitor dependence and action irreversibility. *Acad Mgmt J* **35**: 539–570.
8. Barnett WP and Hansen MT (1996). The red queen in organizational evolution. *Strat Mgmt J* **17**: 139–158.
9. Selznick P (1949). *TVA and the Grass Roots.* Berkeley, CA: University of California Press.
10. Hannan MT and Freeman J (1984). Structural inertia and organizational change. *Am Sociol Rev* **49**: 149–164.
11. Amburgey T, Kelly D and Barnett WP (1993). Resetting the clock: The dynamics of organizational change and failure. *Admin Sci Q* **38**: 51–73.
12. Senge PM and Sterman JD (1992). Systems thinking and organizational learning: acting locally and thinking globally in the organization of the future. *Eur J Opl Res* **59**: 137–150.
13. Sterman JD (1989). Modeling managerial behaviour: Misperception of feedback in a dynamic decision making experiment. *Mgmt Sci* **25**: 321–339.
14. Morecroft JDM, Larsen ER, Lomi A and Ginsberg A (1995). The dynamics of cooperation and competition for shared resources. *Sys Dynam Rev* **11**: 151–177.
15. Mosekilde E, Larsen E and Sterman JD (1991). Coping with complexity: Deterministic chaos in human decision making behaviour. In: J Casti and A Karlqvist (eds.), *Beyond Belief: Randomness, Prediction and Explanation in Modern Science*. Boston, MA: CRC Press.

16. Gresov C, Haveman H and Oliva T (1993). Organizational design, inertia and the dynamics of competitive response. *Organ Sci* **4**: 181–208.
17. March JG (1982). Footnotes to organizational change. *Admin Sci Q* **26**: 563–597.
18. Hall R (1976). A system pathology of an organization: The rise and fall of the Saturday Evening Post. *Admin Sci Q* **21**: 185–211.
19. Masuch M (1985). Vicious circles in organizations. *Admin Sci Q* **30**: 14–33.
20. Morecroft JDM (1988). System dynamics and microworlds for policymakers. *Eur J Opl Res* **35**: 301–320.
21. Morecroft JDM and Sterman JD (Eds) (1992). Modelling for learning. *Eur J Opl Res* **59**: 1–230.
22. Hanneman RA, Collins R and Mordt G (1995). Discovering theory dynamics by computer: experiments on state legitimacy and imperialist capitalism. *Sociol Methodol* **25**: 1–46.
23. Sterman JD (1985). The growth of knowledge: Testing a theory of scientific revolutions with a formal model, *Technol Forecasting and Social Change* **28**: 93–122.
24. Sastry MA (1997). Problems and paradoxes in a model of punctuated organizational change. *Admin Sci Q* **42**: 237–275.
25. Baum JAC and Singh JV (1994). Organization-environment coevolution. In: AC Baum and JV Singh (Eds.) *Evolutionary Dynamics of Organizations*. New York, NY: Oxford University Press, pp 379–402.
26. Brittain JW (1994). Density-independent selection and community evolution. In: AC Baum and JV Singh (eds.), *Evolutionary Dynamics of Organizations*. New York, NY: Oxford University Press, pp 355–402.
27. March JG (1991). Exploration and exploitation in organizational learning. *Organ Sci* **2**: 71–87.
28. Kelly D and Amburgey T (1991). Organizational inertia and momentum: A dynamic model of strategic change. *Acad Mgmt J* **34**: 591–612.
29. Nelson RR and Winter SG (1982). *An Evolutionary Theory of Economic Change.* Cambridge, MA: Belknap Press.
30. Hannan MT and Freeman J (1989). *Organizational Ecology.* Cambridge, MA: Harvard University Press.
31. Stinchcombe AL (1965). Social structure and organizations. In: JG March (Ed.) *Handbook of Organizations*. Chicago, IL: Rand McNally.
32. Haveman H (1992). Between a rock and a hard place. *Admin Sci Q* **37**: 48–75.

33. Freeman J, Carroll GR and Hannan MT (1983). The liability of newness: Age dependence in organizational death rates. *Am Sociol Rev* **48**: 692–710.
34. Delacroix J and Swaminathan A (1991). Cosmetic, speculative and adaptive organizational change in the wine industry: A longitudinal study. *Admin Sci Q* **36**: 631–661.
35. Carley K (1992). Organizational learning and personnel turnover. *Organ Sci* **3**: 20–46.
36. Krackhardt D and Stern R (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychol Q* **51**: 123–140.
37. Thompson J (1967). *Organizations in Action.* New York: McGraw Hill.
38. Tushman M and Anderson P (1986). Technological discontinuities and organizational environments. *Admin Sci Q* **31**: 439–465.
39. Tuma N and Hannan MT (1984). *Social Dynamics.* New York: Academic Press.
40. Sutton RI and Straw BM (1995). What theory is not. *Admin Sci Q* **40**: 371–384.
41. Masuch M (1995). Computer models. In: N Nicholson (Ed.) *Encyclopedic Dictionary of Organizational Behaviour*. London: Blackwell, pp 91–92.
42. DiMaggio PJ (1995). Comments on "what theory is not". *Admin Sci Q* **40**: 391–397.
43. Mollona E *and* Lomi A (1997). Three Experiments on the Organizational Dynamics of Resource Accumulation. *Working Paper*. London Business School.
44. Weick KE (1995). What theory is not, theorizing is. *Admin Sci Q* **40**: 385–390.
45. Levinthal D (1992). *Competitive Forces and Population Dynamics*. Unpublished manuscript, University of Pennsylvania, Philadelphia.
46. Mezias S and Lant T (1994). Mimetic learning and the evolution of organizational populations. In: J Baum and J Singh (Eds.) *Evolutionary Dynamics of Organizations*. New York, NY: Oxford University Press, pp 179–198.
47. Levinthal D (1990). Organizational adaptation: Environmental selection and random walks. In: J Singh (Ed.) *Organizational Evolution: New Directions*. Newbury Park: Sage, pp 201–233.
48. Hannan MT and Ranger-Moore J (1990). The ecology of organizational size distributions: A micro-simulation approach. *J Math Sociol* **15**: 67–90.
49. Hannan MT and Carroll G (1992). *Dynamics of Organizational Populations.* New York: Oxford University Press.

50. Levinthal D (1991). Random walks and organizational mortality. *Admin Sci Q* **36**: 397–420.
51. Petersen T and Koput K (1991). Density dependence in organizational mortality: Legitimacy or unobserved heterogeneity? *Am Sociol Rev* **56**: 399–409.
52. Bruderer E and Singh JV (1996). Organizational evolution, learning and selection: A genetic-algorithm-based model. *Acad Mgmt J* **39**: 1322–1349.
53. Lomi A and Larsen ER (1996). Interacting locally and evolving globally: A computational approach to the dynamics of organizational populations. *Acad Mgmt J* **39**: 1287–1321.
54. Peli G, Bruggerman J, Masuch M and O'Nualláin B (1994). A logical approach to organizational ecology. *Am Sociol Rev* **59**: 571–593.
55. Carroll G and Hannan MT (1995). *Organizations in Industry.* New York: Oxford University Press.
56. Barron J, West E and Hannan MT (1994). A time to grow and a time to die: Growth and mortality of credit unions in New York City, 1914–1990. *Am J Sociol* **100**: 381–421.
57. Wittenberg J (1992). On the very idea of a system dynamics model of Kuhnian science. *Sys Dynam Rev* **8**: 21–33.
58. Barlas Y (1992). Comments on "On the very idea of a system dynamics model of Kuhnian Science". *Sys Dynam Rev* **8**: 43–47.
59. Sterman JD (1992). Response to "On the very idea of a system dynamics model of Kuhnian science". *Sys Dynam Rev* **8**: 35–42.
60. Radziki MJ (1992). Reflections on "On the very idea of a system dynamics model of Kuhnian Science". *Sys Dynam Rev* **8**: 49–53.
61. Forrester JW and Senge PM (1980). Tests for building confidence in system dynamics models. In: AA Legasto, JW Forrester and JM Lyneis (Eds.) *System Dynamics.* Amsterdam: North-Holland, pp 209–228.

Management Attitudes, Learning and Scale in Successful Diversification: A Dynamic and Behavioural Resource System View

J.D.W. Morecroft

Introduction

Can a strategy of diversification lead to superior financial performance? Do diversified companies outperform firms that focus on a single core business? These are questions that have attracted much attention from strategy academics. They are also important questions for corporate leaders. Several decades of careful statistical research show that it is difficult to make a clear-cut empirical case for financial advantages or disadvantages from diversification (see [1, 2] for leads into this extensive area of strategy literature). It remains a puzzle why empirical studies are inconclusive despite persuasive arguments suggesting that financial performance should improve, especially in the case of related diversification. Without better understanding of this puzzle there is little verifiable practical advice that the research community can offer to business executives who are intent on pursuing a diversification strategy.

Journal of the Operational Research Society (1999) **50**(4)

J.D.W. Morecroft (✉)
London Business School, London, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_3

The motive for diversification often stems from dissatisfaction with the performance of the current business or the realisation that opportunities for improved performance may exist in other areas of business in which the firm does not currently compete, (see [3] Chapter VII on *The Economics of Diversification*). Central to management decision making is the judgement of whether new business is relatively more attractive (profitable) than the core business for any new investment the firm wants to undertake. At first glance such a judgement may seem straightforward, but in reality it is quite subtle. Who is to say when a portfolio is really outperforming a single business, based on what time series evidence, viewed across which time-span? Relative performance (and therefore investment) is subjective and changing over time. Could it be that researchers and executives are misled by the complexity of such dynamic multibusiness investment decisions? Certainly there is experimental evidence to show that people routinely underperform in dynamic decision-making tasks [4].

The tyre industry in the 1980s provides a good example of an industry that came to be viewed by its own executives as unattractive, thereby prompting a wave of diversification [5, 6], Between 1975 and 1985 total tyre demand fluctuated in the range 160–210 million tyres per year, There was very little growth, severe capacity shortages and surpluses developed, caused by long lead times on capacity expansion and an overhang of old technology bias-ply capacity held by producers unwilling to exit the industry. The result was damaging peaks in rivalry that depressed prices and spoiled firm profitability. The persistence of such adverse trading conditions over many years (coupled with corporate mindsets shaped by the pessimism of the post oil- shock economy) was sufficient to cause leading producers such as Uniroyal, Goodrich and Goodyear to curtail investment in the core tyre business and look for new businesses in which to invest. In all three cases diversification failed to produce sustainable performance advantages, and in Goodyear's case led to a hostile takeover bid.

Two models examine the aggregate policies that control investment in such situations.[1] Simulations show how the success of a diversification strategy depends both on business fundamentals (real profitability and relative scale of the core and non-core) and behavioural traits of typical corporate investment policy: target setting for financial performance, managers' expectations about performance of the core, managers' limited foresight and inherent optimism about the likely future performance of non-core business, and how quickly executives learn about performance in the light of experience, prior beliefs and confidence.

There are some surprising results. Measureable performance advantage from diversification takes a long time to achieve—up to a decade, even when management has picked a clear winner. Optimism and misplaced confidence can perpetuate a losing strategy, but fast learning (about true performance) can prematurely stunt a potential winner. The paper ends with implications of the results for practical diversification strategy and contemporary strategy research, emphasising the need for a dynamic and behavioural view of performance in diversifying firms.

Model of FocusCo, the Focuser

FocusCo is an imaginary tyre maker that is assumed to focus strictly on its core business, rather like Goodyear in the 1970s and early 1980s. Figure 1 shows the investment policy of FocusCo and the feedback loops in which it is embedded. The firm invests according to management's perception of return on the core business relative to an agreed benchmark return, as shown in bold on the right of the figure. The better the performance the more investment. If return exceeds the benchmark then resources in the core business grow which in turn fuels further investment as represented by reinforcing loop R1, because a bigger business invests more than a small one. This reinforcing growth process continues providing that return remains attractive and providing that the rate of investment in loop R1 exceeds the rate of resource depreciation in loop B1. Conversely, if return falls below the benchmark then the core business is starved of investment and resources fall.

Return itself depends on a variety of factors that reflect the feedback consequences of past investment decisions and industry conditions. These additional factors are shown in the remainder of Fig. 1. There is an important balancing loop B2 that results from the direct connection between resources and return. As resources grow then return on those

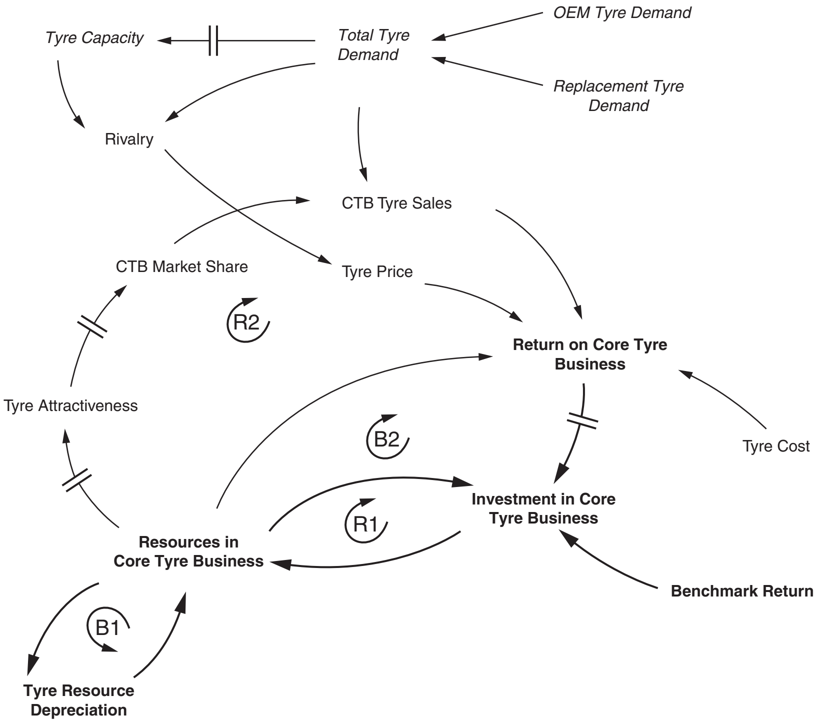


Fig. 1 Investment in the core by FocusCo and feedback consequences

resources falls (assuming that other influences on return such as sales, price and cost remain unchanged). On the other hand, as resources fall, return tends to rise. The balancing loop represents management's efforts to keep return in line with the agreed benchmark. There is also a long term reinforcing loop R2 linking resources to tyre attractiveness and to tyre sales. Greater investment in core resources can lead to a better product, increased market share and more sales. Finally, industry demand and rivalry play an important role in firm performance, as depicted at the top of the figure. Total tyre demand is the sum of replacement demand (car owners replacing worn-out tyres) and demand from original equipment manufacturers OEMs (car makers fitting new tyres on new cars). Rivalry depends on the balance of tyre demand and tyre capacity. When capacity exceeds demand, rivalry is high and manufacturers reduce price in an effort to utilise expensive idle capacity. Typically, in mature capital-intensive industries like tyres, demand is cyclical and capacity adjusts only gradually to changes in demand leading to volatile capacity utilisation and periods of intense rivalry.

Model of DiversiCo, the Diversifier

DiversiCo runs a tyre business like FocusCo, but also diversifies into new business areas when the financial performance of tyres becomes unsatisfactory, rather like Goodyear in the mid-1980s. Figure 2 shows

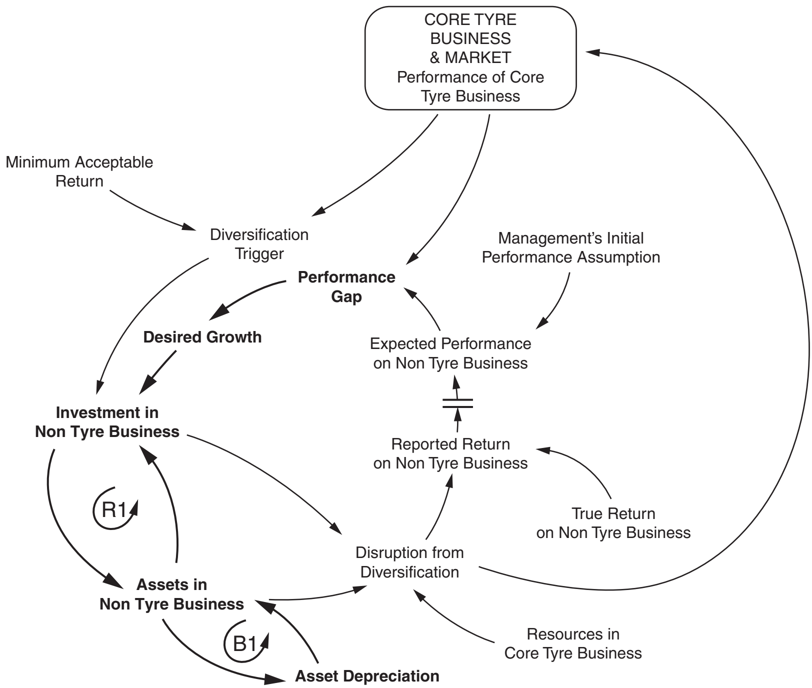


Fig. 2 Investment in the non-core by DiversiCo and feedback consequences

DiversiCo's investment policy in the non-core business. At the heart of the map is a reinforcing loop R1, shown in bold, connecting investment to assets in the non-tyre business. This loop will generate growth in assets providing that the performance of the non-tyre business is judged as superior to the core—a relative judgement. However, the dilemma facing managers is that they do not know for certain how well the new business will perform when it is integrated into the diversified portfolio. Instead they must make do with a judgement of expected performance that blends their initial performance assumption (as originally foreseen at the time of diversification) and reported performance. The blend will differ from company to company depending on the optimism, confidence and foresight of the management team. Moreover, the reported return on the non-tyre business can be distorted in the short to medium term by disruption caused by heavy new demands on management. The faster the rate of diversification, the greater the disruption. The same temporary disruption can also upset the performance of the core business, thereby further confusing the judgement of relative performance.

The overall feedback model of diversification as depicted in Figs. 1 and 2 embodies four specific assumptions about managerial behaviour that can influence diversification and resulting performance of the corporate portfolio:

1. Diversification takes place when the firm's existing business becomes *relatively* less profitable for new investment than new business—a Penrosian view.
2. Relative performance (core *vs* non-core) depends on a subtle managerial judgement that compares a perception of the performance of the core business with a perception of the possible future performance of the targeted non-core business(es).
3. Management's perception of the possible future performance of the non-core business depends on a blend of optimism, confidence and foresight.
4. The faster the rate of diversification, the lower the performance of both the core and non-core businesses due to disruption.

A Closer Look at FocusCo

Resources, Investment and Depreciation in the Core Business

The model of FocusCo takes an aggregate view of strategic investments. The flow of investment to the core business accumulates in a stock of strategic resources as shown in Fig. 3. This stock depletes gradually over time through depreciation. A stock-flow representation is of course fundamental to system dynamics [4, 7–10], but it is also compatible with resource-based thinking in contemporary strategy literature [11–15]. According to the resource-based view, strategic resources are any resources of the firm, tangible or intangible, that underpin sustainable competitive advantage and that are difficult for competitors to copy or buy. In a tyre company like FocusCo they can include not only physical assets such as capital equipment, inventory and staff but also intangible assets such as product quality and brand image. Similarly, investment flows can include not only capital expenditures on machinery, equipment and buildings, but also product development programmes, advertising expenditures, quality initiatives and staff training. The initial stock of strategic resources,

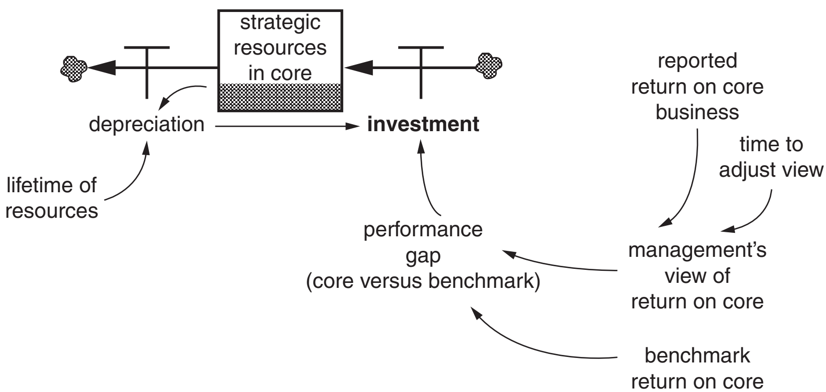


Fig. 3 FocusCo's strategic resources and investment policy for the core

both tangible and intangible, is valued at ten billion dollars, scaled to represent a company like Goodyear. Resources depreciate with an average lifetime of three years that aggregates the long life of physical assets with the fleeting life of intangible assets such as marketing image and tyre tread design.

Figure 3 shows the factors that drive investment in the core business. Further details on these factors can be found in Appendix 1, including equations. At the heart of the investment policy is a comparison of management's view of return relative to the so-called benchmark return. The benchmark is the prevailing view among the firm's leaders about acceptable rates of return for the industry. The benchmark return is initialised at 5.8% per year (calibrated to tyre industry case data in the late 1970s) and evolves as a five-year long-term average of reported return. The performance gap between the management view and the benchmark determines the rate of investment in the core. If management believe future return on the core will be higher than the benchmark then they approve investment at a rate that is greater than resource depreciation. On the other hand, if future return is expected to be lower than the benchmark, then investment projects will be pared back to a rate lower than depreciation, leading to a fall in accumulated resources. The conversion of the performance gap into investment is represented in the model by a non-linear (graphical) formulation which determines the funding required for the core business as a multiplier of resource depreciation.

Return on the Core Business

A vital role for management is to form a balanced view of the likely future return on the business. In Fig. 3, management's view of return on the core depends on the reported return (in the company accounts). If return is falling, say due to a market downturn, then it is management's job to gauge whether the fall is temporary (just a blip in the market), enduring (a business cycle downturn) or permanent (a structural change in the industry and market). This judgement is very important because it is the basis for all subsequent investment approval by the management team of the focusing firm. The time span for such a judgement is represented by a concept called time to adjust view, which is set at two years in the equations (see Appendix 2). The longer the time span, then the more stable is management's perception of the underlying return potential of the business. Return itself comes from the gap between revenue and normal manufacturing cost relative to the value of accumulated resources invested in the core business.

A Closer Look at DiversiCo

The Trigger for Diversification

Diversification begins (in a Penrosian way) when managers in the core business judge that the existing market and industry is no longer attractive. The first step is simply to form a consensus view that diversification is an option worth pursuing. In the model the option is exercised through a diversification trigger (see Appendix 3 for details). The diversification trigger depends on a comparison of management's view of return relative to a minimum acceptable return. The trigger can be either '0' (no diversification, stick to knitting) or '1' (seriously consider the option of diversification and win political support). A switch of the trigger from 0 to 1 represents the realisation among opinion leaders in the firm that core performance is unsatisfactory (below the minimum). Broad, company-wide, acceptance of diversification follows the opinion leaders with a long time lag of 20 years that represents the time to change opinions across the entire organisation. Of course the firm does not wait a full 20 years to act on the views of opinion leaders. A modest degree of acceptance, less than 50%, will be sufficient to provide the political support for change.

Management's Perception of Non-core Performance

Once there is an acceptance that diversification is a worthwhile option then the search begins in earnest for new lines of business and acquisition targets. A fundamental criterion driving this search is management's shared view of the expected return on potential non-core business. Here is a fundamental policy and information dilemma. How can management know future performance when (by definition) they have little or no experience of the business into which they are diversifying? The truth is they do not know for sure. They can seek expert advice. They can scrutinise the track record of an acquisition target. They can collect information on rivals. But in the end, expected return is a collective judgement subject to the biases, whims, beliefs and pet theories of the team who have the power to act.

In Fig. 4, expected return begins at a value equal to management's initial performance assumption, which can be set anywhere in the range 6–12% per year, depending on the optimism and foresight of the management team. Expected return then adjusts toward the reported return, which itself adapts to the true return on the non-core business that only time and experience can reveal. Knowledge of the true return is further confounded by disruption to performance caused by the process of diver-

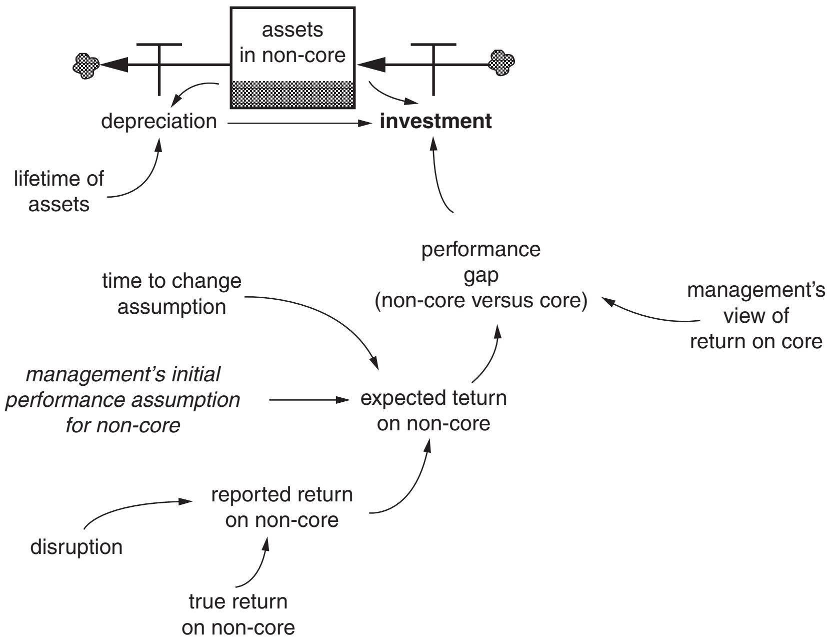


Fig. 4 DiversiCo's assets and investment policy for the non-core

sification itself. The disruption effect is described later. In steady state, the true return can be as high as 8% per year or as low as 6% per year (which is only half the most optimistic initial assumption).

The rate at which expected return adjusts to the true return depends on a crucial managerial trait represented by the 'time to change assumption'. Some management teams have strongly held opinions while others are easily swayed by the course of events. The longer the time to change assumption, the more management hold tenaciously to their initial performance assumption in the face of conflicting evidence from experience (see Appendix 4 for more details). Such confidence has both virtues and dangers in diversification. Unbridled optimism coupled with confidence could lead to a ruinous campaign of inappropriate diversification. On the other hand, stark realism coupled with a limited sense of purpose could prematurely starve a promising new business venture. Simulations of the model will explore the possibilities of varying degrees of confidence in more depth.

Investment Policy and Assets in the Non-core Business

Investment in the non-core business depends on the performance gap between the return generated by the non-core and the core businesses. The bigger the gap the more attractive is the non-core, relative to the core, and so the more funds are allocated to non-core investment. Figure 4 develops this Penrosian view of diversification. Incidentally, the implicit assumption in the model is that adequate funding is available for diversification and there is no competition with the core business for limited funds. Such free availability of funding is quite plausible if the core business is a cash cow or if the firm has a good reputation with banks, financiers and the stock market.

The performance gap drives desired growth of the non-core. The bigger the performance gap the greater is desired growth, though the precise formulation is non-linear (see Appendix 5 for details). When the performance gap is zero then desired growth is zero, even though management has accepted the need for diversification. A significant performance gap of 4% per year elicits a desire to expand the non-core rapidly at an assumed maximum rate of 25% per year. The model is not explicit about whether diversification happens through acquisition or organic growth. However, the desired growth formulation, with its assumed maximum growth rate and smooth process of asset accumulation is probably more appropriate to organic growth (starting from a small acquired base) than to pure acquisition.

Investment in the non-core business is the sum of authorised funds for growth and asset depreciation. The assumption here is that baseline investment in the non-core is sufficient to replace depreciation. This baseline is then augmented by authorised funds for growth. Investment accumulates in a resource stock which is drained by depreciation. Depreciation is proportional to the asset stock and inversely proportional to the lifetime of non-core assets. The longer the lifetime, the lower the rate of depreciation. Lifetime is assumed to be ten years, much longer than in the core business on the assumption that new lines of business are in growth industries with less aggressive imitation.

Disruption from Diversification and Its Effects on Performance

Diversification is inherently a disruptive process. It consumes scarce management time and attention. It introduces new people, new values, new ways of thinking, new markets, new customers and new suppliers to an organisation that has evolved to serve its core market. Such changes are likely to undermine the performance of both the core and non-core business, at least while diversification is in progress and the balance of core and non-core activities is changing.

The model represents disruption as a function of the growth rate of the non-core business and the relative size of non-core assets within the corporate portfolio (see Appendix 6 for details). Growth rate is defined as the ratio of net investment in the non-core to assets in the non-core and measures the pace of diversification. If net investment is zero (in other words, the rate of investment equals asset depreciation), then disruption is zero because diversification has reached an equilibrium. On the other hand, if net investment is positive then the non-core business is still growing and there is potential for disruption. Relative size of the non-core is the ratio of resources in the core to the sum of all resources and assets in the combined business. The assumption is that as the non-core business becomes an ever-greater proportion of the total business, further diversification is less disruptive than at an earlier stage of the diversification programme. At the outset, all resources are in the core business, so the ratio takes its maximum value of 1. But, if for example the non-core grows to be 50% the size of the core (in asset terms) then the disruptive effect of further growth of the non-core is diluted in the ratio 1–1.5, or in other words by two-thirds.

Disruption has an impact on the performance of both the core and non-core businesses. In the non-core it reduces reported return below the underlying true return of the non-core. The formulation is a simple multiplicative function in which reported return is the product of true return and the complement of disruption (1—disruption). So if disruption is zero, because growth of the non-core has ceased, then reported return becomes the true return. On the other hand, if growth of the non-core is proceeding at its maximum rate of 25% per year, and the diversification programme has only just started (so the core is still dominant in the asset portfolio) then reported performance is reduced to 75% of true performance.

Disruption has a different effect on the core. It is assumed to lower the market attractiveness of products in the core business. This effect stems from diversion of management time and attention from the core which results in less effective strategic investment of resources. In practical terms the effectiveness of core activities such as marketing programmes, product development initiatives and pricing campaigns is lowered because management no longer has its eye firmly on the core. To illustrate, take the same example as above where growth of the non-core is proceeding at 25% per year. Assume also that strategic investment in the core has been the same as the industry average. Under these circumstances core products (tyres) become only 75% as attractive to customers as they would otherwise have been in the absence of diversification. The depressing effect on attractiveness is large, though it is not instantaneous. Disruption gradually plays its way through strategic investment programmes. The average time for implementation of such programmes is assumed to be four years. So a policy of rapid diversification, though highly disruptive to the core, may not make itself felt in product or service quality for several years. Even then, the customer will take time to perceive the quality change and switch product or supplier. Conversely, once the damage is done to product quality and image, it takes a long time to reverse.

Simulation Experiments on Diversification

A series of simulation experiments compare the performance of FocusCo and DiversiCo under identical scenarios for demand in the core tyre industry. The results are reported below.[2]

Experiment 1: FocusCo Faces a Sharp Downturn in Tyre Demand

In Experiment 1 FocusCo faces a 20% downturn in tyre demand, starting in year 2 of the simulation and extending to year 20. The benchmark return on the core business is initially 5.8% per annum and by definition FocusCo has no option to diversify. Figure 5 shows the industry scenario and its effect on both FocusCo's revenue and profits.

Tyre demand (line 1 in the top chart) falls in year 2 and then settles at a new stable value for the remainder of the simulation out to year 20. Industry capacity (line 2) gradually declines to equal demand. But there is a long period between years 2 and 10 of the simulation when there is excess capacity in the industry, reflecting the long lifetime of specialised factory machinery and normal exit barriers in a mature capital-intensive industry. As a result of the capacity surplus, industry rivalry (line 3) rises sharply in year 2, hitting a peak of almost 1.5%, before falling back to a neutral value of 1 as the industry's capacity overhang is slowly worked off.

Rivalry causes price to decline and lowers FocusCo's profit margin. Gross profits appear as the gap between core business revenue (line 1 in the lower chart) and normal manufacturing cost (line 2). It is easy to see that profits decline sharply during year 3 of the simulation as a combination of falling demand and falling price eat into revenue. Meanwhile, manufacturing cost falls due to lower volume, but not enough to

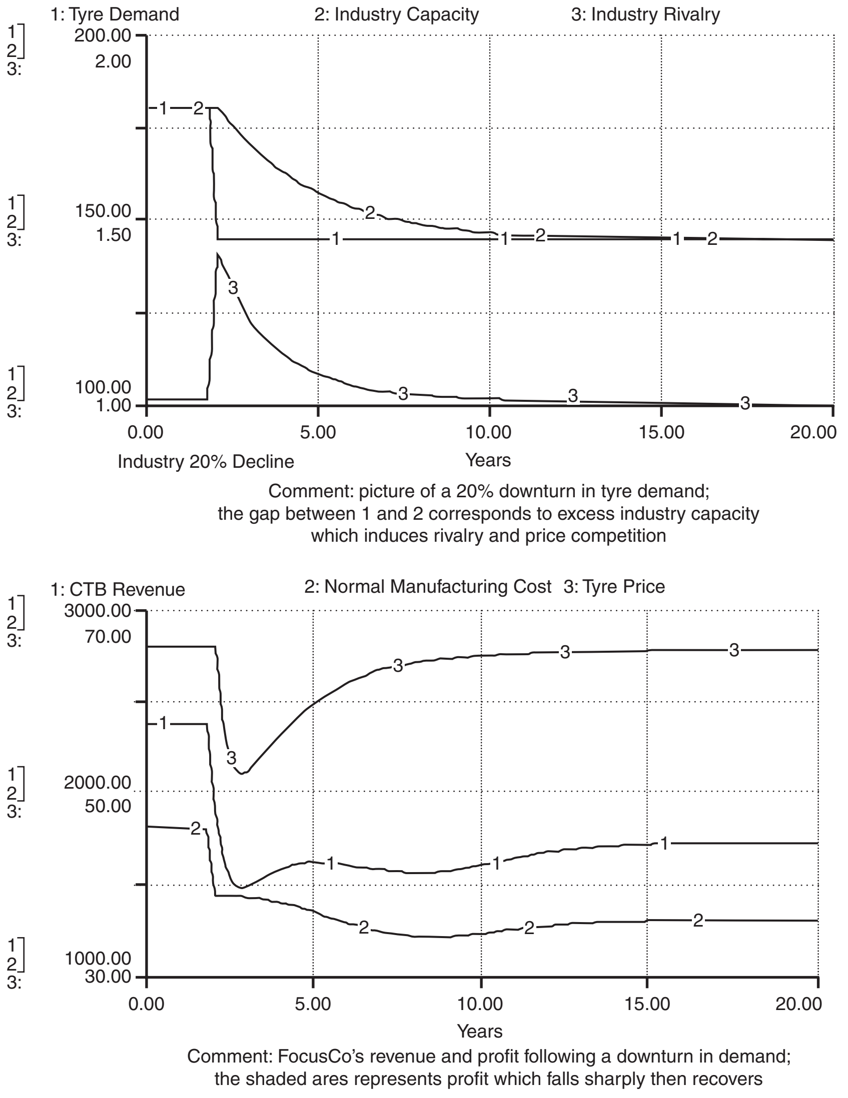


Fig. 5 A thought experiment—a 20% permanent downturn in tyre demand and the financial consequences for FocusCo

compensate for revenue loss. From year 3 onwards profitability recovers as FocusCo reduces costs and as price recovers in the wake of declining rivalry. The vulnerability of profits to a downturn is typical of a mature competitive industry. This was the kind of volatile environment facing executives in Goodyear following the recession caused by oil price hikes of the 1970s.

Figure 6 shows how FocusCo responds to the downturn. Return on the core tyre business, CTB, (line 2 in the upper chart) falls steeply from a starting value of 5.8% and then recovers gradually in the interval between years 3 and 7 as management implement cost-cutting programmes. Meanwhile, market share (line 1 in the upper chart) falls gradually from a starting value of 20% to a low point of about 17% in year 8 of the simulation, before recovering to a new stable value of about 18.5%. This long-term erosion of market share is the result of falling product attractiveness caused by underinvestment. The implicit competitive assumption here is that FocusCo produces a premium-differentiated tyre achieved through heavy investment in product development and marketing. Cuts in such programmes can restore profits, but only at the expense of product differentiation and long-term market share.

Overall resources in the core business decline quite steeply from their initial value of ten billion dollars shortly after the industry downturn (line 1 in the lower chart). The exact resource trajectory can be understood by comparing investment in the core (line 2) with resource depreciation (line 3). Aggregate resources accumulate the difference between investment and depreciation. So when depreciation exceeds investment, resources fall, as in years 2–7. When investment exceeds depreciation then resources rise, as in years 7–15. In equilibrium, resources remain at a constant value and there is an exact balance of investment and depreciation.

Experiment 2: DiversiCo Faces a Downturn in Tyre Industry Demand

In experiment 2 DiversiCo faces the same 20% downturn in tyre demand as FocusCo. But management choose to diversify (by acquisition then organic growth) when return on the core falls below a threshold of 4%

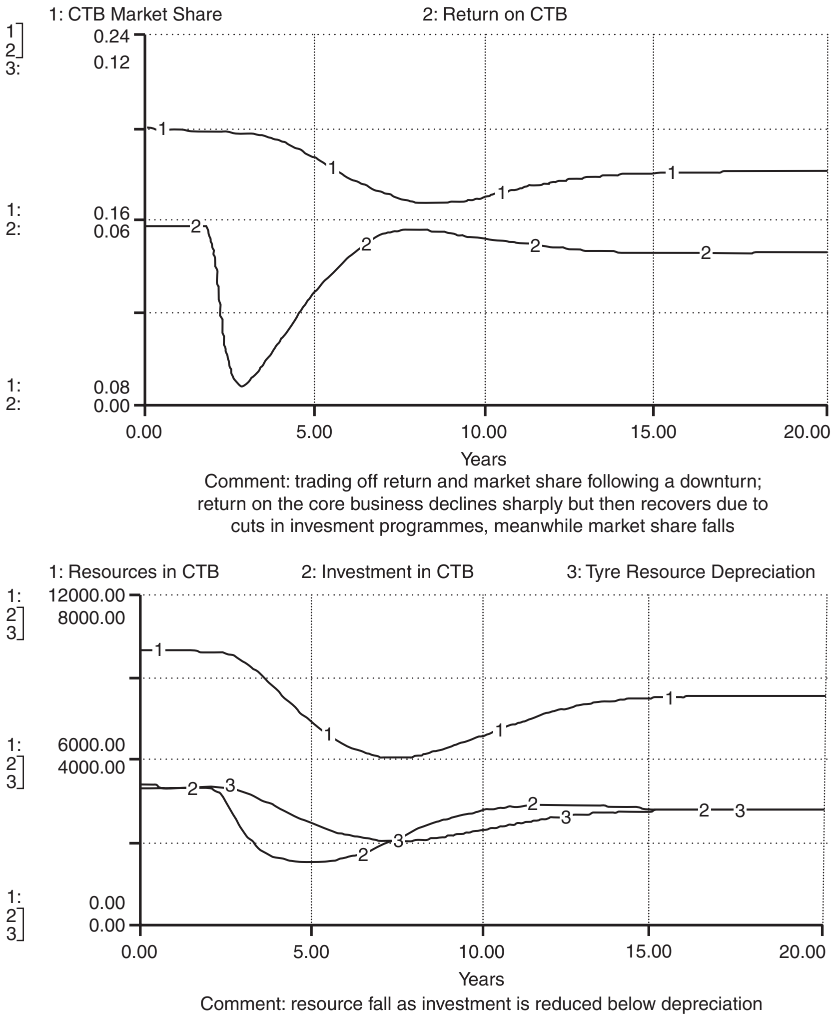


Fig. 6 FocusCo's response to a sustained downturn; reduced investment and downsizing to boost return

per year. The benchmark return on the core business is initially 5.8% per year. The expected return on the non-core business is set at 8% per year and is assumed to be an accurate estimate of the true underlying return which is also 8%.

The top half of Fig. 7 shows the performance gap between DiversiCo's core and non-core businesses during a downturn in the core. In equilibrium the gap is (by assumption) 2.2% per year, which is the difference between the benchmark return on DiversiCo's core business and the true underlying return on the non-core. After the business downturn at the start of year 2, the performance gap begins to rise and reaches a peak value of almost 5% per annum in year 4 due to a dramatic fall in return on the core business. Core return falls below the minimum acceptable threshold of 4% per annum, triggering diversification. As the core business adapts to the downturn, then the performance gap falls from its peak value of 5% back to 2% by year 10. Thereafter the gap rises slowly to about 2.5% by the end of the simulation.

The bottom half of Fig. 7 shows DiversiCo's assets in the non-tyre business (line 2) which grow steadily at a rate of about 25% per year, starting from an initial value of 0.5 billion dollars (line 1). Such growth, when compounded, leads to accumulated non-tyre assets of 5 billion dollars by year 15, and almost 16 billion dollars by year 20. DiversiCo's portfolio undergoes a radical transformation. (Note: the initial non-core assets represent a small legacy from past diversification ventures. For example, Goodyear in the 1970s focused on its core tyre business but also ran a small aerospace business derived from its famous symbol, the Goodyear blimp).

Comparing Performance—DiversiCo vs FocusCo

Figure 8 compares strategic performance of DiversiCo and FocusCo in the core tyre business. In the top half of the figure, DiversiCo (line 2) loses much more market share than FocusCo (line 1), about 4% more by the end of the simulation, despite facing an identical industry downturn. The reason for DiversiCo's relative share loss is a mixture of underinvestment in the core and disruption from diversification. The lower half of

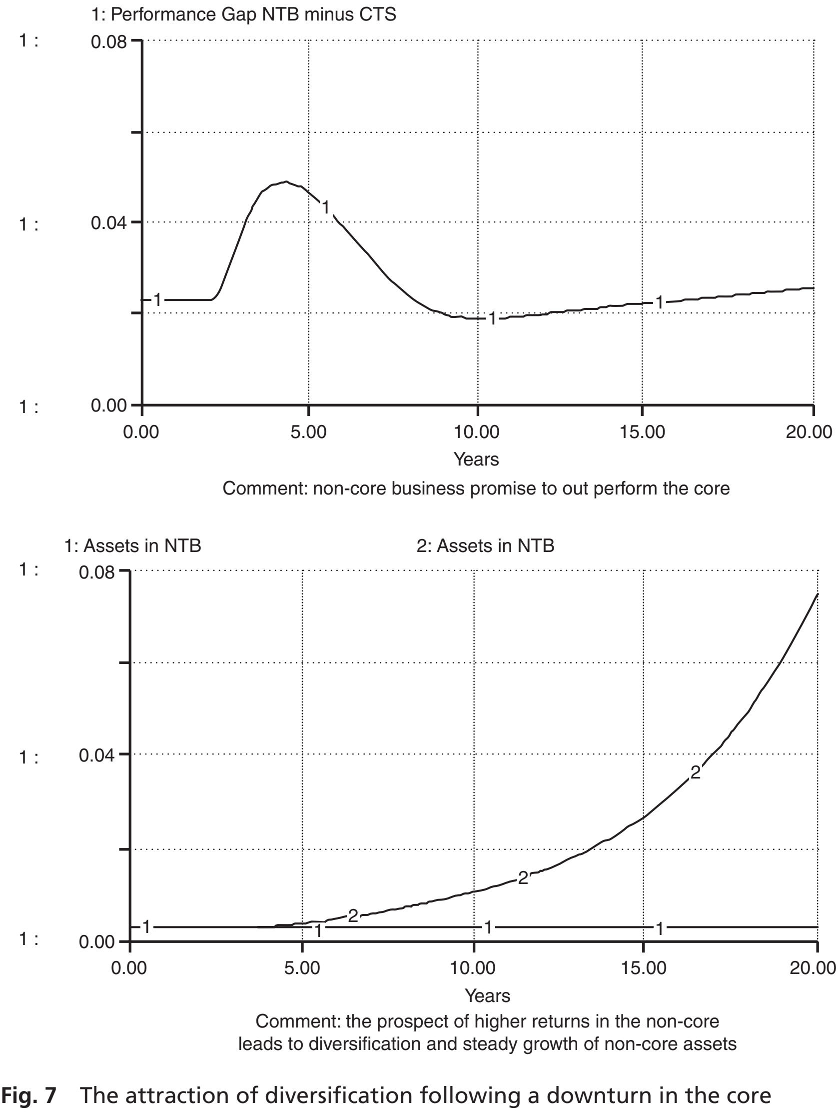


Fig. 7 The attraction of diversification following a downturn in the core

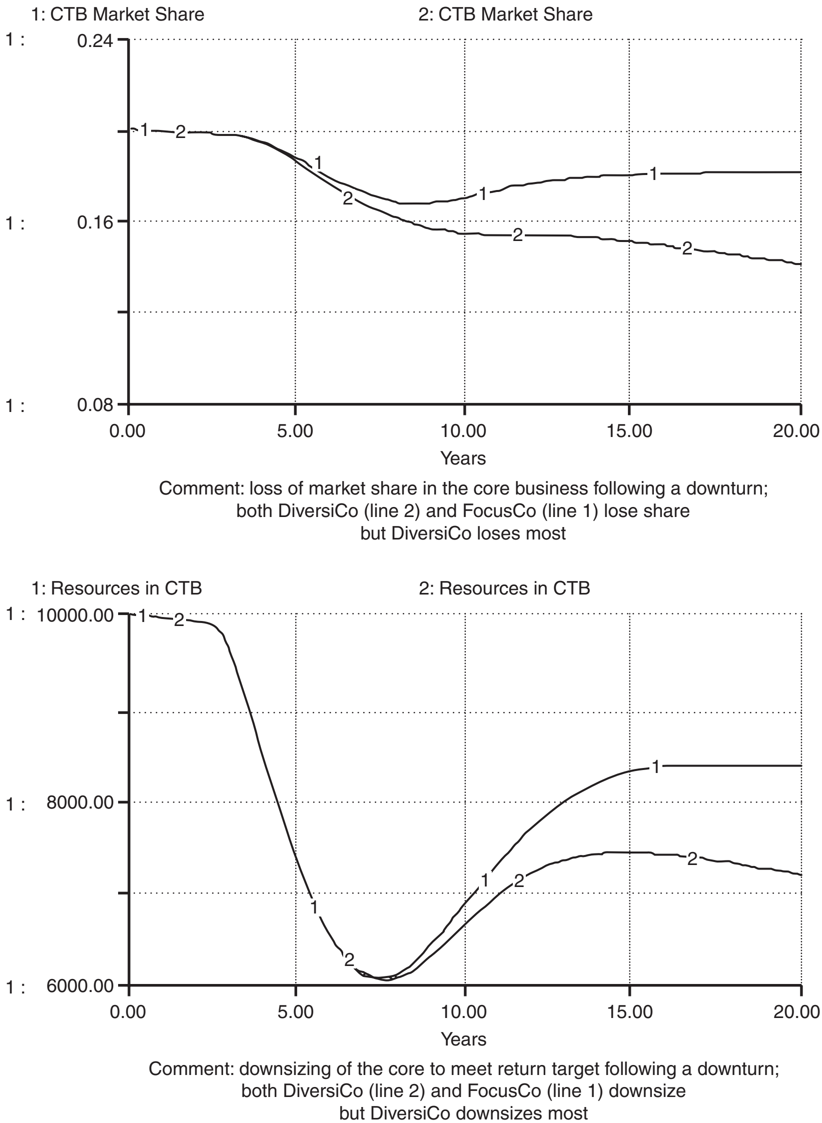


Fig. 8 Comparing strategic performance of DiversiCo and FocusCo in the core

Fig. 8 shows that both DiversiCo (line 2) and FocusCo (line 1) reduce resources in the core tyre business in years 2 through 7 following the downturn. But thereafter, DiversiCo invests less in its core business than FocusCo. Here we see the operation of a self-fulfilling spiral of decline in DiversiCo. A corporate investment policy dominated by financial return starves the core business of strategic investments because diversification is undermining the effectiveness of those investments. The result is that the core looks even less attractive relative to the non-core, thereby sustaining a robust weight of opinion in favour of continued diversification. DiversiCo's core steadily loses resources and market share relative to a more focused rival.

What about the overall financial performance of DiversiCo and FocusCo? Figure 9 tells the story. Interestingly, for the first ten years of the simulation the combined return on DiversiCo's core and non-core businesses (line 2) is almost identical to FocusCo's return (line 1). This close similarity of performance occurs even though DiversiCo's non-core business has, by assumption, the potential to outperform tyres. The large initial size of the core coupled with disruption from growth of the

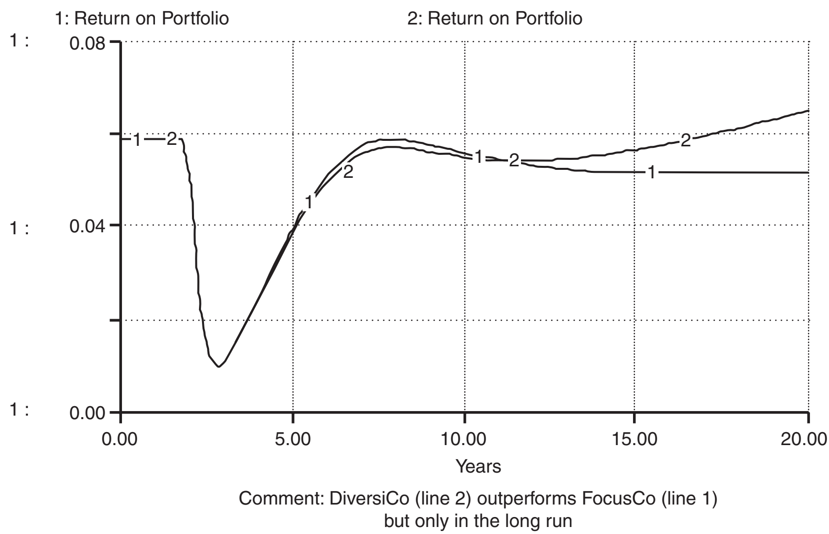


Fig. 9 Comparing financial performance of DiversiCo and FocusCo following a downturn in the core

non-core conspire to mask performance advantages from diversification, or at least defer the advantages for many years. Only in the case where the non-core has a massive and sustainable return advantage over the core (say 5% or more) are appreciable relative performance advantages likely to be visible in the early years following diversification. Of course the validity of this result hinges on the assumption that disruption is a real phenomenon and that relative returns (core *vs* non-core) matter to the investment policy of the diversifier.

Experiment 3: DiversiCo Faces a Downturn in Tyre Demand and Is Lured into 'Excess' Diversification by Over-optimism

Once more DiversiCo faces a 20% downturn in demand for tyres, with the option to diversify. In this case however, management's initial performance assumption for the non-core is much too optimistic at 12% per annum rather than the 'true' underlying return of 6%. The true return of 6% is deliberately chosen to be little better than the benchmark expected from the core in the long run (5.8%). If these facts were known beforehand by the management team there would be no financial incentive for diversification. But only time and experience reveal the true facts.

Figure 10 shows over-optimism and imperfect foresight leading to excessive growth of the non-core. The top half of the figure compares a falsely optimistic perception of the performance gap (core minus non-core, line 2) with an accurate and unbiased perception (line 1). There is a huge distance of 6% between the two lines at the start. When diversification is triggered in year 3, the distance slowly erodes as management adjust their perception in the face of unrelenting evidence that the non-core does not match their expectations and hopes. But the aura of optimism encourages investment in the non-core as shown in the bottom half of the figure. Assets in the non-core grow to almost 8 billion dollars (8000 million, line 2) by the end of the simulation, fuelled by optimism. Meanwhile the realist diversifier hardly invests at all (line 1) even though there is a climate of opinion in favour of diversification.

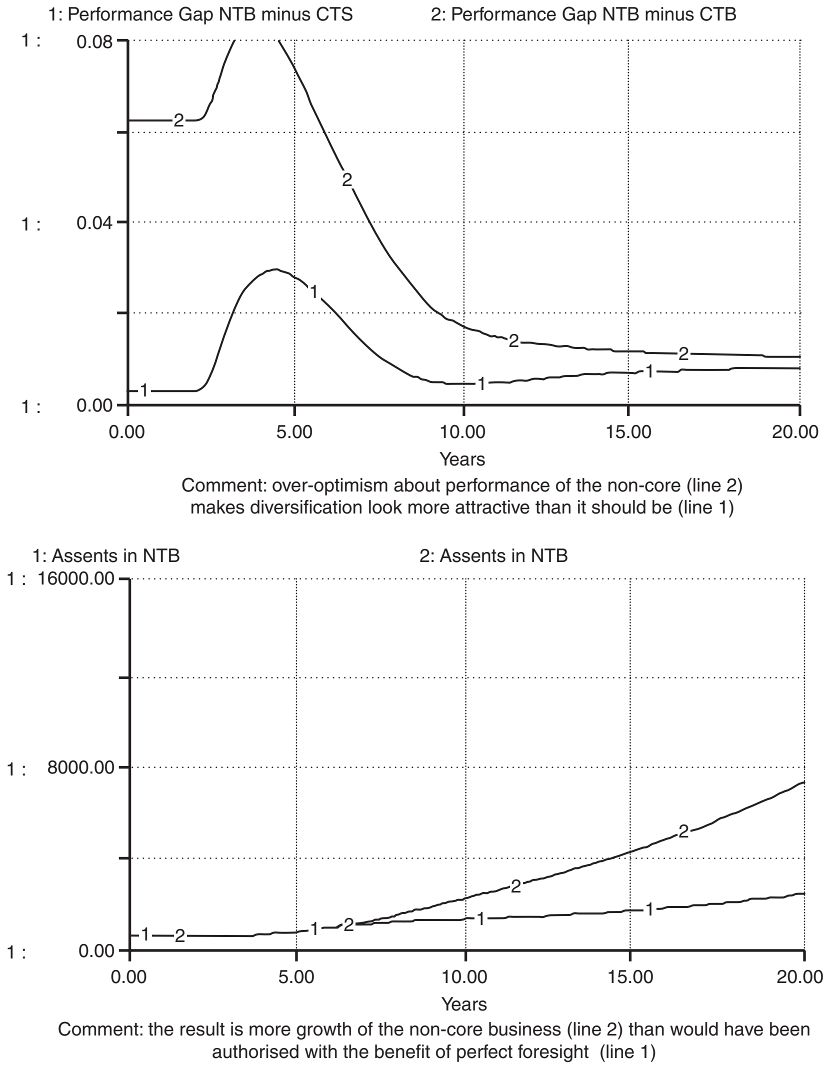


Fig. 10 'Excess' diversification fuelled by over-optimism

Over-optimism in the non-core has knock-on consequences to the core (which are reported here but not shown as simulations). Resources are squeezed in an effort to maintain adequate returns and market share in the core falls by about two percentage points more for the mistaken optimist than for the realist. The unintended victim of overly optimistic diversification policy is the core business itself.

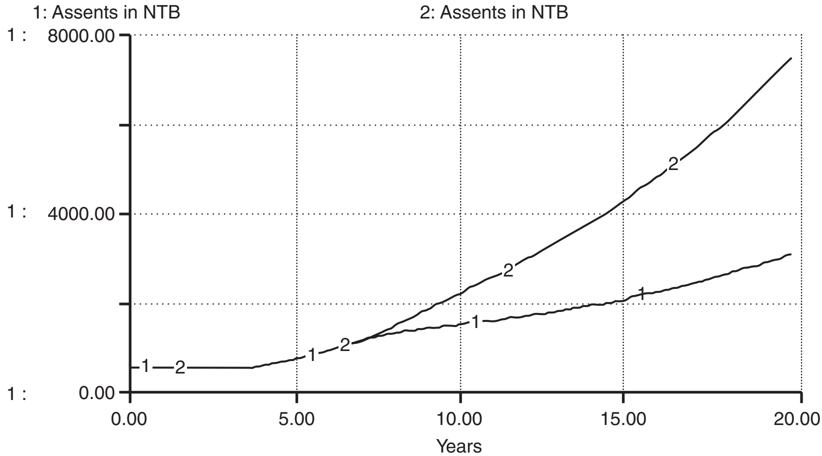
Meanwhile, overall return on DiversiCo's portfolio is virtually the same as return to FocusCo (the chart is not shown here, but is similar to line 1 in Fig. 9). At first glance this similarity of financial outcome may seem surprising. After all, DiversiCo has embarked on a seemingly fruitless investment in a non-core business that experience proves is no better than the core in the long run. On the other hand, it is no worse in the long run and is (by assumption) better during downturns. So in the end DiversiCo just ends up as a mediocre performer like FocusCo, with a different composition of assets and resources. In this case, managerial over-optimism changes the character of the business, but not its financial performance.

Experiment 4: Non-core Growth and the Trade-Off Between Confidence and Learning

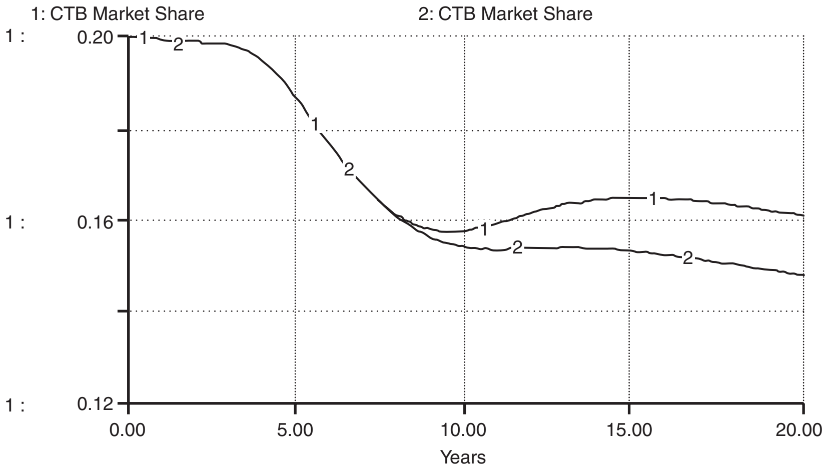
Diversification is an inherently risky policy. The firm ventures into an unfamiliar area of business, serving new customers in new markets, facing new rivals and engaging in new activities. Goodyear's diversification into the gas pipeline business in 1983 illustrates the difficulties: completely new customers such as electric utilities in place of traditional buyers such as car manufacturers and tyre dealers; unfamiliar and powerful rivals such as Exxon and Conoco in place of old faces like Uniroyal and Bridgestone; new activities such as offshore exploration and production in place of well understood activities such as tyre tread design and brand advertising. Management cannot know for certain the relative potential of the new business until they have it in harness, functioning smoothly as part of a diversified portfolio. In the face of so many unknowns a vital role for management is to transmit and maintain confidence in the new enterprise as a basis for justifying investment in the non-core. On the other hand management needs to form a realistic view of the returns that can be expected from the new business and to update this view in the light of experience, in other words, to learn from experience. Confidence and learning can often be at odds with each other and present business leaders with a fundamental dilemma in managing diversification.

Experiment 4 examines the trade-off between confidence and learning as it affects investment and growth of the non-core. Once again the core business faces a permanent 20% downturn in demand, generating poor trading conditions in the core that fuel opinion in favour of diversification. The management team is naturally optimistic about opportunities in the non-core which they initially assume will generate annual returns of 12%. As before the true, but hidden, performance potential is 6% per annum. The experiment compares the implications for growth and performance of variations in management's time to change their assumption about expected return on the non-core. Run 1 of the comparative simulation corresponds to fast learning: members of the management team update their expectations of return over a time horizon of two years. Run 2 portrays confidence: team members are not swayed by short-term results and update their expectations of return over a time horizon of five years.

Confidence boosts growth of the non-core but undermines the core through disruption. Fast learning inhibits growth of the non-core with less disruption to the core. Figure 11 shows the details. Diversification is triggered during year 3. For a period of almost four years, up to year 7, the growth trajectory of the non-core is identical in the two runs (top half of figure). But after year 7 the trajectories diverge significantly. With fast learning (line 1) management soon acknowledge that the initial performance assumption of 12% was much too optimistic. Lower expected return reduces the relative attractiveness of the non-core thereby curtailing the rate of investment. With high confidence (line 2) management maintain a belief that the non-core is capable of high returns—despite contrary evidence from reported returns. To some extent this belief is self-fulfilling. Confidence fuels continued investment in the non-core and rapid growth. Disruption from growth depresses market share in the core (bottom half of Fig. 11) which, through its depressing effect on core profitability, tends to confirm the view that the non-core is much more attractive than the core. Interestingly the overall portfolio return (not shown as a chart) is virtually identical under fast learning or high confidence.



Comment: confidence in diversification boosts growth of the non-core (line 2) above the rate that would be achieved if management learned quickly about true performance to the non-core (line 1)



Comment: fast growth of the non-core from managerial confidence depresses core market share (line 2) below the value that would result if management learned quickly about true performance of the non-core (line 1)

Fig. 11 Managerial confidence and learning and their effect on growth of the non-core and market share of the core

Slightly superior true returns available from the non-core (which reward confidence), are offset by the disruption to core performance caused by rapid growth of the non-core.

Conclusions

Successful diversification is an elusive phenomenon and the conditions that breed success, though much studied, are still not well understood. A common-sense hypothesis is that corporations can reap improved returns by diversifying into related businesses. Relatedness offers opportunities for the corporate centre to exploit synergies and economies of scope not available to undiversified rivals. Numerous statistical studies have set out to test this hypothesis, but with inconclusive results. Relatedness does not seem to matter. This outcome is disappointing for both researchers and policymakers and has led to much debate about the definition and accurate measurement of relatedness. Some believe that a better metric will reveal the expected relationship. But relatedness is only part of the story.

This paper takes a different approach by developing a dynamic and behavioural simulation model of a diversifying firm. The model represents the corporate centre of a diversifying firm as an investor, choosing whether to favour investment in the core or non-core business, based on expected relative financial performance. This simple disaggregation to core and non-core business captures a common (though much overlooked) fact about many diversifying firms: usually the core business is very big and the non-core is small, at least to start. So performance of the portfolio depends on relative scale of its parts which is dynamic and evolving. In the early years, the core dominates portfolio performance, even if the non-core is much better (or worse). Note that at this level of aggregation the role of the corporate centre as match-maker and synergy exploiter is invisible.

Simulations of this (essentially Penrosian) model reveal difficulties for corporate policymakers in evaluating the performance of a diversifying firm. We discover for example that the corporate centre is unlikely to find immediate and measurable financial rewards from diversification (by comparison with a focuser facing identical trading conditions) even when the non-core business is assumed to have superior financial performance to the core. This financial insensitivity stems from relative scale of the businesses (mentioned above) and disruption caused by the process of diversification itself. The greater the potential (steady-state) gap between non-core and core return, the faster diversification proceeds and therefore the greater the disruption from growth. The simulations suggest that portfolio performance results from a dynamic process dependent on the pace of diversification and the initial stock of assets and resources (clearly a much different explanation than performance that results from exploiting relatedness among shared resources, though in reality both happen simultaneously).

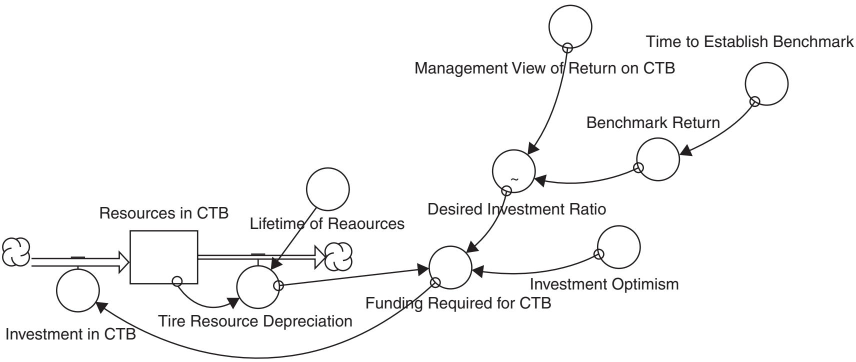
We also discover that behavioural traits of management's investment policy (optimism, limited foresight, confidence, and malleability of prior beliefs) all play an important role in determining the pace of diversification and therefore the dynamics of performance. Overoptimism about future return from the non-core can lead to excessive growth of non-core assets, while fast learning about true performance potential can prematurely stunt growth. Corporate mindsets matter to a diversifying firm—a cognitive view consistent with Prahalad and Bettis [16] in their influential article on dominant logic, diversity and corporate performance.

The results suggest a number of practical issues for closer attention of executives and researchers. Do not expect instant benefits from diversification, even if synergy arguments look good. Relative scale of the non-core is vital and it may take many years to significantly change the composition of the portfolio. Meanwhile, performance is dominated by the core and its dynamic industry forces. In the search for synergies do not overlook dynamic diseconomies such as disruption and experience dilution from growth. Such transient diseconomies can mask or even reverse long-term economies of scope for years. Also, remember that if dynamic diseconomies disrupt the core then diversification can become a self-fulfilling prophecy as the performance of the core is undermined through fast diversification. Finally, in a world of dynamically changing portfolio performance, managerial traits are part of the feedback structure linking investment to performance. In this welter of dynamic and behavioural factors it is not surprising that a clear link between diversification and performance is elusive.

Acknowledgements The author gratefully acknowledges helpful and perceptive criticism from referees on an earlier draft of the paper. Thanks also to Ari Ginsberg whose understanding of corporate strategy and curiosity about system dynamics led to the Goodyear modelling project and many interesting discussions about the role of feedback systems thinking in strategy and case teaching. Finally, thanks to Costas Markides for his advice on interpreting the findings of the diversification literature.

Appendix 1

Map and equations for investment policy and resources in the core business

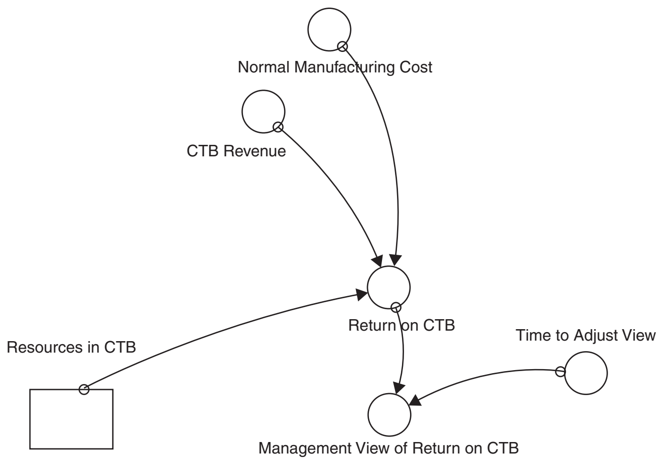


Resources_in_CTB(t) = Resources_in_CTB(t - dt) + (Investment_in_CTB - Tire_Resource_Depreciation) * dt
INIT Resources_in_CTB = 10000 {millions dollars)
Investment_in_CTB = Funding_Required_for_CTB {millions dollars per year) Tire_Resource_Depreciation = Resources_in_CTB/Lifetime_of_Resources {millions dollars per year}
Benchmark_Return = SMTH1(Return_on_CTB, Time_to_Establish_Benchmark, .058) {fraction per year}
Funding_Required_for_CTB = Tire_Resource_Depreciation * Desired_Investment_Ratio * Investment_Optimism {millions dollars per year}
Investment_Optimism = 1

Lifetime_of_Resources = 3 {years}
Management_View_of_Return_on_CTB = SMTH1(Return_on_CTB, Time_to_Adjust_View) {fraction per year}
Time_to_Establish_Benchmark = 5 {years}
Desired_Investment_Ratio = GRAPH(Management_View_of_Return_on_CTB - Benchmark_Return {dimensionless})
(−0.02, 0.478), (−0.016, 0.526), (−0.012, 0.61), (−0.008, 0.706), (−0.004, 0.838), (0.00, 1.00), (0.004, 1.20), (0.008, 1.37), (0.012, 1.47), (0.016, 1.53), (0.02, 1.56)
Comment: in this map the term core tyre business is abbreviated to CTB: the same abbreviation is used in later maps

Appendix 2

Map and equations for management view of return on core business

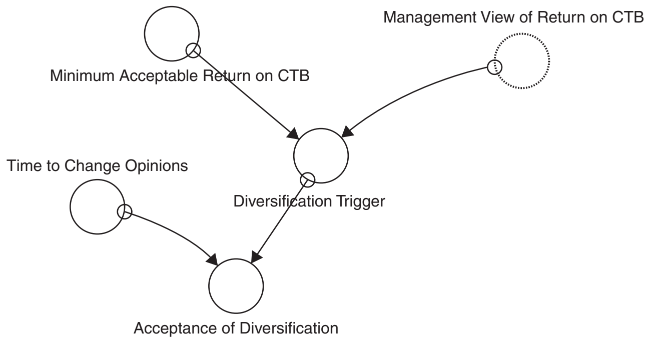


Resources_in_CTB(t) = Resources_in_CTB(t - dt) + (Investment_in_CTB - Tire_Resource_Depreciation) * dt
INIT Resources_in_CTB = 10000 {millions dollars}
CTB_Revenue = Tire_Price * CTB_Tire_Sales {millions dollars per year}

Management_View_of_Return_on_CTB = SMTH1(Return_on_CTB, Time_to_Adjust_View)
{fraction per year}
Normal_Manufacturing_Cost = ((CTB_Tire_Sales * Normal_Tire_Cost {millions dollars per year}
Return_on_CTB = ((CTB_Revenue-Normal_Manufacturing_Cost)/ Resources_in_CTB) {fraction per year}
Time_to_Adjust_View = 2 {years}

Appendix 3

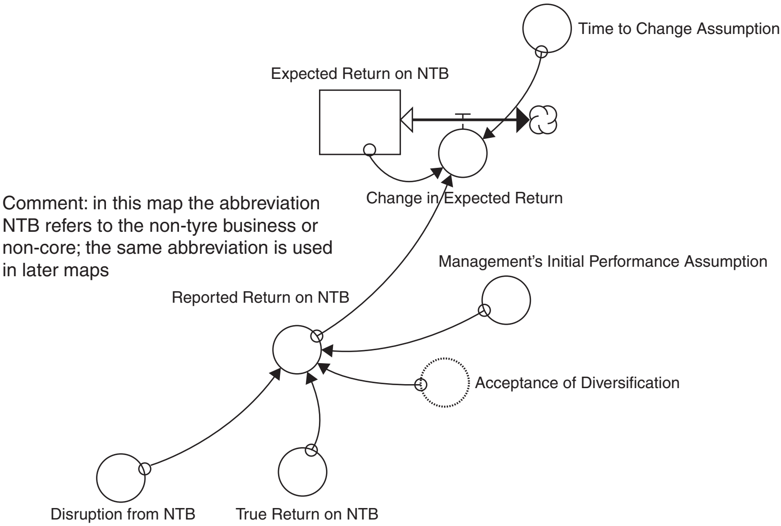
Map and equations for diversification trigger



Management_View_of_Return_on_CTB = SMTH1(Return_on_CTB, Time_to_Adjust_View)
{fraction per year}
Acceptance_of_Diversification = SMTH1 (Diversification_Trigger, Time_to_Change_Opinions)
{dimensionless}
Diversification_Trigger = if Management_View_of_Return_on_CTB < Minimum_Acceptable_Return_on_CTB then (1) else (0) {dimensionless}
Minimum_Acceptable_Return_on_CTB = .04 {fraction per year}
Time_to_Change_Opinions = 20 {years}

Appendix 4

Map and equations for management's perception of non-core performance



Expected_Return_on_NTB(t) = Expected_Return_on_NTB(t - dt) + (Change_in_Expected_Return) * dt

INIT Expected_Return_on_NTB = Management's_Initial_Performance_Assumption

Change_in_Expected_Return = (Reported_Return_on_NTB-Expected_Return_on_NTB)/Time_to_Change_Assumption {fraction per year per year}

Acceptance_of_Diversification = SMTH1 (Diversification_Trigger, Time_to_Change_Opinions)

{dimensionless}

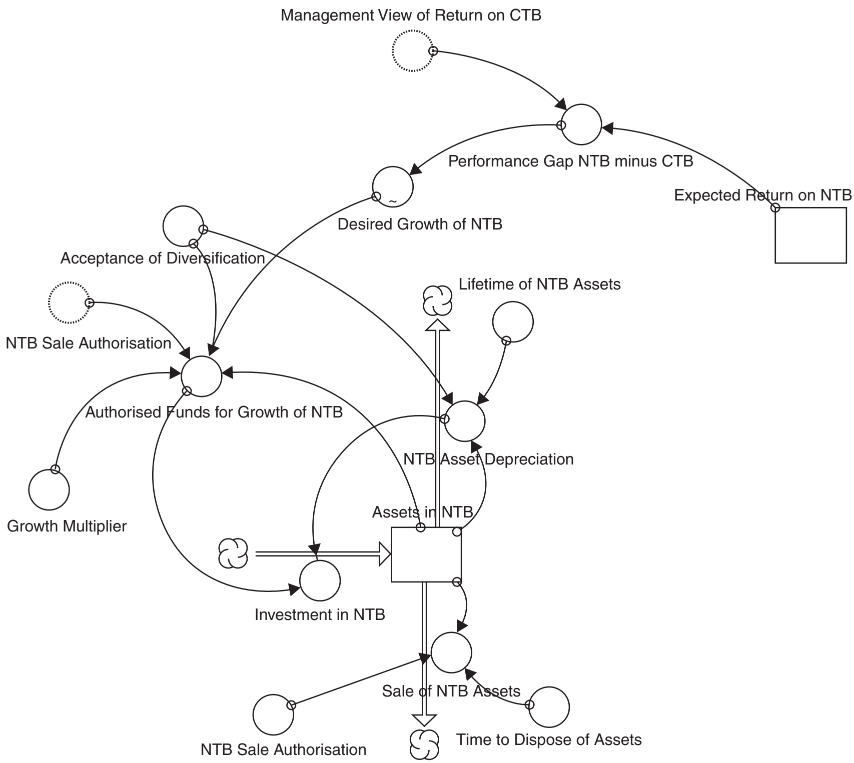
Disruption_from_NTB = ((Investment_in_NTB-NTB_Asset_Depreciation)/Assets_in_NTB)

*(Resources_in_CTB/(Resources_in_CTB+Assets_in_NTB)) {fraction per year}

Management's_Initial_Performance_Assumption = .12 {fraction per year)
Reported__Return_on_NTB = if Acceptance_of_Diversification > 0 then True_Return_on_NTB * (1-Disruption_from_NTB) else Management's Initial_Performance_Assumption {fraction per year}
Time_to_Change_Assumption = 5 {years}
True_Return_on_NTB = .06 {fraction per year}

Appendix 5: Part 1

Map of investment policy and assets in the non-core business



Appendix 5: Part 2

Equations for investment policy and assets in the non-core business

Management__View_of_Return_on_CTB = SMTH1(Return_on_CTB, Time_to_Adjust_View)
{fraction per year}
Assets_in_NTB(t) = AssetsJn_NTB(t - dt) + (InvestmentJn_NTB - Sale_of_NTB_Assets - NTB_Asset_Depreciation) * dt
INIT Assets_in_NTB = 500 (millions dollars}

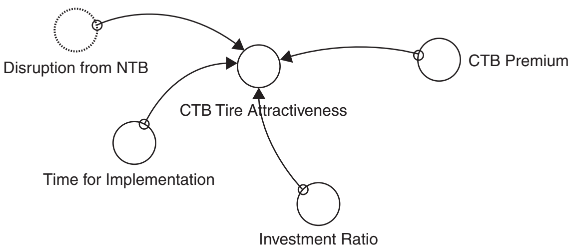
Investment_in_NTB = Authorised_Funds_for_Growth_of_NTB + NTB_Asset_Depreciation {millions dollars per year}
Sale_of_NTB_Assets = Assets_in_NTB*NTB_Sale_Authorisation/Time_to_Dispose_of_Assets
{millions dollars/year}
NTB__Asset_Depreciation = if (Acceptance_of_Diversification>0) then (Assets_in_NTB/Lifetime_of_NTB_Assets) else (0)
{millions dollars per year}
Expected_Return_on_NTB(t) = Expected_Return_on _NTB(t - dt) + (Change_in_Expected_Return) * dt
INIT Expected_Return_on_NTB = Management's_Initial_Performance_Assumption

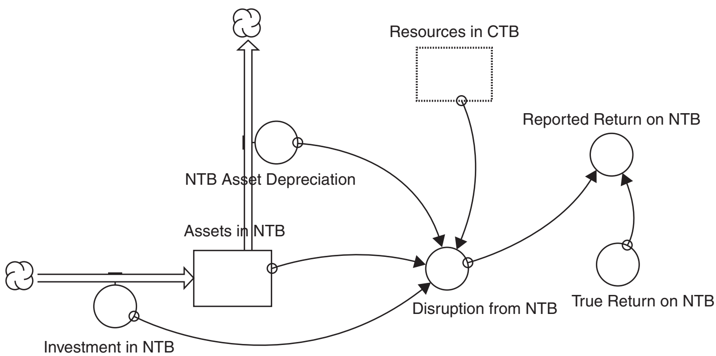
Acceptance_of_Diversification = SMTH1 (Diversification_Trigger, Time_to_Change_Opinions)
{dimensionless}
Authorised_Funds_for_Growth_of_NTB = if (Acceptance_of_Diversification>0 and NTB_Sale_Authorisation = 0) then (Desired_Growth_of_NTB * Assets_in_NTB * Growth_Multiplier) else 0 {millions dollars per year}
Growth_Multiplier = 1 {dimensionless}
Lifetime_of_NTB_Assets = 10 {years}
NTB_Sale_Authorisation = 0 {dimensionless}
Performance_Gap_NTB_minus_CTB = Expected_Return_on_NTB -
Management_View_of_Return_on_CTB {dimensionless}
Time_to_Dispose_of_Assets = 2 {years}

Desired_Growth_of_NTB = GRAPH(Performance_Gap_NTB_minus_CTB {fraction per year}) (−0.02, −0.0725), (−0.015, −0.0625), (−0.01, −0.0475), (−0.005, −0.025), (1.73e-18, 0.00), (0.005, 0.0475), (0.01, 0.102), (0.015, 0.15), (0.02, 0.192), (0.025, 0.217), (0.03, 0.237), (0.035, 0.247), (0.04, 0.25)

Appendix 6: Part 1

Disruption from diversification and effect on performance





Appendix 6: Part 2

Equations for disruption from diversification and effect on performance

Resources_in_CTB(t) = Resources_in_CTB(t - dt) + (Investment_in_CTB - Tire_Resource_Depreciation) * dt

INIT Resources_in_CTB = 10000 {millions dollars)
CTB_Premium = .1 {dimensionless fraction)
CTB_Tire_Attractiveness = SMTH1(Investment_Ratio * (1-Disruption_from_NTB),
Time_for_Implementation)/(1+CTB_Premium) {dimensionless}
Investment_Ratio = Investment_in_CTB/Industry_Standard_investment
Time_for^Implementation = 4 {years}
Assets_in_NTB(t) = Assets_in_NTB(t - dt) + (Investment_in_NTB - Sale_of_NTB_Assets -
NTB_Asset_Depreciation) * dt
INIT Assets_in_NTB = 500 {millions dollars)
Investment_in_NTB = Authorised_Funds_for_Growth_of_NTB + NTB_Asset_Depreciation
{millions dollars per year)
NTB_Asset_Depreciation = if (Acceptance_of_Diversification>0) then
(Assets_in_NTB/Lifetime_of_NTB_Assets) else (0)
{millions dollars per year)
Disruption_from_NT B = ((Investment_in_NTB-NTB)_Asset_Depreciation)/Assets_in_NTB)
*(Resources_in_CTB/(Resources_in_CTB+Assets_in_NTB)) {fraction per year}
Reported_Return_on_NTB = if Acceptance_of_Diversification> 0 then True_Return_on__NTB * (1-Disruption_from_NTB) else Management's_Initial_Performance_Assumption {fraction per year)
True_Return_on_NTB = .06 {fraction per year}

Notes

1. Although system dynamics has long been applied to strategic business problems, there has been surprisingly little published work dealing with the topic of diversification and multi-business firms. Widely cited business dynamics models typically deal with a single business or else a complete industry.
2. A notable exception is the work of Peter Merten [17] and Merten et al. [18] who portrayed diversifying firms as a combination of causal feedback loops (to represent business unit operating policies) and 'intelligent logical

loops' (to represent corporate strategy). In this approach the intelligent logical loops switch on new, latent pieces of feedback structure depending on the performance of the business portfolio relative to corporate goals.

FocusCo and DiversiCo are hypothetical firms. However, initial values for variables such as demand, capacity, market share, price and profitability are scaled to data from case studies by Ginsberg [5] describing the US tyre industry and Goodyear. Readers should be aware that the models are not tightly calibrated to real time series data. Nor are they intended to replicate the turbulent history of Goodyear or the tyre industry in the 1980s. By design, the models deviate from the case in several important respects: (1) Goodyear faces highly cyclical demand in both the OE and replacement markets, whereas the models use a simple one-time step downturn in demand. The downturn is a thought experiment to isolate the dynamics of firm performance. (2) The Goodyear case deals with diversification through repeated acquisition whereas DiversiCo best fits diversification through organic growth starting from an initial acquisition. (3) In reality, the repercussions of Goodyear's diversification strategy extend to the stock market and a powerful corporate raider in the person of the late sir James Goldsmith. In contrast, DiversiCo's simulated strategy escapes the attention of the stock market. (Nevertheless, it is feasible to represent additional stakeholders. See Ginsberg and Morecroft [6] for a review of feedback loops that might link core and non-core businesses, such as DiversiCo's, to a stock market and corporate raider).

References

1. Markides CC and Williamson PJ (1994). Related diversification, core competencies and corporate performance, *Strat Mgmt J* **15**: 149–165.
2. Robins J and Wiersema MF (1995). A resource based approach to the multi-business firm: empirical analysis of portfolio interrelationships and corporate financial performance, *Strat Mgmt J* **16**: 277–299.
3. Penrose E (1959). *The Theory of the Growth of the Firm.* London: Basil Blackwell.
4. Sterman JD (1989). Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment, *Mgmt Sci* **35**: 321–339.
5. Ginsberg A (1995). Transformation of the US tyre industry and Goodyear on the skids, case studies, Stern School of Business: New York University.

6. Ginsberg A and Morecroft JDW (1997). Weaving feedback systems thinking into the case method: an application to corporate strategy, *Mgmt Learning* **28**: 455–473.
7. Forrester JW (1961). *Industrial Dynamics*. Portland, OR: Productivity Press (originally published by MIT Press, Cambridge MA).
8. Forrester JW (1992). Policies, decisions and information sources for modeling, *Eur J Opl Res* **59**: 42–63.
9. Morecroft JDW and Hines JH (1986). Strategy and the representation of structure. *Working Paper WP-1721-85*. Revised April 1986, MIT Sloan School of Management, Cambridge MA.
10. Warren KD and Langley PA (1999). The effective communication of system dynamics to improve insight and learning in management education, *J Opl Res Soc* **50**: 396–404.
11. Wernerfelt B (1984). A resource-based view of the firm, *Strat Mgmt J* **5**: 171–181.
12. Dierickx I and Cool K (1989). Asset stock accumulation and sustainability of competitive advantage, *Mgmt Sci* **35**: 15041511.
13. Barney J (1991). Firm resources and sustained competitive advantage, *J Mgmt* **17**: 99–120.
14. Warren KD (1996). The dynamics of building resources for competitive advantage, *Financial Times, 'Mastering Management'*, issue 17, March 1996.
15. Warren KD (2000). *Competitive Strategy Dynamics*. Chichester: Wiley (forthcoming).
16. Prahalad CK and Bettis RA (1986). The dominant logic: a new linkage between diversity and performance, *Strat Mgmt J* **7**: 485–501.
17. Merten PP (1991). Loop-based strategic decision support systems, *Strat Mgmt J* **12**: 371–386.
18. Merten PP, Loffler R and Wiedman KP (1987). Portfolio simulation: a tool to support strategic management, *System Dynamics Review* **3**: 81–101.

Relevance Assumed: A Case Study of Balanced Scorecard Development Using System Dynamics

H.A. Akkermans and K.E. van Oorschot

Trusting his images, he assumes their relevance;
Mistrusting my images, I question their relevance.
Assuming their relevance, he assumes the fact;
Questioning their relevance, I question the fact.
When the fact fails him, he questions his senses;
When the fact fails me, I approve my senses.
Robert Graves, from *In Broken Images*

Journal of the Operational Research Society (2005) **56**(8), 931–941.
https://doi.org/10.1057/palgrave.jors.2601923
Published online 22 December 2004.

H.A. Akkermans (✉)
Department of Management, Tilburg School of Economics and Management, Tilburg University, Tilburg, The Netherlands

Eindhoven University of Technology, Eindhoven, The Netherlands

K.E. van Oorschot
Minase BV, Tilburg, The Netherlands

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_4

Introduction

Some 2 decades ago, Harvard Professors Johnson and Kaplan renounced conventional financial measures as the right way to control company performance in their book *Relevance lost: The Rise and Fall of Management Accounting* [1]. Instead, they introduced an integrated set of financial and non-financial measures, which has become known under the label of the balanced scorecard [2–4]. Since then, performance management in general, and the balanced scorecard (BSC) approach in particular, have risen to prominence in both the business world and in academia. Balanced scorecard concepts are now also studied in several areas of management research, such as organization studies [5], operations management [6, 7] and information systems [8].

The reasons for this sudden rise to prevalence appear obvious in retrospect. On the one hand, there is the appeal of simplicity: no longer do managers have to work their way through piles of statistics, but can focus on monitoring some 15–20 key indicators instead. On the other hand, there is the strength of interdisciplinarity. In the past, all the relevant inputs from different functional areas had to be translated into financial data. Some functions, such as marketing, may have been significantly better suited to do this than others, such as R&D or operations. But, regardless of how successful this translation was, it remained a conversation in a 'foreign language' [9] for non-financial managers. With a BSC, managers have now a more acceptable common language to discuss issues in.

Despite or perhaps in response to its popularity, the BSC concept has had its share of criticism. Broadly speaking, these criticisms can be seen as the other side of the coin for the before-mentioned advantages of the BSC. Yes, the idea of only a few process indicators that point at key leverage points of the system is very attractive [10, 11]. But, how can one be sure that the few ones selected are indeed the right ones? Should there be more? Or less? And, do they all work in the same direction or might they counteract each other? Moreover, if they are the right variables, what are the correct values to target for? And, within what time frame should these be achieved? From a theoretical perspective, these are not trivial questions. Nor are they, from a practitioner's perspective, trivial to managers implementing a BSC, as we will observe further on. Rather than *assume* their scorecard was relevant, the managers in the case study we will be describing were eager to go through a rigorous process of *questioning* its relevance, as in the poem by Robert Graves this article derives its title from.

A comparable weakness is inherent to the apparent advantage of interdisciplinarity of BSCs. If a BSC is to reflect all the different relevant perspectives on the business, then all the stakeholders that represent these perspectives should be actively involved in its development. How else can one be sure that all the relevant viewpoints are represented in the BSC? But then, how does one organize a process in which a group of people with inherently different perspectives, goals and constraints and, indeed, languages, can find agreement upon just a few numbers as the basis for a joint strategy?

In this paper, we suggest that system dynamics (SD) can be an effective way of overcoming these limitations to BSC development. After a brief recapitulation in the next two sections of what BSC and SD entail, we describe the development of a BSC with system dynamics for one of the business units of Interpolis, a leading insurance company in the Netherlands. We introduce the case setting next and then describe how a 'strategy map' [12, 13] of the BSC was developed in close collaboration with the management team. We then outline the SD simulation model that was developed from this map and look at a number of policy experiments that were conducted with this model, which led to further refinements in the final BSC. In the concluding sections we consider limitations of this research and pros and cons of different OR-based approaches for BSC development.

Current BSC Theory and Practice

The Business scorecard is a performance measurement system introduced by Kaplan and Norton [2, 3]. According to these authors, a BSC addresses shortcomings of traditional performance measurement systems that relied solely on financial measures. To overcome this, Kaplan and Norton introduced three additional measurement categories that cover non-financial aspects. The result is a scorecard that translates the vision and strategy of a business unit into objectives and measures in four different areas:

1. *The financial perspective*: how the company wishes to be viewed by its shareholders;
2. *The customer perspective*: how the company wishes to be viewed by its customers;

3. *The internal business process perspective*: in which processes the company must excel in order to satisfy its shareholders and customers;
4. *The organizational learning and growth perspective*: which changes and improvements the company must achieve to implement its vision.

The 'balance' of the scorecard is reflected in its mix of lagging (outcome measures) and leading (performance drivers) indicators, and of financial and non-financial measures.

Recently, Kaplan and Norton have developed the notion of a strategy map as a complementary concept next to the BSC. A strategy map links measures of process performance, or key performance indicators (KPIs), together in a causal chain that leads through all four perspectives: measures of organizational learning and growth influence measures of internal business processes, which, in turn, act upon measures of the customer perspective, which ultimately drive financial measures [12, 13]. Causal chains or causal diagrams provide a medium by which people can externalize mental models and assumptions and enrich these by sharing them [14–17]. In fact, one of the hidden strengths of a balanced measurement framework, in particular of the BSC, may well be that it forces management teams to explore the beliefs and assumptions underpinning their strategy [18].

The BSC concept originates from the USA. There it has been applied successfully across many industries and within the public sector. It has also been delivered to an international audience and on a multidisciplinary front [19]. For example, Malmi [20] found that the logic of the BSC was appealing to many companies in Finland. Wisniewski and Dickson [21] describe its application to a police force in Scotland. From a functional perspective, researchers from different management fields have made contributions. The management accountancy aspect of the BSC has been considered by, for example, Newing [22] and Nørreklit [23]. Also in the operations management field the BSC is well-known [6, 7, 24, 25]. From a general strategy perspective, the BSC has been described by for example Mooraj et al. [26] Hudson et al. [27].

Partly, the success of the scorecard can be explained by good timing and marketing. Managers were clearly frustrated with traditional measurement systems at the time when the BSC was promoted in the *Relevance Lost* book by Johnson and Kaplan and in articles in the Harvard Business Review. However, apart from timing, one key strength of the BSC is that its appearance is so agreeably simple. It suggests that with only a few well-balanced numbers one can monitor the performance of an entire company. Another key strength of the BSC concept is that it can serve as a bridge between different fields, both financial and non-financial ones.

Next to the well-published successes of BSC, a number of inherent weaknesses have been reported in the literature. Interestingly, the advantages of the BSC mentioned in the previous paragraph can also be interpreted as disadvantages. If all one has is a small set of indicators, how can one be sure that these are the right ones; can one be assured of their relevance? And, if the BSC development process offers opportunities for bridging different fields, how does one organize this effectively? Regarding the first weakness, it has been stated repeatedly that the BSC concept provides no mechanism for maintaining the relevance of defined measures [23]. Neely et al. [6] found that the problem for managers is usually not identifying what could be measured, but reducing the list of possible measures to a manageable (and relevant) set. An additional concern here is that the concept of causality is not in all implementations of BSCs equally well developed. In their more recent work on strategy maps, Kaplan and Norton emphasize the importance of showing how improvements in one area lead to improvements in others [12, 13]. Nørreklit [23] questions the existence of such causality in most BSCs currently being used. Moreover, instead of unidirectional causal relationships, he believes that the relationship is much rather one of interdependence, of bi-directional causality. Tan et al. [28] agree and stress that simply looking at different measures simultaneously is not enough. The linkages between them must also be understood.

Regarding the second advantage, that of bridging different fields, this is in practice easier said than done. For instance, what to do when performance indicators of different fields counteract or thwart each other? The bridging to be performed can be both hierarchical and horizontal. Regarding the former, Mooraj et al. [26] state that the BSC concept fails to identify performance measurement as a two-way process, as it focuses primarily on top-down performance measurement. Hudson et al. [27] agree to this and write that BSCs have a lack of integration between the top-level, strategic scorecard, and operational-level measures. Regarding the latter, Mooraj et al. [26] discuss that BSCs do not consider the extended value chain in which employee and supplier contributions are also highlighted. Neely et al. [6] argue that the BSC is not able to answer one of the most fundamental questions for managers: what are the competitors doing? So, to summarize, the advantages of checking just a few numbers related to different fields may become a disadvantage when there are many relevant fields to be looked at.

System Dynamics as a Modelling Method for BSC Development

In this article, we suggest the use of system dynamics (SD) as an approach to overcome the limitations to current BSC theory and usage mentioned in the literature. In particular, we suggest a two-stage process of SD modelling for BSC development. We call these two stages the qualitative and the quantitative stage in SD modelling. This distinction is a very common one in SD, see for example, Wolstenholme's overview in the special issue on SD in JORS in 1999 [29].

- *Stage 1*: Elicit qualitative mental models from management of perceived interrelationships using causal loop diagramming (CLDs), resulting in a refined version of a 'strategy map' [12, 13]. From this map, distil key performance indicators (KPIs) and assign preliminary targets.
- *Stage 2*: Translate the causal loop diagram into a quantified simulation model. Calibrate the model using key company data. Test the implicit assumptions about dynamic behaviour in the preliminary BSC on the basis of this simulation model with managers and discuss implications for mental models and BSC.

Not only is the distinction between qualitative and quantitative modelling a common one in the SD literature, but a two-stage approach as suggested here to system dynamics modelling is also the 'normal' way of conducting system dynamics interventions in organizations [14–16, 29–33]. In this article, we show how this generic approach can be tailored to support the process of developing BSCs. We will illustrate how this approach yields advantages compared to more 'conventional approaches', advantages that stem from the systematic approach to 'strategy mapping' that SD offers, as well as advantages that originate from the smooth transition from qualitative models to quantified simulation.

Obviously, SD is not the only method available for improving the conventional way in which most scorecards have been developed so far. There are several 'soft OR' methods that offer similar improvements in rigour in the mapping process [33, 34]. In particular, cognitive mapping (SODA) [34] and soft systems methodology [35] come to mind. Both approaches place great emphasis on identifying causal relations between key variables and constructing 'maps' of these. Both these qualitative approaches can be followed up by a more focused quantitative modelling stage. In particular of SODA there have been reports in the literature of modelling efforts where this approach was combined with SD modelling and simulation [36–38]. Also, both approaches typically employ a process of close collaboration with key problem owners in which maps and models are developed interactively to assure maximal model relevance and client ownership [34, 35]. The other way round, qualitative SD followed by a quantitative stage on the basis of another OR method, also occurs. One published example describes the use of multicriteria analysis, preceded by a first stage of causal loop diagramming [39]. Of course, if SD simulation is deemed most appropriate for the Stage 2 process, a strength of the SD approach becomes that the transition between the causal maps that are developed first and the simulation model to be developed next can be a very smooth and natural one. Problem owners see the same diagrams in the same software package that they used first but now get a 'back-end' to it, equations and graphs which build upon their intuitive grasp of the model after the first stage.

There is one school of thought in operations and strategic management that sets system dynamics apart from all other OR methods for BSC development. Authors such as Warren [40, 41] stress the importance of distinguishing explicitly time delays and accumulations in BSCs. Regarding time delays, some indicators are leading, others are lagging. Changes can be made to inputs and changes can be observed in outputs [28]. In 'conventional' BSCs this distinction in time delays is not made as explicitly and rigorously as in SD models. Regarding accumulations, in the strategic management literature on the resource-based view of the firm, the concept of accumulation as a driving force for firm performance has been stressed repeatedly [42, 43]. Such resources, both tangible and intangible, grow or decline gradually over time, and it is these accumulations that really drive organizational performance according to this literature. Therefore, one should make it clear what those accumulations really are and what drives their behaviour over time. The stocks-and-flows concept in SD addresses both time delays and accumulations in a rigorous manner, while most other BSC and OR methods do so considerably less.

The Case Study

The case study described in this article concerned a business unit of Interpolis. Interpolis (www.interpolis.nl) is one of the leading insurers in the Netherlands. Since 1990, the company as a whole forms part of the Rabobank organization, one of the three large banking and insurance conglomerates in the Netherlands. Interpolis employs some 6000 people. 2003 revenues were 5 Billion.

The business unit concerned is called Stichting Rechtsbijstand ('Foundation for Legal Aid', abbreviated from here on as 'SRB'). Some 600,000 consumers have insured their needs for judicial aid with this formally independent organization. Some 300 SRB employees provide this aid. Overall, the demand for legal aid continues to increase considerably. In 2003, 68,000 requests for legal aid were noted, which is up 16% from 2002.

In the years prior to our involvement with this company, SRB had gone through a period of considerable upheaval. First, there had been several changes in top management. Secondly, there was the noted and continued significant increase in demand for its services, as a result of changes in the market and new insurance sales policies with its parent, Rabobank. In response, staff hiring had increased significantly, after a long period of little to no growth. Thirdly, the organization had recently undergone a major restructuring, shifting from a regional structure to a structure according to area of judicial expertise. Fourthly and finally, the management team members were almost all less than a year in their current jobs, including the CEO. All in all, the time seemed right for a serious reorientation on key goals for the future.

This was the background against which our involvement with this company should be situated. We, the authors, formed part of a small group of external consultants who facilitated the management team (MT) of this organization in the development of a BSC. This development process was set up in two stages.

During the first stage, preparatory interviews were conducted with MT members, the results of which were discussed in a half-day workshop where the group engaged in a number of causal loop diagramming exercises. The findings from this workshop were summarized in a so-called work-book [15], which was distributed to the MT members, studied, filled in and sent back by for analysis, leading to input for the next workshop. A simplified and stylized version of the end result of this CLD mapping exercise is shown in Fig. 1 and described in the next section. The intermediate causal loop diagrams were discussed in a final full-day workshop. On the basis of these discussions, an intermediate version of a BSC was generated in a fairly straightforward manner. MT members chose those key indicators from the diagram that they felt would enable management to keep a good grasp of key drivers of performance. The resulting list of measures and objectives was discussed, refined, simplified and finally agreed upon. This resulted in the intermediate version of the BSC that is shown in Table 1.

Perhaps the most important managerial insight that emerged from this stage was that management came to realize how goals that they had first believed to be at odds with each other were, in fact, not mutually exclusive but mutually reinforcing. It was not: choose *either* for customer satisfaction, *or* for employee satisfaction *or* for cost effectiveness, but rather: either we will achieve all three goals *or* none at all. The key linking concept was employee productivity. Higher productivity does mean greater cost effectiveness, but also greater customer satisfaction, as cases are handled sooner, and greater employee satisfaction, as work pressure is less severe.

At the end of stage 1 there was agreement on the content of the BSC. Equally important, there was also agreement on the approach forward. Especially relevant in the context of the current article is that the team felt pleased with their first BSC, but at the same time was very uncertain about its quality. Were these really the right indicators? Had they been complete? And would these indeed all work towards the same goal? To what extent would the chosen indicators really be sufficient to achieve the overall company mission? On the other hand, could this list of KPI's not be shortened further? After all, the fewer dials to watch the easier it becomes to monitor performance effectively. To address these uncertainties, the team decided that a system dynamics simulation model was to be developed to investigate these questions more thoroughly.

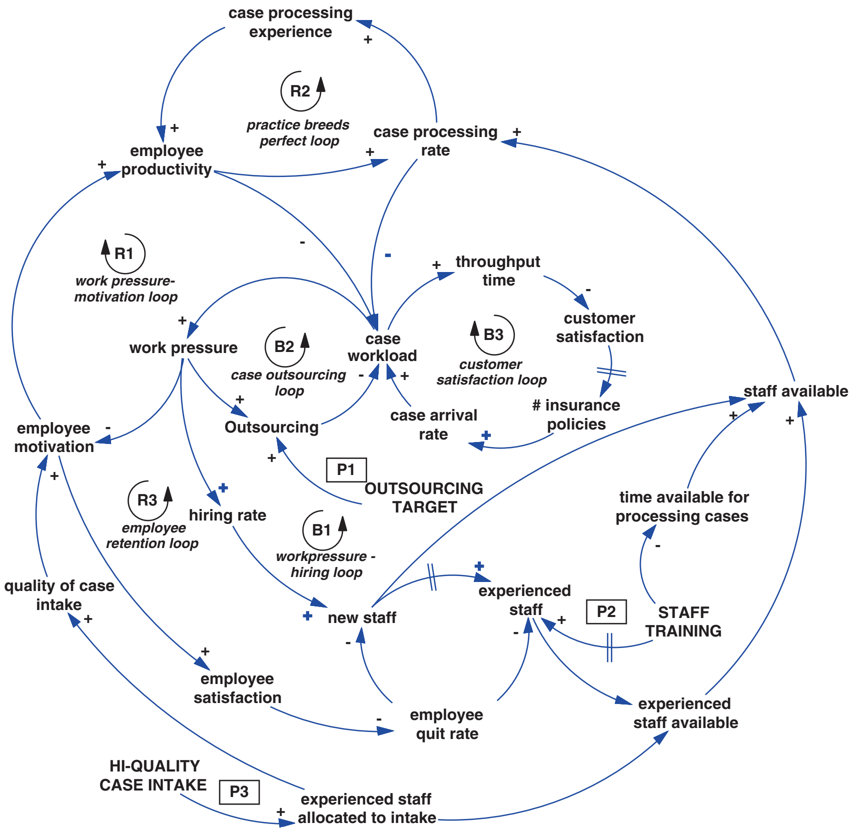


Fig. 1 Causal loop diagram of key interdependencies

In the second stage of the project we, the authors, developed a quantified simulation model for SRB. We started from the causal loop diagrams and the intermediate BSC that had been developed in the first stage. These were sufficient to develop a first skeleton of an SD model. This skeleton was then filled with key company data, which were delivered by two managers from the MT. These two were more closely involved than the others in the subsequent development of the model, critiquing intermediate versions and providing valuable feedback, fulfilling what Richardson and Andersen have called a 'gatekeeper role'.

With the calibrated simulation model that was developed in this manner, we conducted a number of analyses that addressed the questions the

Table 1 Intermediate BSC as developed during the first stage

	Objectives	Measures
Financial perspective	Be able to meet continued demand growth Keep cost levels in line with agreements	Output per employee Percentage outsourcing of cases
Customer perspective	Deliver a good service for a reasonable price Reduce work pressure and employee stress in general	Customer satisfaction Throughput time per case Throughput time per case % of small and easy cases
Process perspective	Improve company agility	Number of successful projects Working at home
Learning and growth perspective	Attract and retain good people Increase collaboration between employees	Employee turnover rates Employee satisfaction Training on the job/coaching Hiring of new staff through referral by colleagues

MT was still grappling with after the first stage. This process led to significant additional managerial insights. Firstly, it was comforting for the team to see that most of the KPIs selected for the intermediate version of the BSC were confirmed by the simulation exercises as key in driving performance in the simulation model.

Secondly, what was less comforting for management to notice was the model's prediction that performance would first deteriorate further before things would get better. The increase in workload that had been building up the past 2 years would stabilize in the year ahead before it would drop significantly. As a result, setting ambitious targets for especially the first half of the coming year was not appropriate. This was a message that none of the action- oriented managers really liked, but one that was key in managing expectations adequately.

A third managerial insight from the simulation effort was that it focused attention on the real leverage points in the company, on those policies that really mattered to performance and away from the ones that were less likely to be high-impact. Among these were, as we will also see further on, the two-edged sword of outsourcing of cases to third- parties and the mixed blessing of additional staff training. Unexpectedly high benefits were found to result from another policy initiative, which was to install a high-quality case intake process staffed with experienced employees.

All these findings have been discussed with the management team, have been challenged by them, have in some cases been mitigated but nevertheless have broadly been accepted. They have been used to guide the implementation of the BSC approach for the organization as a whole and well as for the various sub-units involved.

Qualitative Model Structure

A simplified and stylized version of the qualitative model that was the end result of the first modelling phase is shown in Fig. 1. In this causal loop diagram, six interconnected key feedback loops are shown that together determine the dynamic behaviour of the model. These are labelled R1–R3 with the 'R' standing for 'reinforcing' or positive feedback loop, and B1, B2 and B3, the 'B' standing for 'balancing' or negative feedback loop [16].

R1. *The work pressure-motivation loop*: A pertinent observation the MT made was that, in recent years, work pressure had gone up as a result of increases in requests for legal aid. More work had to be performed by the same people, so work pressure went up. As a result, employee motivation had dropped. Lower motivation leads to more absenteeism, more sick leaves, and, in general, to a reduction in productivity (defined as # cases handled per person per period). Productivity changes affect again the workload, and hence, the work pressure, leading to a vicious cycle of low employee motivation and high work pressure.

R2. *Practice breeds perfect-loop*: How do employees become more proficient in processing cases? Obviously, by training, but mostly by learning-on-the-job. A law degree does not automatically make one an ideal staff member at SRB. For instance, a solution that is legally quite sophisticated may not be the most desirable option from a client or managerial perspective, as it may require long handling times and a great deal of personal attention. In practice, a quick and straightforward settlement may be preferable. Judgments such as these are best learned on the job. Interestingly, the more cases one has handled, the more proficient and productive one becomes. So, the better one gets, the better one becomes: a clear virtuous cycle of mutually reinforcing experience and productivity.

R3. *Employee retention loop*: Employees that feel pressurized by high workload are less motivated. Unmotivated employees are more likely to leave than motivated ones. A higher quit rate further reduces the organization's capacity to handle cases, which increases work pressure even more and leads to even more staff turnover. Obviously, this reinforcing loop works the same way as well: happy employees will be more inclined to stay, making work loads better manageable and everybody even more happy. In the former setting, this structure is called a vicious cycle, in the latter the same structure becomes a virtuous one.

B1. *Work pressure—hiring loop*: Apart from simply waiting until staff becomes sufficiently proficient in handling higher volumes of work, what else can one do? One obvious solution is to hire more people. This is visualized by loop B1, which goes from work pressure to hiring rate, and from there to new staff, to staff available, from there on to case processing rate, to case work load and back again to work pressure. Obviously, this adjustment process takes some time. In the case of SRB, it was estimated that more staff could effectively be hired after some 3–6 months. Then, getting them up to speed, that is, become experienced, would take 1–3 years.

B2. *Case outsourcing loop*: A more short-term fix to high workload levels was to outsource part of the caseload to outside firms. This reduces work pressure immediately. How much was to be outsourced was a point of debate. One obvious reason was that this was expensive. Another, more subtle one, was that every case outsourced was a learning opportunity missed, in the logic of loop R2.

B3. *Customer satisfaction loop*: Perhaps in theory most powerful balancing loop was the one involving the customer. If, as a result of high workloads, processing times become very long and other aspects of performance deteriorate likewise, consumers may decide that they are better off switching to another insurer. However, in the reality of SRB at that time, this was definitely not an immediate concern. Most people will ask for legal aid very infrequently, and even then handling speed would only be one of many drivers of customer satisfaction.

These six feedback loops describe most, but not all the variables and links in the diagram. In particular, they do not fully address the three specific managerial policies that were being considered by the MT at that time, and that found their way into the intermediate BSC. The first one (labelled as P1 in the diagram) was to strongly increase the target for outsourced cases, so as to relieve workloads immediately, and help reverse the employee retention loop from a vicious into a virtuous cycle. The second proposed policy was to train new staff better and hence to boost the company's ability to handle higher workloads of cases. Obviously, when employees are on training they cannot be handling cases at the same time, so this increase in training would come at the expense of staff availability.

Finally, there was a plan to improve the quality of the case intake process dramatically, so that every new case would be allocated to precisely the right person in the organization, as these differed widely in skill levels, fields of experience, hobby's, work style and the like. This was a job that would require the refined judgment of experienced staff, who were at the same time badly needed for handling cases. So, every policy had its potential downside, and from the conceptual discussions during Stage 1 of the modelling process it became clear that quantification would be needed to provide final answers.

Quantitative Model Structure

In the second stage of the BSC definition, a quantified SD simulation model of the insurance company was developed. In this section, the structure of this model is described. As a result of reasons of confidentiality and size, we will not describe the model in full detail. The quantitative relationships between variables are not given in this paper. However, the full model (with fictitious numbers) is available from the authors upon request.

The simulation model is based on the causal loop diagrams described above. Quantified system dynamics modelling asks for specification of the main flows and accumulations, or stocks, in the system. In this case, the two main flows are the *processing of cases* and the flow of employees. The most important stocks are, for the first flow, cases being processed and, for the second flow, employees, both new and experienced. These two main flows are drawn in Figs. 2 and 3. The causal loops connect these two main flows in various ways. In the next paragraphs the contents of the model are explained.

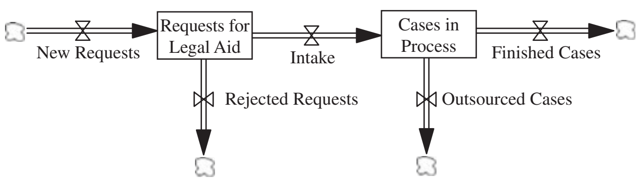


Fig. 2 Stocks-and-flows diagram for cases

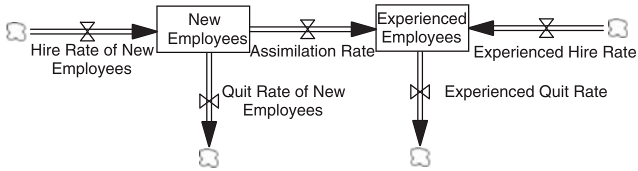


Fig. 3 Stocks-and-flows diagram for employees

Flow of Cases

The legal aid that SRB provides can only be supplied after a so-called intake process. In this process it is decided whether the request for legal aid of a customer can be given by the company. When the intake is accepted, a specific customer file (a 'case' in legal terms) is made and this case is allocated to an employee for further processing. Occasionally, when employees are too busy, a case can be outsourced. When the intake is rejected, the request of the customer is discarded without further processing taking place. This flow of cases is represented in stocks-and-flows notation [16] in Fig. 3.

Flow of Employees

Figure 3 shows the employee flow for new and experienced employees. This distinction is necessary because both turnover rates and productivity differ for new and experienced employees. New employees become experienced after a certain assimilation time. More experience means higher productivity. So, the productivity of experienced employees is considerably higher than that of new employees. However, experienced employees are also needed to train new employees. The time that experienced employees 'lose' decreases their productivity, defined as the number of cases that they can handle per period.

These two main flows, cases and employees, interact in many ways. Most of the feedback loops that were described in the preceding section link these two parts of the model. For instance, loop R3 states that higher workload (flow of cases) leads to more work pressure and hence to more employee turnover (flow of employees). Loop B1 says that higher case load (flow of cases) and more work pressure lead to more aggressive hiring (flow of employees). Quantification of most of these loops is straightforward for a trained SD model builder. One exception is loop R2, the 'practice breeds perfect' loop. In more specific terms, this part of the model calculates productivity. It is based on learning curve theory, in which experience is linked with productivity. For our calculations, we referred to the model provided by Sterman [16], p. 507. His key definition is:

$$\text{Productivity} = \text{Reference Productivity} \times \left(\frac{\text{Average Experience}}{\text{Reference Experience}} \right)^{c}$$

Average Experience of either experienced or new employees is the total experience (expressed in working years) divided by the number of experienced or new employees. Experience can be gained through processing of cases, but experience can also be lost. People forget relevant knowledge and new developments in the insurance sector cause experience to become obsolete. This is expressed in the model by an *experience decay rate*. Furthermore, experience is also lost when employees leave the company.

Reference Productivity is the productivity attained at the *Reference Experience* level. For example, in the simulation model Reference Experience is about 7 working years for experienced employees and 0.2 working years for new employees. Reference Productivity for experienced employees is almost double the reference value for new employees. The exponent c in the computation of the productivity determines the strength of the learning curve and is equal to

$$c = \ln\left(1 + f_p\right) / \ln\left(2\right)$$

where f_p is the fractional change in productivity per doubling of effective experience [12].

Model Analysis

The simulation effort was conducted because the team was insecure about the quality of the balanced scorecard that had been developed on a qualitative basis. After quantification, what were the lessons learned? One overall conclusion was that the performance indicators that had been selected were broadly speaking the right ones, but that the chosen performance improvement indicators were not straightforwardly positive. So, aspects such as employee productivity, throughput time, customer and employee productivity and employee turnover rates were indeed confirmed to be key drivers of performance, also in the quantitative version of the model. Figure 4 shows the behaviour over time for some of the key performance indicators identified. This graph partly replicates history and partly predicts future behaviour, as the modelling effort took place towards the end of Year 2. Although precise time series data were missing for the two preceding years, behaviour was found to be broadly in line with available knowledge of past performance. The gradual build-up of work pressure over time is clearly visible, and the increases in throughput time that are the logical consequence of that as well. Less straightforward but quite explainable is the gradual decrease of employee productivity (measured in Cases per Employee per Year, so $C/(E/Y)$). As there had been a considerable increase of new and inexperienced staff and productivity of new staff is only around half of that of experienced staff, average productivity would have to drop.

Overall Performance: Gradual Progress, But First Worse-Before-Better

What Fig. 4 also shows is that progress would not be made immediately. For the first half of Year 3, the model predicted that performance would still become somewhat worse, and that only around the start of Year 4 work pressure and throughput times would have gone down significantly. So, the good news was that a turnaround appeared to be in the making; the bad news was that SRB was not quite there yet. This finding resulted for management in the uncomfortable recommendation of setting performance targets for the first quarters of the coming year not higher but lower than last year's performance, which is not a natural inclination of most managers.

Existing Policies Insufficient for Swift Progress

Figure 4 shows performance with existing policies. These included policies for hiring new staff. So far, those had been mainly reactive: when

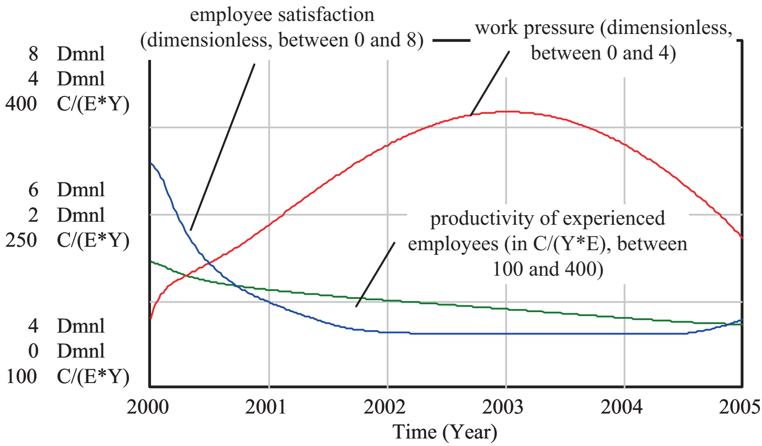


Fig. 4 Work pressure, throughput time and employee productivity over time

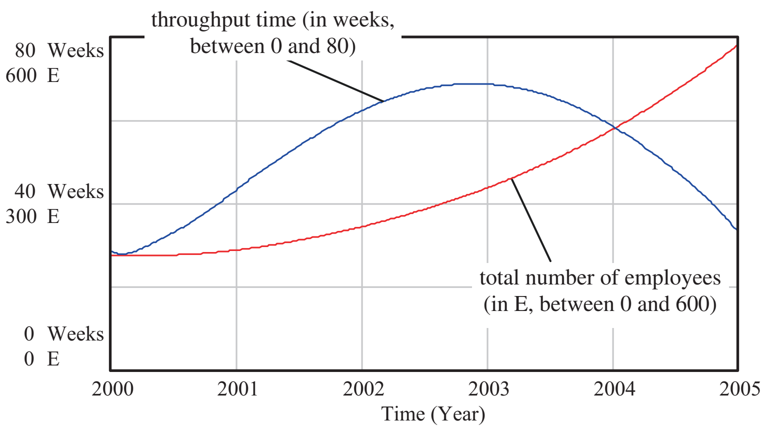


Fig. 5 Delays in increasing hiring rates in response to increased demand rate

work pressure went up, more employees were hired. This also becomes apparent from Fig. 5, where some key performance indicators for the flow of employees are visualized. From this graph it is clear that there are several delays involved. When it is finally noted that work pressure is structurally increasing, it still takes some time before employees are actually hired and then even longer before they become productive and the effect of hiring them becomes visible. For example, in Year 0 the total number of employees was 207. In Year 3 this number was estimated at 330. Between Year 0 and 3 throughput time and work pressure (see Fig. 4) were still increasing.

Unclear Benefits from Proposed Policies

The rate of improvement with existing policies was hard to swallow for management. What could be done to speed up progress? One obvious remedy was to outsource more cases. The system dynamics model showed that indeed, in the short term, outsourcing has a positive effect on work pressure, as shown in Fig. 6. In Year 2, almost 15% of all cases are outsourced and, indeed, between Years 1 and 2 work pressure is somewhat lower than in the scenario of no outsourcing. However, after Year 3, work pressure becomes slightly higher in the outsourcing scenario. This is because, when cases are outsourced, employees cannot gain experience from those cases, and in the long run these missed opportunities have a negative effect on their productivity and, consequently, on work pressure.

This experiment showed that outsourcing might well be a two-edged sword. When applied selectively to special settings (one example included

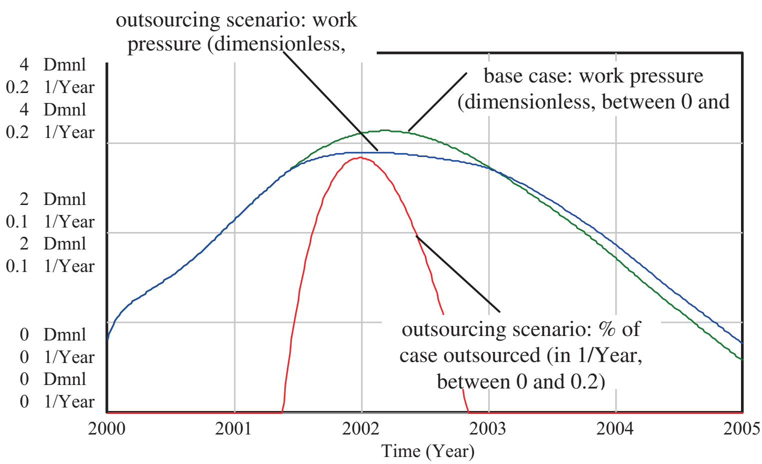


Fig. 6 Effect of outsourcing on work pressure

a department with very experienced staff, so with limited learning opportunities, but a very high case load and a tight job market), outsourcing could work well to alleviate short-term pressures. However, when applied too lavishly it might be counter-productive in the long run. So, it was considered better not to make '% outsourced' a KPI in itself.

There was a similar story to tell for the policy of increased staff training. More staff training would improve productivity and motivation, but it would also reduce the time available to learn from handling real cases. Again, this proposed policy was found to have unclear benefits and was therefore removed from the BSC.

Unexpected Leverage from a Counter-Intuitive Policy

The importance of employee productivity had been repeatedly stressed and confirmed. The crucial importance of experienced staff for overall performance had also been noted several times. As more and more junior staff would come on board, the higher productivity of the experienced employees would be all the more a strategic asset. So, an initiative that would limit the time that experienced staff could spend on handling cases could well be labelled as 'counter-intuitive'. And yet, this was what the initiative of the high-quality case intake process was all about.

When new cases come in, they have to be evaluated and assigned to specific staff members. Experienced employees not only can 'read' a case very early on, they are also familiar with the strengths and weaknesses, preferences and dislikes of their fellow staff members. When more experienced employees are involved in the intake process, it is likely that the fit between cases and employees will improve. The better this fit, the higher employee satisfaction becomes. This operational measure was tested; the results are shown in Fig. 7. In the base case, 50% of the intake of cases is handled by experienced employees (and consequently 50% by new employees). In the high-quality intake-scenario, experienced employees do 75% of the intake. Figure 7 shows that this policy has indeed a significant positive effect on work pressure, and consequently also on employee productivity and throughput times.

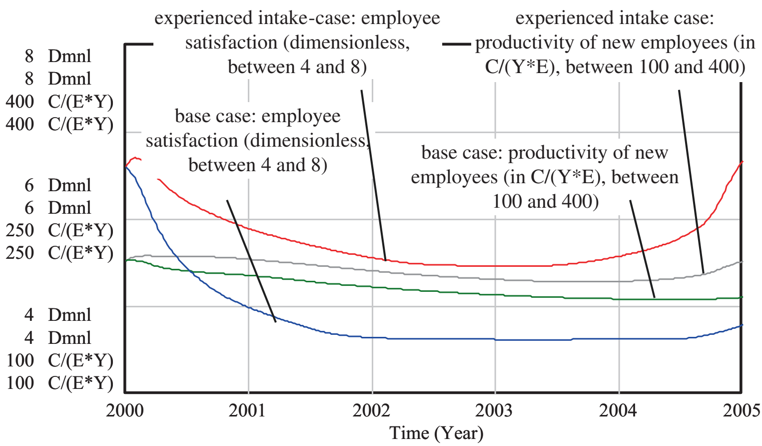


Fig. 7 The effect of installing a high-quality intake process

Discussion

Limitations: Modelling of 'Mental Models', Not of the 'Real World'

There are many limitations to the study we could mention and discuss in this section, for instance that since it reports on a single case, generalizability of findings is problematic. One limitation that is specifically relevant in this case is the distinction one has to make between modelling 'mental models' and modelling 'the real world'. The process that the Interpolis management team went through focuses on making explicit the mental models of the individual managers, on sharing them, on challenging their internal consistency and on aligning them. What this approach did *not* aim or claim to do is to model the 'real world', independent of what the managers' perception of this real world was. The philosophical dimension of this distinction we will not solve in this article, as this goes back all the way to Plato's cave and there still remain two camps of academics: those who insist that all models are social constructions of reality and those who believe that there are at the very least significant elements of objectivity in all social system models [34, 44].

Fortunately, the practical side to this limitation is easier resolved than the philosophical one. We stress that developing a rigorous model of real-world business processes through direct observation is a laudable, but fairly time-consuming process. The approach that we have presented here is intended to supplement a strategic decision-making process. So, the fair comparison to be made is not between the model that one develops through the process we have outlined and some theoretical 'optimal' model. Rather, one should set a BSC development approach informed by system dynamics against a conventional approach of developing BSCs and then see if the benefits of using SD outweigh the drawbacks.

The Added Value of System Dynamics to BSC Development

Management of SRB and Interpolis has remained quite positive about the process described in this article. Generally speaking, the future has unfolded pretty much as had been predicted during the modelling effort. So, after an initial period of seemingly little progress in performance, in fact of further deterioration of throughput times, considerable improvements in operation could be noted, especially in the second half of the year. To what extent was this due to the use of the system dynamics approach? This is a question that is difficult to answer objectively, but SRB management did indicate that both the causal diagramming workshops and the simulation effort had had clear added value for them.

So, some kind of causal diagramming exercise seems worthwhile. Obviously, this is pretty much what Kaplan and Norton themselves have been stressing recently with their notion of 'strategy maps'. And, as observed earlier on, there are other mapping techniques such as SODA and SSM that one could employ equally well. We do not see a distinctive 'competitive advantage' to the use of SD here, rather an approach that is in line with current best practice in BSC development.

The use of simulation modelling, and SD simulation in particular, is far less frequently attempted in BSC development and yet yielded considerable additional insights in this case. In particular, the quantification effort helped to be focused on the importance of a good appreciation of time delays and accumulations. The MT found it essential to recognize that adjusting to the sharp increase in customer demand took a significant amount of time, and that the case load that had accumulated over the past few years would not disappear overnight, even when staffing levels would be in line with current demand rates.

It is doubtful if understanding the importance of delays and accumulations, for which SD seems eminently suited, is essential in all BSC development settings. Certainly, this single case cannot prove that. But, in general, one would expect that specific problem settings ask for specific tools. The case of Santos et al. [39] describes a setting where multicriteria analysis seemed most appropriate at the end of a qualitative modelling stage where SD was applied. We hope that, in the future, BSC development practitioners will continue to 'mix and match' methods where appropriate, and use screwdrivers for screws, hammers for nails, rather than regarding everything as a nail since all they have is a hammer.

Conclusion

Some 2 decades ago, balanced scorecard pioneers Johnson and Kaplan [1] stated that conventional management accounting deserved the label *'Relevance lost'*. As an alternative, they introduced the concept of the BSC, in which the organization tries to focus on a small number of truly relevant indicators to monitor and improve performance. This article has looked at a setting of BSC development for management of one business unit of the Dutch insurer Interpolis. These managers found the BSC concept helpful to arrive at a list of financial and non-financial performance measures, which they saw as the most important ones. However, they were uncertain if these were really the right measures to monitor. Rather than assuming that their scorecard was correct, this management team went through a system dynamics-based approach that was both thorough and practically feasible to question the relevance of the measures it contained.

The use of system dynamics has proved to be very beneficial in this process. The use of causal diagramming was very instrumental during the first stage of modelling in identifying key variables and their causal interrelations. The use of SD simulation modelling was essential in arriving at a proper appreciation of the importance of time delays and accumulations in the key business processes of handling legal cases and of attracting and retaining employees.

System dynamics remains, from the palette of systems interventions available, the technique that, in terms of quantitative modelling, has from the onset been developed to 'boldly go where no one has gone before' in areas where reliable data and theoretical models are lacking but nevertheless the need for simulation, for scenario analysis, is clearly apparent. Surely, balanced scorecard development fits this description well. Although other techniques and approaches will be more appropriate in some cases, SD remains a good choice to test for relevance in a wide variety of BSC development settings.

References

1. Johnson HT and Kaplan RS (1986). *Relevance Lost: The Rise and Fall of Management Accounting*. Boston: Harvard Business School Press.
2. Kaplan RS and Norton DP (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review* 70: 71–79.
3. Kaplan RS and Norton DP (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
4. Kaplan RS and Norton DP (2001). Leading change with the balanced scorecard. *Financial Executive* 17: 64–66.
5. Dinesh D and Palmer E (1998). Management by objectives and the Balanced Scorecard: will Rome fall again? *Management Decision* 36: 363–369.
6. Neely A, Gregory M and Platts K (1995). Performance measurement system design: a literature review and research agenda. *International Journal of Operations and Production Management* 14: 80–116.
7. Bourne M et al. (2000). Designing, implementing and updating performance measurement systems. *International Journal of Operations and Production Management* 20: 754–771.
8. Martinsons M, Davison R and Tse D (1999). The balanced scorecard: a foundation for the strategic management of information systems. *Decision Support Systems* 25: 71–88.
9. Hill T (1989). *Manufacturing Strategy. The Strategic Management of the Manufacturing Function*. London: Macmillan.
10. Ackoff RL (1981). *Creating the Corporate Future. Plan or be Planned for.* Chichester: Wiley.

11. Forrester JW (1992). Policies, decisions and information sources for modelling. *European Journal of Operational Research* 59: 42–63.
12. Kaplan RS and Norton DP (2000). Having trouble with your strategy? Then map it. *Harvard Business Review* 78(5): 167–178.
13. Kaplan RS and Norton DP (2004). *Strategy Maps. Converting Intangible Assets into Tangible Outcomes.* Boston: Harvard Business School Press.
14. Senge P (1990). *The Fifth Discipline. The Art and Practice of the Learning Organisation*. New York: Doubleday Currency.
15. Vennix JAM (1996). *Group Model Building. Facilitating Team Learning Using System Dynamics.* Chichester: Wiley.
16. Sterman JS (2000). *Business Dynamics. Systems Thinking and Modelling for a Complex World.* New York: McGraw-Hill.
17. Haas M de (2000). *Strategic Dialogue: In Search of Goal Coherence.* PhD thesis, Eindhoven University of Technology.
18. Neely A (1998). *Measuring Business Performance.* London: Profile Books.
19. Hepworth P (1998). Weighing it up—a literature review for the balanced scorecard. *Journal of Management Development* 17: 559–563.
20. Malmi T (2001). Balanced scorecards in Finnish companies: a research note. *Management Accounting Research* 12: 207–220.
21. Wisniewski M and Dickson A (2001). Measuring performance in Dumfries and Galloway Constabulary with the balanced scorecard. *Journal of the Operational Research Society* 52: 1057–1066.
22. Newing R (1994). Benefits of a balanced scorecard. *Accountancy* 114: 52–53.
23. Nørreklit H (2000). The balance on the balanced scorecard—a critical analysis of some of its assumptions. *Management Accounting Research* 11: 65–88.
24. Flapper SD, Fortuin L and Stoop PPM (1996). Towards consistent performance management systems. *International Journal of Operations and Production Management* 16(7): 27–37.
25. Kleijnen JPC and Smits MT (2003). Performance metrics in supply chain management. *Journal of the Operational Research Society* 54: 507–514.
26. Mooraj S, Oyon D and Hostettler D (1999). The balanced scorecard: a necessary good or an unnecessary evil. *European Management Journal* 17: 481–491.
27. Hudson M, Smart A and Bourne M (2001). Theory and practice in SME performance measurement systems. *International Journal of Operations and Production Management* 21: 1096–1115.
28. Tan KH, Platts K and Noble J (2004). Building performance through in-process measurement: toward an 'indicative' scorecard for business excel-

lence. *International Journal of Productivity and Performance Management* 53(3): 233–244.

29. Lane DC (1992). Modelling as learning: a consultancy methodology for enhancing learning in management teams. *European Journal of Operational Research* 59: 64–84.
30. Winch GW (1993). Consensus building in the planning process: benefits from a 'hard' modelling approach. *System Dynamics Review* 9: 287–300.
31. Akkermans HA (2001). Renga: a systems approach to facilitating inter-organisational network development. *System Dynamics Review* 17: 179–194.
32. Rosenhead J (ed) (1989). *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict.* Chichester: Wiley.
33. Pidd M (1996). *Tools for Thinking. Modelling in Management Science.* Chichester: Wiley.
34. Eden C (1989). Using cognitive mapping for strategic options development and analysis (SODA). In: Rosenhead (ed). *Rational Analysis for a Problematic World*. Chichester, UK: Wiley, pp 21–42.
35. Checkland P (1981). *Systems Thinking, Systems Practice*. Chichester: Wiley.
36. Eden C (1994). Cognitive mapping and problem structuring for system dynamics model building. *System Dynamics Review* 10: 257–276.
37. Eden C, Wiliams T, Ackermann F and Howick S (2000). On the nature of disruption and delay (D&D) in major projects. *Journal of the Operational Research Society* 51: 291–300.
38. Howick S (2003). Using system dynamics to analyse disruption and delay in complex projects for litigation: can the modelling purposes be met? *Journal of the Operational Research Society* 54(3): 222–229.
39. Santos SP, Belton V and Howick S (2002). Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations and Production Management* 22(11): 1246–1272.
40. Warren K (2002). *Competitive Strategy Dynamics.* Chichester: Wiley.
41. Warren K (2003). *The Critical Path. Building Strategic Performance Through Time.* London: Vola Press.
42. Wernerfelt B (1984). A resource-based view of the firm. *Strategic Management Journal* 5(2): 171–180.
43. Dierickx I and Cool K (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science* 35(12): 1504–1511.
44. Flood RL and Jackson MC (1991). *Creative Problem Solving: Total Systems Intervention*. Chichester: Wiley.

Interpersonal Success Factors for Strategy Implementation: A Case Study Using Group Model Building

Rodney J. Scott, Robert Y. Cavana, and Donald Cameron

Introduction

Success factors for strategy implementation include commonly reported outcomes of group model building—indeed, it was this simple observation that led the authors to conduct the research described in this paper. Interpersonal success factors for strategy implementation are communication quality (Hambrick and Cannella 1989), insight (Wooldridge and Floyd 1990), consensus (Floyd and Wooldridge 1982) and commitment (Kim and Mauborgne 2005). A review of 107 papers revealed that these are commonly reported outcomes of group model building interventions

Journal of the Operational Research Society (2015) **66**(6), 1023–1034.
https://doi.org/10.1057/jors.2014.70
Published online 9 July 2014.

R.J. Scott (✉) • D. Cameron
School of Agriculture and Food Sciences, The University of Queensland, Brisbane, QLD, Australia

R.Y. Cavana
Victoria University of Wellington, Wellington, New Zealand

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_5

(Rouwette et al. 2002), suggesting possible applicability. Yet little has been written about how group model building can support implementation (Sterman 2000). Group model building is not a strategy implementation plan (it does not address all of the success factors for strategy implementation, nor enact change in and of itself), but the literature suggests that it may make a positive contribution.

On the basis of this coincidence of success factors, a case study was conducted. This observational study involved New Zealand public servants completing a short group model building workshop to plan how predetermined strategic objectives would be implemented. This paper presents the initial evaluation of that case study, using an established survey method. A change in circumstance prevented the completion of the planned intervention. The paper discusses what we can conclude from the initial results, and presents a case for why this area is a rich opportunity for the group model building community that warrants further research.

This paper is of interest to academics in informing their research agenda, to group model building practitioners in exploring a potential setting (strategy implementation) for their practice, and for the management community in identifying a new tool to support strategy implementation.

Strategy Development and Implementation

Most strategies fail, with authors generally claiming failure rates between 50 and 90% (Kiechel 1982, 1984; Gray 1986; Judson 1991; Nutt 1999; Kaplan and Norton 2001; Raps 2005; Sirkin et al. 2005). One study demonstrated that most managers believe that the difficulty of implementing a strategy surpasses the difficulty of formulating it (Zairi 1996). Despite this, management (and research) attention remains focussed almost exclusively on improving strategy development (Raps 2005). Strategies can fail for many reasons (Raps 2005), but many of these are through factors internal to the group or organisation, rather than external (Nutt 2002). Authors have long urged co-creation and meaningful participation in strategy development (Floyd and Wooldridge 1982; Guth and Macmillan 1986; Nutt 1999, 2002). Several have gone further and cautioned against making a distinction between strategy development and implementation, preferring instead to consider strategy as a practice (Whittington 1996), or as an opportunity for organisational learning (De Geus 1997) or double-loop learning (Argyris 1989).

There is some evidence that application of these methods has led to improved success rates (Walsh et al. 2002; Sila 2007). Nonetheless, many strategies remain products of episodic and top-down decisions (Lyneis 1999), with resultant barriers to employee acceptance (Noble 1999; Nutt 2002). This paper addresses the common problem of how to implement predetermined strategic objectives with little prior employee involvement in their creation.

Implementation Success Factors

Literature on the nature of strategy implementation and the reasons for its success and failure is not well-organised or agreed (Yang et al. 2008). Conversely, the success factors that predict effective strategy implementation are the subject of more agreement (Skivington and Daft 1991; Noble 1999). Strategy implementation literature can be divided into two broad categories: structural views, and interpersonal process views (Skivington and Daft 1991).

Structural views include organisation (re-)structure (Bain 1968; Miles and Snow 1978; Porter 1980; Drazin and Howard 1984; Gupta 1987) and control mechanisms (Jaworski and MacInnis 1989; Jaworski et al. 1993) including monitoring systems (Daft and Macintosh 1984), and performance management systems (Kaplan and Norton 1996). Organisation structure and control mechanisms are direct tools available to managers in shaping their organisation (Skivington and Daft 1991), and include popular templates and models (Kaplan and Norton 1996, 2001; Langfield-Smith 1997).

However, as strategies are executed by people, a range of interpersonal and cognitive factors may also be critical (Noble 1999), and may be further divided into a number of subcategories: communication quality, insight, consensus and commitment. The theory of planned behaviour

(Ajzen 1991) suggests strong interrelationship and a logical sequencing of these success factors; communication quality fosters insight and consensus, and insight and consensus contribute to commitment (Rouwette 2003). This paper focuses on the interpersonal success factors, as the methods for supporting these success factors are less well-understood and agreed (Noble 1999; Yang et al. 2008).

Communication Quality

Many authors identify the impact of communication quality between managers and staff on strategy cognition, and therefore implementation (Argyris 1989; Sandy 1991; Workman 1993; Kim and Mauborgne 2005). Different authors have focussed on vertical communication between leaders and staff (Fidler and Johnson 1984; Robertson and Gatignon 1986; Johnson and Frohman 1989) and horizontal communication between peer groups (Hambrick and Cannella 1989).

Insight

Several authors explore the importance of novel insight by employees in creating effective strategy implementation (Wooldridge and Floyd 1989; Redding and Catalanello 1994; Baum and Greve 2001; Tang 2011).

A centrally created, highly detailed plan faces several limitations. Business units have detailed knowledge of their subject area that may not be known to central planners (Floyd and Wooldridge 1982), and face operating environments that may be subject to change (Hrebiniak and Snow 1982; Hrebiniak and Joyce 1984). While consensus and adherence may be useful in high-level goals, employees must interpret how to put these goals into practice within the context of their own detailed area (Floyd and Wooldridge 1982; Bonoma 1984, 1986; Bonoma and Crittenden 1988).

Employee insight in co-creating lower-level actions has also been positively associated with commitment to the strategy (Bonoma and Crittenden 1988). This leads to the popular diffusion approach where high-level guidance is provided centrally, and business units create more detailed planning in a 'trickle-down' manner (Fidler and Johnson 1984; Robertson and Gatignon 1986; Kaplan and Norton 2004). Effective strategy implementation is supported by providing employees with the opportunity to combine their knowledge with the direction provided by the strategy, to produce new insights (Crittenden and Crittenden 2008).

Consensus

The degree of unanimity and agreement of a group is positively associated with implementation success (Floyd and Wooldridge 1982; Schweiger et al. 1989; Wooldridge and Floyd 1990; Noble 1999). Low agreement is associated with implementation failure (Guth and Macmillan 1986; Huy 2011). Agreement between different business units is often to ensure coordinated deployment of the strategy (Wooldridge and Floyd 1990). However, premature agreement (before sufficiently understanding the problem, or without consideration of alternate solutions) is associated with poorer decision-making (Schweiger et al. 1989).

Commitment

The level of dedication to strategy implementation predicts implementation effectiveness through both intensity of commitment (Nutt 1983; Wooldridge and Floyd 1989; Kim and Mauborgne 2005), and durability of commitment (Bourgeois 1980; Nutt 1983, 1986, 1990; Bourgeois and Brodwin 1984). The hand-over of strategy from senior to middle management can be problematic—middle management may be apathetic to strategies they have not been involved in developing (Whitney and Smith 1983).

Open Issues in Strategy Implementation

The interpersonal success factors described above are the subject of strong agreement in the strategy implementation literature (Noble 1999), but the methods for achieving these factors are still unclear (Yang et al. 2008).

This paper explores whether group model building can contribute to achieving these interpersonal factors in a strategy implementation context.

Group Model Building and Strategy

System dynamics has been applied to many disciplines and subject areas (Andersen et al. 2007; Mingers and White 2010). One area in which system dynamics has been particularly prevalent is in strategy development (Pidd 2004; Rouwette 2011). Some have argued that the reason for this applicability is the complex and interrelated choices that strategy presents (Broman et al. 2000; Aligica 2005; Houchin and MacLean 2005).

Despite strong links to strategy development, the use of group model building to support strategy implementation remains an area requiring further research (Sterman 2000). One study attempts to use system dynamics as a communication tool to build support for strategy implementation (Snabe and Größler 2006), rather than as a participatory modelling process.

On the basis of the review by Andersen et al. (1997) of the existing group model building literature, Rouwette et al. (2002) identified a number of outcomes associated with group model building interventions. These occur at four levels:

- individual: positive reaction, insight, commitment and change in behaviour;
- group: mental model alignment, communication quality and consensus;
- organisation: system changes and system results;
- method: efficiency and further use.

These 11 outcomes include the four interpersonal success factors for strategy implementation: communication quality, insight, consensus and commitment to a decision. This suggests potential applicability for group model building in supporting effective strategy implementation. The case study described below was developed to test this applicability, focussing on the hand-over of strategy from senior to middle management for implementation.

Case Study

The leadership team from a large New Zealand government agency had developed a long-term strategy, and then turned their attention to how it would be implemented. The strategy included a 20-year vision, and four strategic objectives: maximising export opportunities; improving sector productivity; increasing sustainable resource use; and protecting from biological risk. Particular implementation concerns from senior staff included that: the strategy may be poorly understood, or there may be differences in interpretations; no plan exists for the actions that the organisation should take to realise the intent set out in the strategy; and those responsible for implementing the strategy did not participate in its development, and therefore may not feel a sense of ownership.

Senior management tasked a facilitator with working with a range of opinion leaders and 'influencers' (Patterson et al. 2008) in the organisation (middle managers and subject experts) to solve three problems, each of which had behavioural and interpersonal aspects: creating a common understanding of the strategy, creating agreed implementation actions for the strategy and increasing commitment to the strategy. The goals of the intervention were therefore different from those of most case studies in the group model building literature, where the client is seeking robust decisions or system changes (Andersen et al. 1997). In this setting, communication, insight, consensus and commitment were mandatory; robust policies were a secondary concern.

This context touches upon many of the themes and challenges in achieving interpersonal success factors. Communication quality must be achieved vertically (hand-over) and horizontally (shared understanding); the case study involves middle managers and subject experts interpreting the vertical transmission of the strategy, and communicating horizontally with each other to develop shared understandings and agreed plans. Insight has been identified as important in strategy diffusion; the case study involves participants who were expected to be involved in detailed implementation. Consensus on many implementation actions is required between different business units; the case study includes participants from across the business including policy, operations and support functions. Finally, commitment often breaks down at the hand-over from senior to middle management; the case study includes only employees who were not involved in the creation of the strategy.

The facilitator gained senior management agreement to use group model building techniques to convert the strategic objectives into plans for action. Senior managers identified 52 participants based on their perceived level of peer influence, their role in implementation and their subject expertise. The facilitator was an employee of the organisation with experience facilitating qualitative group model building workshop. The employee had used these tools in approximately 20 workshops (as facilitator or participant) prior to the case study, and also had some general facilitation experience.

The strategy consisted of four separate objectives, and participants were split into four groups to work independently on what actions should be taken to implement each objective. Each group participated in a 3-h facilitated workshop using qualitative system dynamics techniques. The workshops contained five main elements as follows:

1. defining the problem or situation (15 min),
2. identifying variables (30 min),
3. describing behaviour over time of the main variables (30 min),
4. constructing causal loop diagrams (75 min), and
5. identifying leverage points (30 min).

These elements are commonly described and relatively easy for a novice group to use (Richardson and Pugh 1981; Sterman 2000; Maani and Cavana 2007). There were two important differences between the workshop elements as commonly described, and as used in this study. First, the causal loop diagrams were completed without participants labelling polarity. The first group quickly identified limitations in the use of polarity (in distinguishing between information and conserved flows—Richardson 1986, 1997) when constructing their causal loop diagram.

To reduce confusion, the facilitator instructed participants to discuss polarity when linking variables, but they were not recorded on the diagram (more consistent with the approach utilised by Senge 1990). Second and consequently, loops could not be labelled as reinforcing or balancing. The facilitator instructed participants to find and trace loops in the diagram, and to explain the behaviour of the loop through narrative. In previous workshops, the facilitator had used polarity and labelled loops as balancing or reinforcing. The causal loop diagrams were constructed using a white-board (drawing causal relationships) and post-it notes (variable names), and computers were not used during the workshop. An example output causal loop diagram is included in Fig. 1. More information on these sessions is available in a previous publication (Scott et al. 2013, and associated online supplementary material).

The facilitation style was focussed on ensuring that the participants followed the process rules. Participants completed all of the workshop elements themselves (the facilitator acted as a guide but did not contribute directly). Most participants had no prior exposure to group model building or system dynamics.

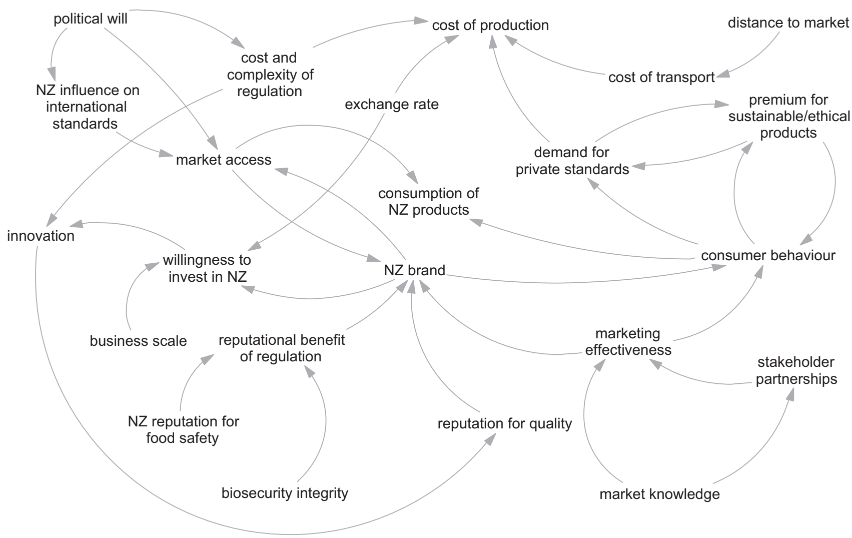


Fig. 1 Causal loop diagram for group 2 'What are the factors that influence New Zealand's export opportunities in the food and fibre sectors?'

Survey Questionnaire

Most studies have used anecdotal or descriptive evidence in evaluating group model building—only a small number attempted quantitative assessment (Rouwette et al. 2002). This study uses a range of survey questions administered at the conclusion of the workshop (the 'CICC' questionnaire, Vennix et al. 1993). These questions have been used before in evaluating group model building interventions (Akkermans et al. 1993; Vennix et al. 1993; Vennix and Rouwette 2000; Rouwette 2011). Other studies have used similar questions (McCartt and Rohrbaugh 1989; Huz et al. 1997) or a subset of the questionnaire (Dwyer and Stave 2008). When this survey was previously used in combination with semi-structured interviews, the interviews revealed no significant new information (Rouwette and Vennix 2006).

This method assesses the level of consensus, but not the quality of that consensus. Premature consensus can result in poorer decisions (Schweiger et al. 1989). One theoretical paper proposes a method for evaluating the degree to which discussion of problems precedes discussion of solutions in group model building (sequential analysis—Franco and Rouwette 2011), which could be an area for future study.

The questionnaire included: demographic information; Likert-scale questions on workshop outcomes; scaled ratings of different elements of the workshop; and open questions with space for written answers. The demographic questions concerned participants' age, gender, education, length of employment and level within the organisation (see Table 1).

The Likert-scale questions measured participants' rating of how the workshops contributed to communication quality, insight, consensus and commitment. Twenty-three questions evaluated these outcomes in absolute terms (e.g., 'My insight into the problem has increased due to the modelling process'), and a further seven questions evaluated outcomes in the workshop compared with 'normal meetings or conferences in which you discuss similar problems' (e.g., 'These meetings give more insight compared with normal meetings'). Each question used a five-point scale from 'strongly agree' to 'strongly disagree'.

Questions for each of the outcomes were assessed for scale reliability using Cronbach's α (a measure, between 0 and 1, of internal consistency between multiple questions evaluating the same factor, which treats any

Table 1 Demographics of participants

	All participants (*n* = 52)	Completed questionnaire (*n* = 40)
Age		
Mean	46 years	46 years
Range	29–64 years	31–64 years
No response	2	2
Length of employment		
Mean	11 years	10 years
Range	1–40 years	1–40 years
Gender		
Male	32	27
Female	18	13
No response	2	0
Organisational level		
Directors	5	3
Group manager	16	15
Team manager	6	2
Non-manager	25	19
Highest qualification		
Postgraduate	37	29
Undergraduate	14	10
Completed secondary	1	1

covariance among items as true score variance—Allen and Yen 2002). As some questions were phrased negatively (e.g., 'We could *not* reach a consensus'.), results were normalised so that in all cases, the value 5 was associated with strong agreement that the outcome was achieved. One question ('The model developed in the session is my own'.) had a correlation of less than 0.20 with the rest of the scale, and was removed (Allen and Yen 2002). The remaining questions all had Cronbach's α of 0.74 or higher, which the authors considered acceptable (Kline 1999, see Table 2).

Further questions evaluated the contribution of different elements in the workshop. The scaled ratings of individual components of the workshop used an 11-point scale from 'was of no use whatsoever, obstructed the session' (−5) to 'contributed very much' (+5). In each case, participants were asked to specify how much each aspect contributed to the overall effect of the workshop. This scale has featured in previous studies

Table 2 Scale reliability of CICC questionnaire

Outcome	Cronbach's α
Communication quality	0.77
Insight	0.76
Consensus	0.77
Commitment	0.74

that use the CICC questionnaire (e.g., Rouwette 2011), but some questions were altered to apply to the methods used in these case studies:

1. The opportunity for open and extensive discussion.
2. The presence of a designated facilitator.
3. The use of behaviour-over-time graphs.
4. The identification of variables.
5. The use of causal diagrams.
6. The identification of leverage points.
7. The use of structured agenda.

The questionnaire included the opportunity for participants to contribute handwritten suggestions to improve the process (Rouwette 2011), with three spaces (boxes) for answers to each of three questions:

1. What were the three best features of the session?
2. What were the three most disappointing features or problems of the session?
3. What specific suggestions would you make if meetings like these were to be organised or held again?

Completed questionnaires were received from 40 of 52 participants (see Table 1). Those who did not complete the evaluation had left the workshop early due to other commitments. As the selection of participants was already non-random, the noncompletions are unlikely to add any new source of error.

As with other studies using this questionnaire, there was no control group. Participants were asked to compare the workshops to a hypothetical 'normal' meeting, and this provides some measure of comparison. A side-by-side comparison would be preferable but this is difficult to achieve in a business setting. Previous studies have used student groups completing hypothetical problems, which may lack external validity (McCardle-Keurentjes et al. 2009), or have used control groups that may not be comparable to the treatment group (Huz 1999; Dwyer and Stave 2008).

Questionnaire results are self-reported, and individuals are typically unreliable reporters of their cognitive and behavioural change (Nisbett and Wilson 1977), and this has previously been noted as a limitation of this questionnaire (Rouwette 2011). Other authors have begun to experiment with methods that do not rely on self-reported change, but these are labour-intensive and have not yet been replicated (McCardle-Keurentjes et al. 2008; Franco and Rouwette 2011; Van Nistelrooij et al. 2012; Scott et al. 2013).

Results

The experimental design was not based on strictly formed *a priori* hypotheses. The data are analysed in several ways, increasing the potential for familywise error (false discoveries due to testing multiple hypothesis—Hochberg and Tamhane 1987). A statistical correction for testing multiple hypotheses (e.g., Bonferroni correction, Shaffer 1995) cannot be completed due to the absence of strictly formed hypotheses. The quantitative data were analysed using common statistical methods. A Kolmogorov-Smirnov test was used to confirm that results were normally distributed, and where significance is discussed, this was measured using a two-tailed Student's t-test (Stephens 1974) to compare the recorded results against a neutral response ('neither agree nor disagree').

Survey Results from Likert Questions

A mean score significantly higher than neutral was recorded for all four outcome areas ($p < 0.01$ compared to 'a/d = neither agree nor disagree' for communication quality, insight, consensus and commitment—see Table 3). This was consistent with results from three other published results using the same tool (Vennix et al. 1993; Vennix and Rouwette 2000; Rouwette 2011). In other studies, qualitative-only workshops were

Table 3 Likert questionnaire results by outcome-area (all $p < 0.01$ above neutral response)

	n	Mean
Communication quality	40	4.04
Insight	40	3.81
Consensus	40	3.68
Commitment to conclusions	40	3.66

1 = strongly disagree that the outcome was achieved, 3 = neither agree nor disagree ('neutral response'), 5 = strongly agree the outcome was achieved

Table 4 Likert questionnaire results compared to a normal meeting (all $p < 0.01$ above neutral response)

	n	Mean
More insight	39	4.05
Faster insight	39	3.88
Better communication	39	4.07
Faster consensus	39	3.85
More clear consensus	39	3.80
Faster commitment	39	3.43
More commitment	39	3.57

1 = strongly disagree compared to a normal meeting, 3 = neither agree nor disagree ('neutral response'), 5 = strongly agree compared to a normal meeting

associated with lower levels of consensus and commitment (Rouwette et al. 2002), but results in this study were comparable to previous studies that included quantitative components.

Survey Results Comparing Group Model Building to a 'Normal Meeting'

Again, a mean score of higher than neutral was recorded for all four outcome areas (communication quality, insight, consensus and commitment) compared to a hypothetical 'normal' meeting ($p < 0.01$ compared to 'a/d = neither agree nor disagree' for communication quality, insight, consensus and commitment—see Table 4). These case studies did not include a control group and therefore do not establish the changes that would be associated with a normal meeting—asking participants to compare the workshop to a hypothetical meeting is one way to investigate this difference. The results indicate that the participants felt the process was significantly more effective than a hypothetical 'normal' meeting. This was broadly consistent with results reported in previous studies (Vennix et al. 1993; Vennix and Rouwette 2000); however, Vennix et al. (1993) had reported an ambiguous result for the speed of commitment generation.

Survey Results Relating to Different Workshop Elements

Questions that asked participants about different elements in the workshop showed strong support for six of seven elements ($p < 0.01$, compared with '0 = did not obstruct, but was of no use either'—see Table 5). For the seventh question (the use of behaviour-over-time graphs), there was a less significant result ($p < 0.05$), meaning that participants felt that the use of behaviour-over-time graphs contributed only marginally to the sessions. The results of this section are not directly comparable with other studies as different workshop elements were used; however Vennix et al. (1993) also report positive ratings for elements in common between both studies: opportunity for open discussion (mean = 3.42), use of causal loop diagrams (mean = 3.46) and the presence of a facilitator (mean = 3.80).

Relationship Between Demographic Data and Survey Results

Demographic data were compared with the results from the Likert-scale questionnaire (results for communication quality, insight, consensus and commitment), and results for each of the workshop elements, using

Table 5 Questionnaire results for different workshop elements

	n	Mean
Opportunity for open discussion	37	+3.26**
Presence of a facilitator	37	+3.10**
Use of behaviour-over-time graphs	30	+1.59*
Identification of variables	40	+3.43**
Use of causal loop diagrams	39	+3.43**
Identification of leverage points	38	+3.45**
Use of structured agenda	35	+3.03**

(–5 = no use, +5 = contributed very much) *$p < 0.05$ above '0', **$p < 0.01$ above '0' a linear regression analysis. Non-managers were more likely to rate the presence of a facilitator and the use of a structured agenda as contributing to the outcomes of the workshop, but these were seen as positive elements by both managers and non-managers. This result had not been anticipated, but may be explained as a way for less powerful participants to ensure their views are considered. Many authors have explored the ability of an independent facilitator to reduce the effect of power imbalances between participants (Schwartz 1994; Heron 1999; Tropman 2003; Rees 2005). One pilot study found group model building is associated with 'power-levelling' (reducing the impact of power imbalances on communication—Van Nistelrooij et al. 2012). Comparing group model building with other facilitation techniques (and the effect of power imbalances in each setting) may be an area for further exploration.

Other relationships shown in Table 6 are not easily explained from the literature, and have not previously been reported using the CICC questionnaire. Previous studies had observed a relationship between participants' rating of the presence of a facilitator and the commitment generated in the workshop (Vennix et al. 1993; Vennix and Rouwette 2000), which could not be established in the case studies in this paper. These previous studies report other relationships that could not be replicated either by each other or the case studies in this paper. The large number of potential comparisons between the different data sources increases the possibility for familywise error (see at the beginning of the Results section, above), and therefore the relationships identified in Table 6 should be considered only as potential starting points for further study.

Open Questions

Participants were asked to describe the three best features, three most disappointing features and make three suggestions for how to make the workshops better.

The most popular features were the participants' ownership of the causal loop diagrams (identified by 18 out of 33 respondents), the

Table 6 Relationships between Likert-scale results, demographics and ratings of different workshop elements

	This paper	Vennix and Rouwette (2000)	Vennix et al. (1993)
Positive relationships between data sources	• Non-manager, and presence of a facilitator • Non-manager, and use of a structured agenda • Older age, and use of causal loop diagrams • Older age, and the identification of leverage points • Post-graduate qualifications, and increased consensus • Post-graduate qualifications, and increased commitment	• Open discussion, and communication in student groups • Causal diagrams, and communication in manager groups • Causal diagrams, and insight in student groups • Causal diagrams, and consensus (all groups) • Open discussion, and commitment in student groups • Presence of a facilitator, and commitment in manager groups	• Presence of a facilitator, and commitment • Open discussion and insight • Visible projection of diagrams, and shared vision

communication between participants (15 of 33), diverse participation (12 of 33) and the presence of a facilitator (10 of 33).

Participants described the duration of the workshop (3 h) as too short (6 out of 24 respondents), too long (3 of 24) and about right (1 of 24). Those that described the workshop as too short mentioned a ‘rushed conclusion’ or ‘not being able to take it (the process) as far as we could’. Those that described the workshop as too long mentioned that it was ‘exhausting’ and ‘tiring’. The only repeated suggestion for improvement was that pre-reading should have been provided to participants so they knew what to expect from the workshop process (7 out of 22 respondents). Other suggestions included ‘more guidance on identifying variables’, ‘reduced scope’, ‘ensure … all the right people (are present)’ and suggestions regarding the workshop venue.

Post-workshop Events

Following the workshop, a major merger and restructure was announced in the case study organisation. This meant that the strategy was not implemented in the manner intended. This also prevented the assessment of implementation success factors over time. Both the intensity and the duration of commitment are important in strategy implementation (Nutt 1983). This paper considers only the intensity of that commitment. Doyle and Ford (1998) proposed that a key challenge for the group model building community was to establish the stability of any changes brought about by brief intervention.

Discussion

The literature describing strategy implementation is fragmented and poorly supported by quantitative evidence (Yang et al. 2008), although there is more agreement on success factors that predict effective strategy implementation (Noble 1999).

Strategy implementation literature and group model building literature exhibit remarkable coincidences. Interpersonal success factors for strategy implementation overlap with reported outcomes of group model building. This provides a theoretical basis for applying group model building to support strategy implementation.

The four success factors are reported to be related to each other: communication quality fosters insight and consensus, and insight and consensus contribute to commitment (Rouwette 2003), but it remains unclear why and how group model building supports these outcomes (Rouwette et al. 2011).

To understand why group model building may be particularly suited to increasing support for existing strategy decisions, it is necessary to delve deeper into a range of reported cognitive biases in the psychology literature. Group model building creates conditions for several cognitive biases that would appear to support agreement and commitment in the hand-over of strategy from senior to middle management for implementation. A similar approach has been used to explain the success of multiple scenario development—that certain aspects of the process reinforce certain cognitive biases to counteract others (Schoemaker 1993). The biases that support agreement and commitment in the hand-over of strategy for implementation fall into four main categories: endowment/empowerment; assembly completion; competence/effectance; and tactile interaction.

Some biases apply to the context for the case study, regardless of the process used. Entrusting middle management with planning the implementation through group model building may create an endowment effect, where individuals prefer things of which they have been given ownership (Kahneman et al. 1990). Transferring ownership increases the power of participants, and this has been shown to increase feelings of engagement (empowerment leadership—Conger and Kanungo 1988).

Conversely, several elements particular to the group model building process are likely to support agreement and commitment in this context. The case study represents an example of not only endowment, but of partial assembly. The IKEA effect' is a cognitive bias where individuals place a disproportionally high value on things that they partially created (Mochon et al. 2012; Norton et al. 2012). Using group model building to plan strategy implementation provides an opportunity for completion of a partially assembled product, which is thought to increase both agreement (measured as thoughts of positive attributes, Carmon et al. 2003) and commitment (measured as positive affect and emotional attachment, McGraw et al. 2003).

Group model building may provide conditions for effectance motivation. Group model building can be taxing on participants—in the case study, it was described as 'exhausting' and 'tiring'. Individuals are likely to have more positive feelings for objects created through great effort (Aronson and Mills 1959). Novice participants in group model building workshops are required to learn several new skills through their participation, which may result in a novelty effect, where performance initially improves as a result of increased interest in the novelty of new techniques (Clark and Sugrue 1988). Causal loop diagrams can appear foreign and complex, yet they can be created by novice participants. The identification of leverage points for interventions provides a sense of achievement—participants quickly came to an agreement that emerged mysteriously from the complexity. Individuals place a higher value on experiences where they are able to demonstrate competence (Franke and Piller 2004), and are more supportive of conclusions that they associate with successful completion of a complex task (Bandura 1977). The greater the (apparent) complexity, the greater the positive association (Thompson and Norton 2011). The combination of empowerment and competence creates the conditions for effectance motivation (White 1959), the desire of individuals to feel effective in the world.

Paper-based group model building provides the opportunity for interaction with a tangible representation (Black and Andersen 2012), and may support a touch-bias. The process used in the case study involved participants interacting with post-it notes (identification of variables), behaviour over time graphs and causal loop diagrams. Their interaction was extensive, tactile, and involved manipulating objects as well as moving around the room. Individuals experience a greater sense of ownership and positive affect through physical touch and physical manipulation of an object (Peck and Shu 2009). The workshop process engaged visual, auditory and kinesthetic learning styles (Barbe et al. 1979) through the use of visible graphical products, group discussion, and the handling and manipulation of sticky-labels by participants. Group processes that engage multiple senses are associated with improved learning outcomes (Dunn et al. 2002; Lujan and DiCarlo 2006). This may be worth considering when choosing between manual and electronic methods.

Group model building may be well-suited to building agreement and commitment to past strategy decisions by reinforcing certain biases to counteract others. Group model building provides opportunities for endowment, empowerment, assembly completion, effort, competence, effectance and tactile interaction.

Group model building literature proposes several explanations for how group model building effects change (Rouwette and Vennix 2006; Scott 2013). These varied mechanisms aim to describe different aspects of the participatory process. Early proposals focussed on what the individual has learned (Richmond 1993; Richardson et al. 1994; Maani and Maharaj 2003; Thompson 2009). More recent research has focussed on the interactions between participants (Vennix et al. 1996; Franco 2006; Rouwette et al. 2011; Black and Andersen 2012; McCardle-Keurentjes et al. 2008; Van Nistelrooij et al. 2012; Black 2013). Cognitive biases describe a kind of intermediate theory that supports those two approaches: why participants might be predisposed to engage positively with the process and with each other.

Limitations of the Case Study

This case study involved New Zealand public servants from a single organisation; it is not clear how results from this study would translate to other organisations or to the private sector. The results are self-reported by participants, which may not be an accurate representation of what actually occurred (Rouwette 2011). Participants compare results to a hypothetical normal meeting, but a direct comparison would be preferable (Shadish et al. 2001).

The implementation success factors are reported to be predictive (Noble 1999). The case study measures these implementation success factors immediately after a short workshop. This provides some empirical basis for applying group model building to support strategy implementation. However, this evidence is less compelling in the absence of follow-up evaluation; otherwise the workshop outcomes may only be fleeting. Disruption to the original experimental design prevented the collection of this follow-up evidence.

What remains is a theoretical basis for suspecting applicability, and the promising beginnings of an empirical basis for the claim that group model building can help in a major unresolved management problem. The coincidence of reported strategy implementation success factors and group model building outcomes (the CICC framework) provides a ready-made evaluation approach for a longitudinal study in applying group model building to support strategy implementation. Group model building workshops to plan how to implement predetermined strategic objectives (assembly completion) may be a new niche area for applying system dynamics principles and techniques. Further research in this area could be transformative to group model building practice.

References

Ajzen I (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes* **50**(2): 179–211.

Akkermans H, Vennix JAM and Rouwette EAJA (1993). Participative modelling to facilitate organizational change: A case study. *Proceedings of the 1993 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

Aligica PD (2005). Scenarios and the growth of knowledge: Notes on the epistemic element in scenario building. *Technological Forecasting and Social Change* **72**(7): 815–824.

Allen MJ and Yen WM (2002). *Introduction to Measurement Theory*. Long Grove: Waveland Press.

Andersen DF, Richardson GP and Vennix JAM (1997). Group model building: Adding more science to the craft. *System Dynamics Review* **13**(2): 187–203.

Andersen DF, Vennix JAM, Richardson GP and Rouwette EAJA (2007). Group model building: Problem structuring, policy simulation and decision support. *Journal of the Operational Research Society* **58**(5): 691–694.

Argyris C (1989). Strategy implementation: An experience in learning. *Organizational Dynamics* **18**(2): 4–15.

Aronson E and Mills J (1959). The effects of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology* **59**(2): 177–181.

Bain JS (1968). *Industrial Organization*. 2nd edn. New York: Wiley.

Bandura A (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* **84**(2): 191–215.

Barbe WB, Swassing RH and Milone MN (1979). *Teaching Through Modality Strengths: Concepts and Practices*. Columbus, OH: Zaner-Bloser.

Baum JAC and Greve HR (2001). *Multiunit Organizations and Multiunit Strategy: Advances in Strategic Management*. Oxford: Elsevier.

Black LJ (2013). When visuals are boundary objects in system dynamics work. *System Dynamics Review* **29**(2): 70–86.

Black LJ and Andersen DF (2012). Using visual representations as boundary objects to resolve conflict in collaborative model-building approaches. *System Research and Behavioural Science* **29**(2): 194–208.

Bonoma TV (1984). Making your marketing strategies work. *Harvard Business Review* **62**(2): 69–76.

Bonoma TV (1986). Marketing subversives. *Harvard Business Review* **64**(6): 113–118.

Bonoma TV and Crittenden VL (1988). Managing market implementation. *Sloan Management Review* **29**(2): 7–14.

Bourgeois LJ (1980). Performance and consensus. *Strategic Management Journal* **1**(3): 227–248.

Bourgeois LJ and Brodwin DR (1984). Strategic implementation: Five approaches to an elusive phenomenon. *Strategic Management Journal* **5**(3): 241–264.

Broman G, Holmberg J and Robert KH (2000). Simplicity without reduction: Thinking upstream towards the sustainable society. *Interfaces* **30**(3): 13–25.

Carmon Z, Wertenbroch K and Zeelenberg M (2003). Option attachment: When deliberating makes choosing feel like losing. *Journal of Consumer Research* **30**(1): 15–29.

Clark RE and Sugrue BM (1988). Research on instructional media, 1978–1988. In: D Ely (ed). *Educational Media Yearbook,* 1987–1988. Denver, CO: Libraries Unlimited, pp 19–36.

Conger JA and Kanungo RN (1988). The empowerment process: Integrating theory and practice. *Academy of Management Review* **13**(3): 471–482.

Crittenden VL and Crittenden WF (2008). Building a capable organization: The eight levers of strategy implementation. *Business Horizons* **51**(4): 301–309.

Daft RL and Macintosh NB (1984). The nature and use of formal control systems for management control and strategy implementation. *Journal of Management* **10**(1): 43–66.

De Geus A (1997). *The Living Company*. London: Nicholas Brealey.

Doyle DK and Ford DN (1998). Mental models concepts for system dynamics research. *System Dynamics Review* **14**(1): 3–29.

Drazin R and Howard P (1984). Strategy implementation: A technique for organizational design. *Columbia Journal of World Business* **19**(2): 40–46.

Dunn R, Beaudry J and Klavas A (2002). Survey of research on learning styles. *California Journal of Science Education* **2**(2): 75–98.

Dwyer M and Stave K (2008). Group model building wins: The results of a comparative analysis. *Proceedings of the 2012 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

Fidler LA and Johnson JD (1984). Communication and innovation implementation. *Academy of Management Review* **9**(4): 704–711.

Floyd SW and Wooldridge W (1982). Managing strategic consensus: The foundation of effective implementation. *Academy of Management Executive* **6**(4): 27–39.

Franco LA (2006). Forms of conversation and problem structuring methods: A conceptual development. *Journal of the Operational Research Society* **57**(7): 813–821.

Franco LA and Rouwette EAJA (2011). Decision development in facilitated modelling workshops. *European Journal of Operational Research* **212**(1): 164–178.

Franke N and Piller F (2004). Value creation by toolkits for user innovation and design: The case of the watch market. *Journal of Product Innovation Management* **21**(6): 401–415.

Gray DH (1986). Uses and misuses of strategic planning. *Harvard Business Review* **64**(1): 89–97.

Gupta AK (1987). SBU strategies, corporate-SBU relations, and SBU effectiveness in strategy implementation. *Academy of Management Journal* **30**(3): 477–500.

Guth WD and Macmillan IC (1986). Strategy implementation versus middle management self-interest. *Strategic Management Journal* **7**(4): 313–337.

Hambrick DC and Cannella AA (1989). Strategy implementation as substance and selling. *Academy of Management Executive* **3**(4): 278–285.

Heron J (1999). *The Complete Facilitator's Handbook*. London: Kogan-Page.

Hochberg Y and Tamhane AC (1987). *Multiple Comparison Procedures*. New York: Wiley.

Houchin K and MacLean D (2005). Complexity theory and strategic change: An empirically informed critique. *British Journal of Management* **16**(2): 149–166.

Hrebiniak L and Joyce WF (1984). *Implementing Strategy*. New York: Palgrave Macmillan.

Hrebiniak L and Snow CC (1982). Top management agreement and organizational performance. *Human Relations* **35**(12): 1139–1158.

Huy QN (2011). How middle managers' group-focus emotions and social identities influence strategy implementation. *Strategic Management Journal* **32**(13): 1387–1410.

Huz S (1999). *Alignment from group model building for systems thinking: measurement and evaluation from a public policy setting*. PhD thesis, State University New York, New York.

Huz S, Andersen DF, Richardson GP and Boothroyd R (1997). A framework for evaluating systems thinking interventions; an experimental approach to mental health system change. *System Dynamics Review* **13**(2): 149–169.

Jaworski BJ and MacInnis DJ (1989). Marketing jobs and management controls: Toward a framework. *Journal of Marketing Research* **26**(4): 406–419.

Jaworski BJ, Stathakopoulos V and Krishnan HS (1993). Control combinations in marketing: Conceptual framework and empirical evidence. *Journal of Marketing* **57**(1): 57–69.

Johnson LL and Frohman AL (1989). Identifying and closing the gap in the middle of organizations. *Academy of Management Executive* **3**(2): 107–114.

Judson AS (1991). Invest in a high-yield strategic plan. *Journal of Business Strategy* **12**(4): 34–39.

Kahneman D, Knetsch JL and Thaler R (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy* **98**(6): 1325–1348.

Kaplan RS and Norton DP (1996). *The Balanced Scorecard*. Cambridge: Harvard Business School Press.

Kaplan RS and Norton DP (2001). *The Strategy-focused Organization—How Balanced Scorecard Companies Thrive in the New Business Environment*. Cambridge: Harvard Business School Press.

Kaplan RS and Norton DP (2004). *Strategy Maps.* Cambridge: Harvard Business School Press.

Kiechel W (1982). Corporate strategists under fire. *Fortune* **106**(13): 34–39.

Kiechel W (1984). Sniping at strategic planning. *Planning Review* **12**(5): 8–11.

Kim WC and Mauborgne R (2005). *Blue Ocean Strategy.* Cambridge: Harvard Business School Press.

Kline P (1999). *The Handbook of Psychological Testing.* London: Routledge.

Langfield-Smith K (1997). Management control systems and strategy: A critical review. *Accounting, Organizations and Society* **22**(2): 207–232.

Lujan HL and DiCarlo SE (2006). First-year medical students prefer multiple learning styles. *Advances in Physiology Education* **30**(1): 13–16.

Lyneis JM (1999). System dynamics for business strategy: A phased approach. *System Dynamics Review* **15**(1): 37–70.

Maani KE and Cavana RY (2007). *Systems Thinking, System Dynamics—Managing Change and Complexity.* 2nd edn. New Zealand: Pearson Education.

Maani KE and Maharaj V (2003). Links between systems thinking and complex decision making. *System Dynamics Review* **20**(1): 21–48.

McCardle-Keurentjes MH, Rouwette EAJA and Vennix JAM (2008). Effectiveness of group model building in discovering hidden profiles in strategic decision-making. *Proceedings of the 2008 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

McCardle-Keurentjes MH, Rouwette EAJA, Vennix JAM and Jacobs E (2009). Is group model building worthwhile? Considering the effectiveness of GMB. *Proceedings of the 2009 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

McCartt A and Rohrbaugh J (1989). Evaluating group decision support effectiveness: A performance study of decision conferencing. *Decision Support Systems* **5**(2): 243–253.

McGraw AP, Tetlock P and Kristel O (2003). The limits of fungibility: Relational schemata and the value of things. *Journal of Consumer Research* **30**(2): 219-229.

Miles RE and Snow CC (1978). *Organizational Strategy, Structure, and Process.* New York: McGraw-Hill.

Mingers J and White L (2010). A review of the recent contribution of systems thinking to operational research and management science. *European Journal of Operational Research* **207**(3): 1147–1161.

Mochon D, Norton MI and Ariely D (2012). Bolstering and restoring feelings of competence via the IKEA effect. *International Journal of Research in Marketing* **29**(4): 363–369.

Nisbett R and Wilson T (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* **84**(3): 231–259.

Noble CH (1999). The eclectic roots of strategy implementation research. *Journal of Business Research* **45**(2): 119–134.

Norton M, Mochon D and Ariely D (2012). The 'IKEA effect': When labor leads to love. *Journal of Consumer Psychology* **22**(3): 453–460.

Nutt PC (1983). Implementation approaches for project planning. *Academy of Management Review* **8**(4): 600–611.

Nutt PC (1986). Tactics of implementation. *Academy of Management Journal* **29**(2): 230–261.

Nutt PC (1990). Strategic decisions made by top executives and middle managers with data and process dominant styles. *Journal of Management Studies* **27**(2): 173–194.

Nutt PC (1999). Surprising but true: Half the decisions in organizations fail. *Academy of Management Executive* **13**(4): 75–90.

Nutt PC (2002). *Why Decisions Fail. Avoiding the Blunders and Traps that Lead to Debacles*. San Francisco: Berrett-Koehler.

Patterson K, Grenny J, Maxfield D, McMillan R and Switzler A (2008). *Influencer: The Power to Change Anything.* New York: McGraw-Hill.

Peck J and Shu SB (2009). The effect of mere touch on perceived ownership. *Journal of Consumer Research* **36**(3): 434–447.

Pidd M (2004). Contemporary OR/MS in strategy development and policy-making: Some reflections. *Journal of the Operational Research Society* **55**(8): 791–800.

Porter ME (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* New York: The Free Press.

Raps A (2005). Strategy implementation—An insurmountable obstacle? *Handbook of Business Strategy* **6**(1): 141–146.

Redding JC and Catalanello RC (1994). *Strategic Readiness.* San Francisco: Jossey-Bass.

Rees F (2005). *The Facilitator Excellence Handbook*. 2nd edn. San Francisco: Pfeiffer.

Richardson GP (1986). Problems with causal-loop diagrams. *System Dynamics Review* **2**(2): 158–170.

Richardson GP (1997). Problems in causal loop diagrams revisited. *System Dynamics Review* **13**(3): 247–252.

Richardson GP and Pugh AL (1981). *Introduction to System Dynamics Modeling with DYNAMO.* Cambridge: MIT Press.

Richardson GP, Andersen DF, Maxwell TA and Stewart TR (1994). Foundations of mental model research. *Proceedings of the 1994 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

Richmond B (1993). Systems thinking: Critical thinking skills for the 1990s and beyond. *System Dynamics Review* **9**(2): 113–133.

Robertson TS and Gatignon H (1986). Competitive effects on technology diffusion. *Journal of Marketing* **50**(3): 1–12.

Rouwette EAJA (2003). *Group Model Building as Mutual Persuasion.* Nijmegen: Wolf Legal Publishers.

Rouwette EAJA (2011). Facilitated modelling in strategy development: Measuring the impact on communication, consensus and commitment. *Journal of the Operational Research Society* **62**(5): 879–887.

Rouwette EAJA, Korzilius H, Vennix JAM and Jacobs E (2011). Modeling as persuasion: The impact of group model building on attitudes and behaviour. *System Dynamics Review* **27**(1): 1–21.

Rouwette EAJA and Vennix JAM (2006). System dynamics and organizational interventions. *Systems Research and Behavioral Science* **23**(4): 451–466.

Rouwette EAJA, Vennix JAM and Van Mullekom T (2002). Group model building effectiveness: A review of assessment studies. *System Dynamics Review* **18**(1): 5–45.

Sandy W (1991). Avoid the breakdowns between planning and implementation. *Journal of Business Strategy* **12**(5): 30–33.

Schoemaker PJH (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal* **14**(3): 193–213.

Schwartz RM (1994). *The Skilled Facilitator: Practical Wisdom for Developing Effective Groups.* San Francisco: Jossey-Bass.

Schweiger DM, Sandberg WR and Rechner PL (1989). Experiential effects of dialectical inquiry, devil's advocacy, and consensus approaches to strategic decision making. *Academy of Management Journal* **32**(4): 745–772.

Scott RJ (2013). Mechanisms for understanding mental model change in group model building. *Proceedings of the 2013 International Society for the Systems Sciences Conference*, International Society for the System Sciences, York.

Scott RJ, Cavana RY and Cameron D (2013). Evaluating immediate and long-term impacts of qualitative group model building workshops on participants' mental models. *System Dynamics Review* **29**(4): 216–236.

Senge P (1990). *The Fifth Discipline: The Art and Practice of Learning.* New York: Doubleday.

Shadish WR, Cook TD and Campbell DT (2001). *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* 2nd edn. Wadsworth: Cengage Learning.

Shaffer JP (1995). Multiple hypothesis testing. *Annual Review of Psychology* **46**(1): 561–584.

Sila I (2007). Examining the effects of contextual factors on TQM and performance through the lens of organisational theories: An empirical study. *Journal of Operations Management* **25**(1): 83–109.

Sirkin HL, Keenan P and Jackson A (2005). The hard side of change management. *Harvard Business Review* **83**(10): 109–118.

Skivington JE and Daft RL (1991). A study of organizational 'framework' and 'process' modalities for the implementation of business-level strategic decisions. *Journal of Management Studies* **28**(1): 45–68.

Snabe B and Größler A (2006). System dynamics modelling for strategy implementation—case study and issues. *Systems Research and Behavioral Science* **23**(4): 467–481.

Stephens MA (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* **69**(347): 730–737.

Sterman JD (2000). *Business Dynamics—Systems Thinking and Modeling for a Complex World.* Boston: Irwin/McGraw-Hill.

Tang F (2011). Knowledge transfer in intra-organization networks. *Systems Research and Behavioral Science* **28**(3): 270–282.

Thompson JP (2009). *How and under what conditions client learn in system dynamics consulting engagements.* PhD thesis, Strathclyde Business School, Scotland.

Thompson DV and Norton MI (2011). The social utility of feature creep. *Journal of Marketing Research* **48**(3): 555–565.

Tropman JE (2003). *Making Meetings Work: Achieving High Quality Group Decisions*. 2nd edn. Thousand Oaks: Sage.

Van Nistelrooij LPJ, Rouwette EAJA, Vestijnen I and Vennix JAM (2012). Power-levelling as an effect of group model building. *Proceedings of the 2012 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

Vennix JAM and Rouwette EAJA (2000). Group model building. What does the client think of it now? *Proceedings of 2000 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

Vennix JAM, Akkermans HA and Rouwette EAJA (1996). Group model-building to facilitate organizational change: An exploratory study. *System Dynamics Review* **12**(1): 39–58.

Vennix JAM, Scheper W and Willems R (1993). Group model building. What does the client think of it? *Proceedings of the 1993 International System Dynamics Conference*, System Dynamics Society, Chestnut Hill.

Walsh A, Hughes H and Maddox DP (2002). Total quality management continuous improvement: Is the philosophy a reality? *Journal of European Industrial Training* **26**(6): 299–307.

White RW (1959). Motivation reconsidered: The concept of competence. *Psychological Review* **66**(5): 297–333.

Whitney JC and Smith RA (1983). Effects of group cohesiveness on attitude polarization and the acquisition of knowledge in a strategic planning context. *Journal of Marketing Research* **20**(2): 167–176.

Whittington R (1996). Strategy as practice. *Long Range Planning* **29**(5): 731–735.

Wooldridge B and Floyd SW (1989). Research notes and communications: Strategic process effects on consensus. *Strategic Management Journal* **10**(3): 295–302.

Wooldridge B and Floyd SW (1990). The strategy process, middle management involvement, and organizational performance. *Strategic Management Journal* **11**(3): 231–241.

Workman JP (1993). Marketing's limited role in new product development in one computer systems firm. *Journal of Marketing Research* **30**(4): 405–421.

Yang L, Sun G and Eppler MJ (2008). *Making strategy work: A literature review on the factors influencing strategy implementation*. ICA Working Paper 2. University of Lugano, Switzerland.

Zairi M (1996). *Effective Benchmarking—Learning from the Best*. London: Chapman & Hall.

Part II

Methodological Developments

An Overview of Strategy and Tactics in System Dynamics Optimization

B. Dangerfield and Carole Roberts

Introduction

From quite early days in the development of System Dynamics, practitioners have been interested not only in producing an adequate model, but also in improving its performance in some respects. In short they have become interested in the concept of optimization of their model.

As an illustration, Hall [1] in his admirable early study showed how a system dynamics model could be created that reflected the business fortunes of the US magazine, the old *Saturday Evening Post*. In his discussion on policies he remarked on a (then recent) study by Nelson and Krisbergh [2] that involved interfacing a sophisticated razor search procedure with Forrester's Urban Dynamics model [3] for the purpose of identifying a better policy mix according to a weighted objective function. Hall stated

Journal of the Operational Research Society (1996) **47**(3)

B. Dangerfield (✉) • C. Roberts
Department of Management, University of Bristol, Bristol, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_6

that: ‘This seems to offer promise as a tool for optimal policy making in dynamic interactive multi-objective feedback systems …’. Since that time optimization in system dynamics has progressed. Not least, this has been aided by the fast-growing power of modern PCs. This paper attempts to examine what developments there have been, chart their performance and look to the future. A comprehensive list of references on system dynamics optimization is gathered and we review the *modus operandi* when performing optimization using search routines. This material is relevant regardless of the particular software implementation being employed.

The Nature of System Dynamics Optimization

Optimization covers two distinct lines of model development. First, there is the problem of calibrating the model by fitting suitable model variables to past time series data. Secondly, there is the issue of gauging a policy’s performance relative to a criterion.

Optimization to Fit Data

Whilst system dynamics is primarily concerned with comparative simulation runs, in order to understand better the behaviour of the system and to evaluate the effect of various possible policy interventions, a reasonable fit of the model to past data is often desirable as a means of reinforcing confidence in the model on the part of client management. Historical data is not always available, for instance when designing a new system, but where it is, then there are now no computational problems in determining the set of parameters that offer the best fit. Furthermore, if one model containing a certain structural feature yields a better historical fit to the data than another, which does not include that feature, but which is the same in all other respects, then some support is clearly offered towards the realism of that structure. Additional supporting evidence as to the existence of that feature might then be unearthed by conducting a direct survey or by statistical analysis of independent cross-sectional data.

Policy Optimization to Improve System Performance

In a business model, for instance, one may be interested in maximizing profit in the face of unknown demand. In this context, optimization determines how parameters guiding the recruitment or lay-off of the workforce, together with parameters associated with the control of inventory, should be set in order that the cost side of the profit equation is minimized. If, additionally, sales are affected by the length of the delivery delay, then policies adopted by the production department are impinging in the external market and the optimization problem is a lot more difficult than one where the sales function is entirely exogenous. Whatever situation prevails, the problem is to ensure the best possible performance in the face of an uncontrollable force, using whatever controllable variables are to hand.

Optimization Techniques: Optimal Control

It is acknowledged that system dynamics has its roots in control engineering and that Forrester [4] wished to devise an approach to analysing management systems using methods that were more amenable than those routinely employed by scientists and engineers in designing physical systems. He was influenced by Tustin [5] who had applied the ideas of the control engineer to economic systems.

Control engineering has developed to include the study of optimal control and the concepts involved have been applied to management as well as physical systems (see for instance Bensoussan et al. [6], Kamien and Schwartz [7] and Sethi and Thompson [8]). It is perhaps not surprising then to learn that research has been published where the methods of optimal control theory have either been applied to system dynamics models or considered in their context. Examples of this work are reported by Burns and Malone [9], Sharp [10], Mohapatra [11] and Kivijärvi and Tuominen [12]. This latter reference is discussed further below. Thus far, the work seems to be restricted only to analysing textbook models that are dominated by negative feedback.

In addition, another class of methods, called modal control theory—where the performance criterion is the degree of stability of the system—has been explored by Mohapatra [13], Mohapatra and Sharma [14] and by Talavage [15].

In the 1970s, Peterson [16] developed a program called GPSIE, which incorporates Kalman filtering techniques primarily as a way of estimating or indirectly measuring system states using numerical data. This worked even if the available data were noisy and/or incomplete. The tools included in GPSIE were based on methods of full-information maximum likelihood via optimal filtering and are discussed in Schweppe [17]. Subsequently, these tools have been made available in the VENSIM [18] software.

A feature of optimal control methods is the complexity of the mathematics involved in order to determine the analytical solution. Consequently, very few managers in industry would be equipped with the necessary skills, and, should they confront a problem that is amenable to dynamic optimization, they would probably respond with a less than optimal (but possibly satisfactory) policy choice. Furthermore, the methods of optimal control can require certain simplifying assumptions to be made, such as linear systems, low order models and quadratic performance criteria. If model reduction is required in order to utilize the techniques, then this is a serious constraint that can undermine the realism of the model. It is vitally important for the analyst not to lose the confidence of the client, which might arise from perceived shortcomings in the tools at the analyst's disposal.

Optimization Techniques: Algorithmic Search

An alternative approach to the optimization of system dynamics models has been put forward by Keloharju [19]. This approach overcomes the objections to optimal control theory mentioned above but at the possible expense of accuracy. Keloharju's idea was to take a heuristic pattern search algorithm and graft this directly onto a standard system dynamics simulation language. In computing terminology, the model itself became a subroutine to the search algorithm, each call to that routine producing another run of the model. The result returned by the model would be the value of the objective function for the model run with a given parameter set. The search algorithm would then automatically adjust the parameter set in the light of the result obtained and another call to the subroutine would follow. The search in parameter space is controlled according to the heuristic that was first devised by Hooke and Jeeves [20] and has been used by Buffa and Taubert [21] in order to tackle the aggregate scheduling problem in production management via a direct pattern search process.

Taubert [22] has assessed the performance of the pattern search algorithm in comparison with two other routines: conjugate gradient and variable metric. He discovered that the Hooke and Jeeves heuristic was computationally superior and, quoting Wilde [23], he attributed this to its ability to find and follow ridges in *n*-dimensional space, thereby reducing the effective dimensionality of the problem. Taubert found that the optima obtained by the three methods varied by only $792 out of $290,000. What set the pattern search heuristic apart, was the speed with which it located the optimum.

However, because it is a heuristic procedure, pattern search cannot—in contrast to optimal control methods—be certain of locating the best parameter set. This risk has to be set against the benefits of flexibility and greater transparency of the approach.

Even in relatively small models, the parameter space is vast. Consider a model of 30 parameters, each of which can take on a range of ten possible discrete values (a very conservative estimate). Then the number of possible parameter combinations is 10 [30]. If a computer could evaluate one million of these per second it would take approximately 3.17×10^{16} years to work through all the combinations. The age of planet earth is approximately 4.6×10^{9} years.

History of Pattern Search in System Dynamics

The work inaugurated by Keloharju over the period 1975–1982 resulted, firstly, in the Hooke and Jeeves algorithm being grafted onto the DYNAMO [24] system dynamics program to create SDRDYN

[25] and, subsequently, onto the DYSMAP simulation software. Krallmann [26] had also reported the integration of a razor search (an enhancement of the pattern search procedure) algorithm with the DYNAMO software in 1976 but, along with the work of Nelson and Krisbergh [2] mentioned earlier, no standalone software tool appears to have emerged as a result.

At that time, mainframe computers were the vogue and the resultant hybrid program produced by Keloharju's group became known as DYSMOD (Dynamic Systems Modeller, Optimizer and Developer) [27]. It ran only on Hewlett-Packard mainframes. Subsequent work at the University of Salford in 1989 has resulted in it being ported to the PC environment. It is implemented to run on any PC equipped with a 386 (or above) chip. The software operates by pre-compiling the model into FORTRAN source code and then executing the resultant program. As a consequence, it is necessary for the PC to have a FORTRAN compiler loaded, to which the DYSMOD output is directed. Presently it is set up to operate with Salford's FTN77 compiler.

Keloharju and Wolstenholme [28, 29] have already reported the basic details of how DYSMOD operates, together with some simple examples and a substantial case study based on a problem described in a text book, and it is not proposed to repeat this material here.

Since 1989, two other system dynamics software programs have appeared, each of which features an optimization facility. These are COSMOS [30] and VENSIM [18]. The former is based directly on a hill-climbing search procedure in the same way as DYSMOD, while the latter, besides including the Kalman filtering techniques mentioned earlier, incorporates randomized search, grid search and vector search for the purposes of either parameter estimation or the optimization of system performance.

Parameter optimization by direct search is intuitively easy to follow and this can lead people to believe that little skill is required in order to use it. Whilst there is no doubt that it involves less mathematical skill than is required when working with optimal control methods, the user still requires a degree of understanding of the nature of the problem.

Tactics for Search-Based Optimization

First, the ranges assigned to the search around each parameter need careful consideration. An overall optimal result is optimal only insofar as the ranges proffered are adequate; a wider or totally different range for a parameter may lead to an improved objective function. However, a better objective function accompanied by a clearly implausible parameter value is unacceptable. This serves to underline that, usually, some prior knowledge of sensible ranges will be available and this knowledge should be utilized. Putting huge ranges on every parameter and then expecting the software to locate sensible optima is facile. A far better method is to restrict the (unknown) range of feasibility and then expand this if the search reveals that the floor or ceiling of the range is restricting exploration. The user should consider whether, for instance, a value of zero for the minimum of the range is reasonable. If this value is selected during the search and the parameter is a multiplier in a denominator, then a fatal error will ensue and the search procedure will immediately terminate.

Furthermore, if the parameter is a delay constant then it is well known that such values must be somewhat larger than the simulation time step (DT) in order to prevent mathematical instabilities in the model. Remembering that, under optimization, every parameter in the search is now a variable, should ensure that the lowest point on the range of a delay parameter is not inconsistent with the rule of thumb which states that DT should be set to, at most, half the value of the smallest first order delay parameter in the model. Some prefer the more stringent requirement that would make DT, at most, one-quarter of the value of the smallest first-order delay parameter, although this doubles the number of separate computations involved in the execution of the simulation.

A step multiplier exists, which is a parameter of the DYSMOD search algorithm itself. It controls the size of the step which determines the next value of the model parameter in the exploratory search. For instance, suppose a model parameter has an initial value of 8 and a range of 6 to 10 is put around this. If the step multiplier takes a value of 0.3 (the default) then trial values of $8 - (4 \times 0.3) = 6.8$ and $8 + (4 \times 0.3) = 9.2$ will be selected.

A check is then made to ensure that these exploratory search values are not outside the user-specified range. If a coarser search is required then a larger value should be selected, although the algorithm subsequently adjusts the value of the step multiplier automatically.

The number of iterations is not something that need tax the user these days given the power of modern PC-based CPUs. The search converges relatively rapidly. This is reflected in Taubert's [22] work previously referred to and also in the experience of the authors. Consider, for instance, Fig. 1, which shows the behaviour of an objective function over only 60 iterations. (Iteration zero is the base run prior to commencing the search process.) This was taken from an optimization for which over 500 iterations were originally requested. It was conducted by the authors with a view to fitting a model of AIDS spread in the UK to historical quarterly data. If sufficient CPU power is available, then thousands of iterations can be contemplated; this is certainly possible with moderately sized models run on 486-based CPUs. The search procedure always stores the best parameter combination found to date, so even if an excessive number of iterations causes the objective function to make an excursion

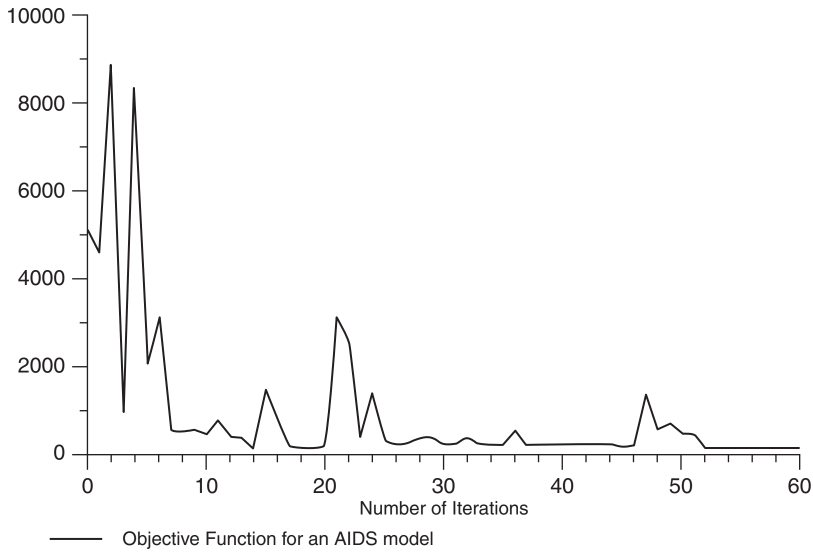


Fig. 1 Rapid convergence of the objective function during optimization

from the point to which it had converged, no crucial results are lost. What is particularly important though is to avoid too few iterations and it should be appreciated that, if optimizing a table function, each point in that function for which a *Y*-value is given is counted as an independent parameter.

Allied to the above is the statistical issue of degrees of freedom. This is a problem only when fitting a model to historical data. If a data set contains n points then a model containing, at most, $n-1$ parameters can be fitted to that data but, in fact, the number of independent parameters in the model should be a lot less than $n-1$ to allow for a reasonable number of degrees of freedom for the error term, thus strengthening significance tests on the parameters. Sometimes it is possible to combine parameters, especially in multiplicative relationships where confounding prevents their separate identification.

The nature of the variable chosen for the historical fit is important. From the set of variables that comprise the model, it is likely that at least one will stand out as being of central importance and a fit to this variable will be carried out. Should another variable also be a candidate for fitting to its simulated counterpart, then this fit should be carried out, preferably independently. A comparison of the relevant parameter set estimated under the first and subsequent fits can then be made. Confidence in this model is considerably enhanced if the two parameter vectors are not too dissimilar.

If it is felt that the search has become lodged in a local optimum, it is often useful to transplant the current optimized parameter values back into the model, redefine the ranges and commence optimization again.

When fitting, it is not appropriate to employ cumulative values of the variables concerned. Visually, a cumulative fit will almost always look good but statistically the approach is flawed since each data point is not independent of those that have come before. Strictly independent realizations of both the simulated and observed variables are a pre-requisite for best statistical practice.

A number of issues have to be resolved when fitting model variables to reported data should the data be reported on a coarser time interval than that under which the model is running. For instance, in modelling the epidemiology of AIDS, the reported incidence is most conveniently

```
A  REPDAT. K=TABHL (REPDATT, TIMM.K, 1982.25-DT, 1991.5-DT ,0.25)

A  TRIGGER. K=0+PULSE(1, STIME+0.25–DT ,0.25)
A  DEVC.K=OCLIP (0,REPDAT.K–EXPDAT.K, TIME.K, 1991.5,
X  1982.25–DT–DT, TIME.K)
A  CHI.K=RATIO(DEVC. K*DEVC.K, EXPDAT.K)
A  LOWEXP.K=CLIP(0,CHI.K, EXPDAT . K,1)
L  SUMCHISQ.K=SUMCHISQ.J+DT/DT*(CHI.J-LOWEXP.J)*TRIGGER.J
N  SUMCHISQ=0
```

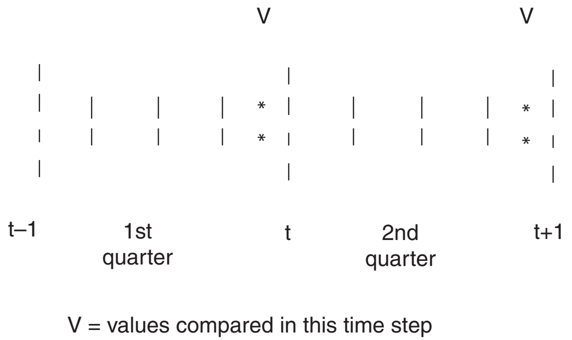


Fig. 2 Comparison of reported quarterly incidence data with model generated data

handled on a quarterly basis in line with the frequency of reported data, yet the DT in the model concerned might be only 1/16 years. If say, 100 new cases of AIDS are reported in one quarter then there are four time slices to choose to allocate these cases. The solution we employ is to assign each data value to the final DT in each quarter (see Fig. 2) and compare it with the instantaneous quarterly incidence computed by the model for that DT [31].

The equations below show the detail of the approach. In addition to the table function for the reported statistics, the equations required to compute the metric chosen for the fitting process are also listed.

A χ^2 statistic is computed in the objective function since a fit equivalent to the principle of maximum likelihood is targeted. The choice of the appropriate objective function is discussed in more detail in the next section.

Strictly speaking, the model's analogue to the quarterly reported incidence of AIDS is the difference between cumulative AIDS cases at quarter t and quarter $t-1$, for all t. However, with system dynamics software, a level is computed at the beginning of each DT and, therefore, even though it is easy to hold the 'old' value of the level at time $t-1$ through the next time step(s), the new value at the level at time t is not known until the start of the interval $(t, t+1)$. Thus, to carry out the differencing between reported and expected cases of AIDS would require that the reported statistic be pushed forward one reporting interval in order to marry up with the appropriate expected value derived as suggested. This is cumbersome and, furthermore, means that the first reporting period in the simulation is devoid of any 'observed-expected' value whatsoever. The inadequacies of this approach led to a preference for the one based on incidence data, explained above and depicted in Fig. 2.

The TRIGGER variable is employed to ensure that the model accumulates in SUMCHISQ only the CHI values from the final instantaneous expected incidences in each quarter. This mechanism is similar to the PICK macro developed by Sterman [32] to handle the same problem. The CHI values are obtained by computing DEVC (DEViations of Cases) squaring this value and dividing by the expected number of cases. DEVC computes the deviations between reported and expected values and restricts this to only those values arising at or between the limits of the data employed, namely 1982 (Q1) and 1991 (Q2). The OCLIP function used in the equation for DEVC effects the necessary restriction. The '-DT-DT' term is required because the selected time step for the comparison is one DT before the end of the quarter and the operational test between the fifth and sixth (and third and fourth) arguments of an OCLIP function is *greater than or equal to*. It is required that the differencing of reported and expected cases should not be carried out until time reaches one DT before the first quarter of 1982. The ability to restrict fitting to a specific data range, despite the model covering a longer period either side of the data, is a rather useful feature easily implemented in a system dynamics optimization.

Computationally, there can be problems with adoption of χ^2 since the denominator in the formula (the expected number of cases) may be zero, or close to zero, at the beginning of the epidemic. The RATIO function is used to circumvent this problem. In addition, a variable LOWEXP is employed to deal with the consequences of any low expected values of cases. Low is defined as less than 1.0. When this happens, LOWEXP exactly offsets the computed value of CHI and hence no change occurs when χ^2 is being cumulated. The TRIGGER variable ensures this takes place only at every quarterly interval where real data exists and the 'DT/DT' term permits the cumulation of the exact χ^2 values and not (l/DT) th of them.

The direct search method, as implemented in DYSMOD, commends itself to the modeller because it allows table functions to be optimized. These are arbitrary *X*–*Y* relationships commonly employed in system dynamics models and they equate to an array of parameters rather than one single value. Keloharju and Wolstenholme [28] show how a table function can be optimized in a simple diffusion model. What is particularly appealing about a table function, however, is the fact that the x-variable can be replaced by TIME, indeed this is the normal procedure for reading a historical time series into a model. Because of the ability to adapt a table function in this way, it means that parameters can be optimized which are themselves changing with time.

Again, by reference to our work in modelling the epidemiology of AIDS in the male homosexual population, we can illustrate how this powerful feature has been used to estimate the parameter for the partner change frequency [33]. Arising from health education campaigns and general knowledge about how HIV has been spread amongst male homosexuals, a significant behaviour change has been charted since the mid-1980s. Thus, it is certainly inappropriate to include a constant value for the partner change frequency in any AIDS model purporting to cover the evolution of the epidemic since its inception in the late 1970s. Optimization of this particular model is discussed further in the section below on examples.

Calibration Issues

Defining the Objective Function

Various metrics are available in order to conduct the fitting process for the purpose of parameter estimation. However, it is generally agreed by statisticians that maximum likelihood estimates are preferred and thus the selection of the most appropriate metric is tied in with the form of the underlying statistical model.

For instance, Nelder and Wedderburn [34] suggested a goodness-of-fit criterion based on the maximized value of the log-likelihood, known as the deviance (*G* [2]). For data derived from a Gaussian process the deviance is related to the familiar sums of squares met in regression and the analysis of variance. For count data derived from a Poisson process

$$G^2 = 2\sum_i (O_i)\,\text{In}\left(\frac{O_i}{E_i}\right),$$

where O = observed and E = expected cell counts, it is much like the Pearson chi-squared statistic.

Deviance obviously cannot be used to drive the fitting process in an optimization since individual components of the summation can be either negative or positive, unlike the strictly positive values when chi-squared is the metric. However, both deviance and chi-squared can be used as comparators between two model fits, one with n and another with $(n - k)$ parameters. The change in deviance (or chi-squared) can be compared with χ^2 with k degrees of freedom.

In a model of AIDS spread, chi-squared would be an appropriate objective function. For a population of size N with a predicted probability p of an individual contracting AIDS during a particular quarter, the number of new cases follows a binomial distribution with parameters (N, p). However, for large N and small p the binomial is approximated by a Poisson. Hence, the generation of new AIDS cases is assumed to be a Poisson process and chi-squared is thus indicated as an appropriate metric for maximum likelihood estimation.

Having obtained a best-fit parameter vector it is possible to put confidence intervals around the estimates. Within the field of statistical modelling, maximum likelihood methods are routinely adopted both for parameter estimation and the subsequent determination of confidence intervals. Those attempting the best fit of a system dynamics model variable to data can still provide confidence intervals on parameters if they know the underlying statistical processes going on in their model and they have chosen a statistically respectable metric for the fitting process. As suggested above, in modelling the spread of AIDS, adoption of χ^2 as the fitting metric can be shown to yield a fit approximately equivalent to the corresponding maximum likelihood estimates of the parameters. The methods of profile likelihood can then be employed for the confidence intervals. By perturbing each parameter estimate in-turn by no more than 1% and then re-running the model to establish the subsequent effect on the objective function, the bounds on the confidence interval can be established by estimating the standard error as follows:

$$\sigma = \frac{|\theta_1 - \theta_0|}{\sqrt{\Delta\chi^2}},$$

where θ_i are the original and perturbed values of the parameter.

In the situation where optimization is being undertaken on a criterion or performance variable, for example cost or revenue, the objective function can take on a range of possible forms. A common feature, however, is to cumulate an appropriate variable in a level and this is then evaluated at the conclusion of the run. Constrained optimization can also be carried out by invoking a penalty, for example a very large number in a minimization, which is added to the objective function when a variable deviates excessively from its desired value during the search process. In addition, constrained optimization can be adopted in order to force a terminal constraint, for example the avoidance of too low an inventory at the end of the planning horizon in an examination of production planning policies.

Calibrating the Direct Search Approach

A neat comparative study of the direct search and optimal control approaches was conducted by Kivijärvi and Tuominen [12]. They presented a typical management problem and then proceeded to optimize using each method in turn. The problem was relatively simple: given a sales profile defined from time $t = 0$, how should production be planned so as to minimize a quadratic cost function, which included terms equating to production relative to desired production and inventory holdings relative to desired? The inventory was determined solely by the interplay between production and sales: in other words it was a finished goods inventory. The heuristic search produced a very close approximation to that obtained by the optimal control approach. The cost function under the latter method was \$5.02 million whilst the heuristic search approach yielded a figure of \$5.08 million. The exact numerical difference was very small, just over 1%.

We also devised a calibration test on the algorithm. A straight line of the form $y = a + bt$ was chosen with an arbitrary intercept and slope. By using a normal random number generator, a scatter of 41 points was put around the line and these points were then used as input data to the heuristic search, which was charged with ascertaining the two parameters of the 'model', namely the intercept and slope, using the sum of squared deviations as a metric. The original model and the best fitting parameters as determined by the heuristic search are given below. Also shown are the best-fit parameters from a simple linear regression on the synthetic data:

$$
\begin{aligned}
y &= 105.70 + 10.30t \quad \text{model,}\\
y &= 109.48 + 10.40t \quad \text{heuristic search,}\\
y &= 109.51 + 10.40t \quad \text{OLS method.}
\end{aligned}
$$

The ranges for the search were set at (20, 150) for the intercept and (0, 30) for the slope. The ordinary least squares method produces parameter estimates that give the true minimum sum of squared deviations. These estimates are also the maximum likelihood estimates in this case. As can be seen, the parameters estimated by heuristic search were almost identical to these OLS estimates and indeed the reported sum of squared deviations for both fits was 44051.

Optimization by Direct Search: Examples

The following section contains a summary of a number of studies where the DYSMOD pattern search optimization has been employed; in two cases we have repeated experiments already published. However, they serve to illustrate the various ways that the software can be used, such as curve fitting, time-varying parameterization, performance improvement and straight data fitting. Other studies, mentioned for comprehensiveness but unreported here for reasons of brevity, include the early work of Keloharju's group in, firstly, using SDRDYN to optimize a simple retail inventory model [25], secondly to improve on the marketing strategies examined by Kotler [35] as part of a duopolistic new product marketing exercise [36] and, finally, to employ the same software in the optimization of a hypothetical pollution model [37]. Also excluded is Coyle's [38] analysis of an unstable generic inventory model, his 1992 study [39] concerning the optimization of defence expenditure in an aggregated model of land, sea and air warfare (which employed the COSMOS software) and Wolstenholme and Al-Alusi's [40] case-study of heuristic optimization in defence analysis, which also appeared in Reference Wolstenholme [41]. Additionally unreported is a rather complex optimization that we carried out to estimate, from censored data, the form and parameters of the AIDS incubation time distribution [42].

Recent additions to the literature on system dynamics optimization are the two chapters in Coyle [43], which include a general discussion of the nature of optimization in system dynamics modelling, with particular reference to system performance, and two worked examples. One of the examples includes a constrained optimization as referred to in the section on objective functions above.

Apart from the defence application by Wolstenholme and Al-Alusi [40], Coyle's model of land, sea and air warfare together with the two strands of AIDS modelling work undertaken by the present authors, we are not aware of any published studies of system dynamics optimization which go beyond applications to text-book or hypothetical models.

Mathematical Curve Fitting

The first example to be described is typical of the general process of mathematical curve fitting. Although this does not, of itself, involve a system dynamics model, the optimization software can be used as a tool for parameter estimation.

The data relate to the growth of the stock of cars in The Netherlands over the period 1964 to 1989. It was felt that this data could be fitted by a Gompertz model. Parameter optimization by direct search was undertaken in response to a suggested approach to the estimation of Gompertz parameters, which had been promulgated by Franses [44].

A Gompertz model involves three parameters, as follows: $x = \alpha \exp(-\beta \exp(-\gamma t))$. This can be represented using system dynamics notation as:

```
X.K=ALPHA*EXP(-BETA*EXP(-GAMMA*TIME.K))
where TIME is initialised at 1 (representing 1964)
```

The data and the corresponding fit using the χ^2 metric are shown in Fig. 3, which also depicts the best fit arising from the method employed by Franses. The optimal value for ALPHA was 5952.0 (range offered 4000–9000), for BETA 1.9056 (range 0.1, 3.0) and for GAMMA 0.1069 (range 0.005, 1.0).

The utility of the direct search approach is its generality: the structure of the model can be changed and the parameters rapidly re-estimated.

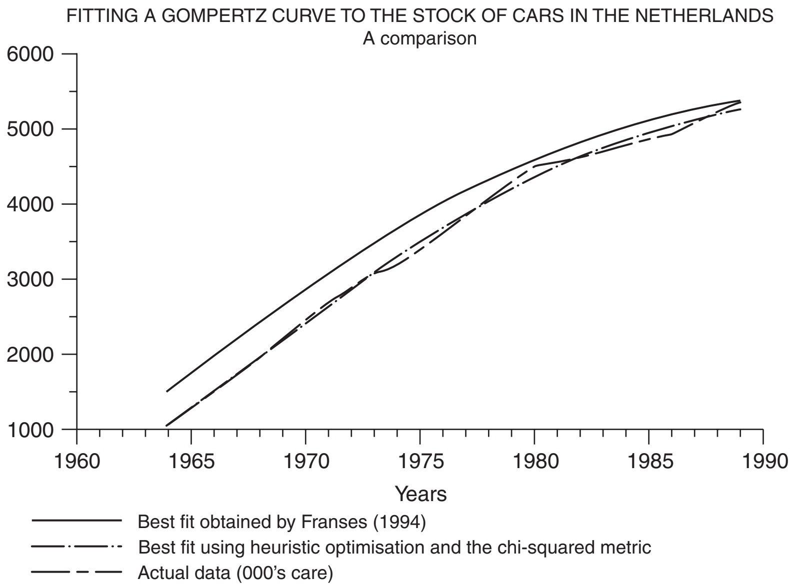


Fig. 3 Two Gompertz curves fitted to the Dutch data on car registrations

This has considerable advantages over methods of parameter estimation, which depend on the form of the hypothesized model, as was the case with the attempted fit to data on the stock of cars described above.

Comparison Between Direct Search and Optimal Control

A second example is that attributed to Kivijärvi and Tuominen [12] and referred to earlier. Their purpose was to estimate the parameters of a model of production, sales and stockholding by both direct search and optimal control methods. Given an exponentially growing sales curve, the control variable was the production rate. Its particular form was identified in the direct search experiment by estimating the parameters of a third degree polynomial.

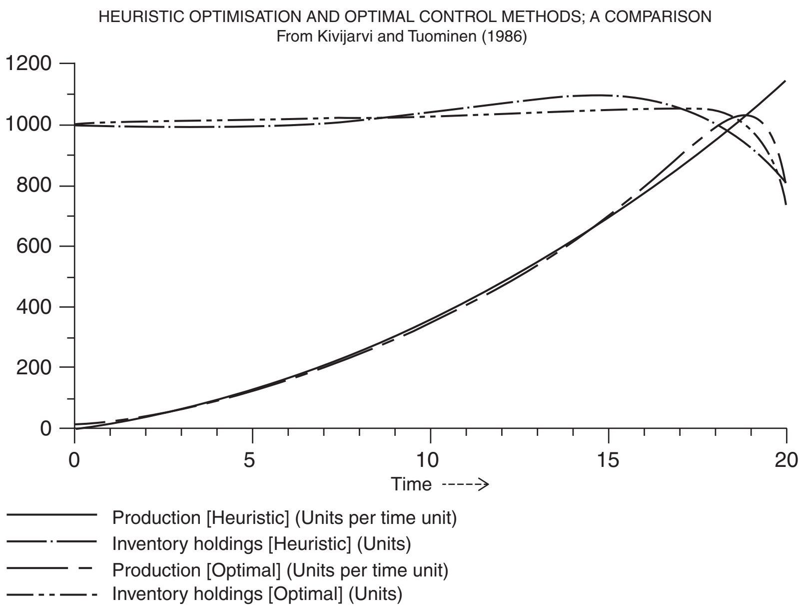


Fig. 4 Production and inventory variables obtained from two different optimization methods

Figure 4 shows the graphical result for the production rate and inventory levels over a period of 20 time units, comparing the results for the heuristic optimization and optimal control methods. As can be seen, there is very close agreement between the two. The fall-off in stockholding towards the end reflects a feature of dynamic optimization techniques applied to this sort of problem. Because the planning horizon is artificially truncated at time $t = 20$, the solution is forced in the direction of a situation where inventory is run down towards the end of the time horizon. This sort of occurrence can be avoided by incorporating a terminal constraint, mentioned above in the section on objective functions.

Another comparison, this time of the objective function (cost) arising from each of the two methods, is shown in Fig. 5. Not surprisingly this too shows extremely close agreement.

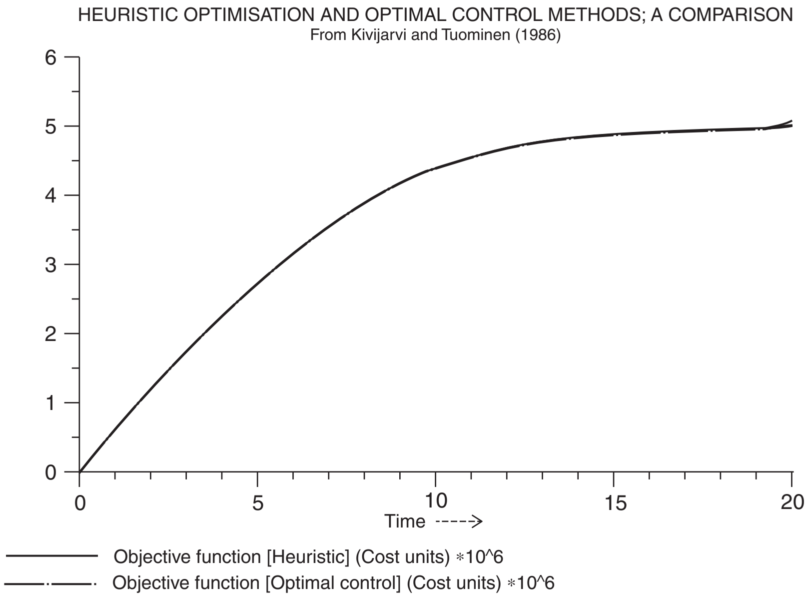


Fig. 5 Comparison of objective functions derived from two different optimization methods

Aggregate Production Planning

As an illustration of where the entire parameter set must be contained within a table function, an example is included on the aggregate production planning problem. Taubert's [22] work on this was mentioned earlier. The problem is: given a set of *n* sales values—either contracted or forecasted—determine the production rate, size of workforce and inventory holdings that will satisfy sales while minimizing costs. It should be added that the cost function in these cases is usually a fairly complex one.

It is worthy of note that system dynamics can be utilized to consider the two extreme strategies capable of being employed for tackling this problem: maintain a steady production rate and let inventory buffer any demand changes or maintain a steady inventory by having production track sales as closely as possible. At this level of consideration the model would show that it is impossible to produce an 'ideal' policy here; the issue involves a trade-off between the vested interests of the production management and planning team on the one hand and those of the finished goods warehouse management on the other [45]. Having demonstrated this conclusion using a conventional system dynamics model, utilization of an optimizing tool would further allow the operational details to be uncovered, assuming the cost function and its coefficients were acceptable to the parties involved. Morecroft [46] recognized this when he referred to a strategic recommendation needing to be 'fleshed out' so as to receive 'operational identity'.

The problem was posed originally by Holt et al. [41] in the context of aggregate production planning in a paint factory. With n = 24 months and using data provided by Taubert [22], for the system dynamics method the problem resolves into optimizing two table functions each containing 24 values—one is for the production rate and the other is for the rate of change of the workforce. The cost function to be minimized is:

$$\begin{aligned} C_t = {} & 340W_t \\ & + 64.3\left(W_t - W_{t-1}\right)^2 \\ & + 0.2\left(P_t - 5.67W_t\right)^2 + 51.2P_t - 281W_t \\ & + 0.0825\left(I_t - 320\right)^2 \end{aligned}$$

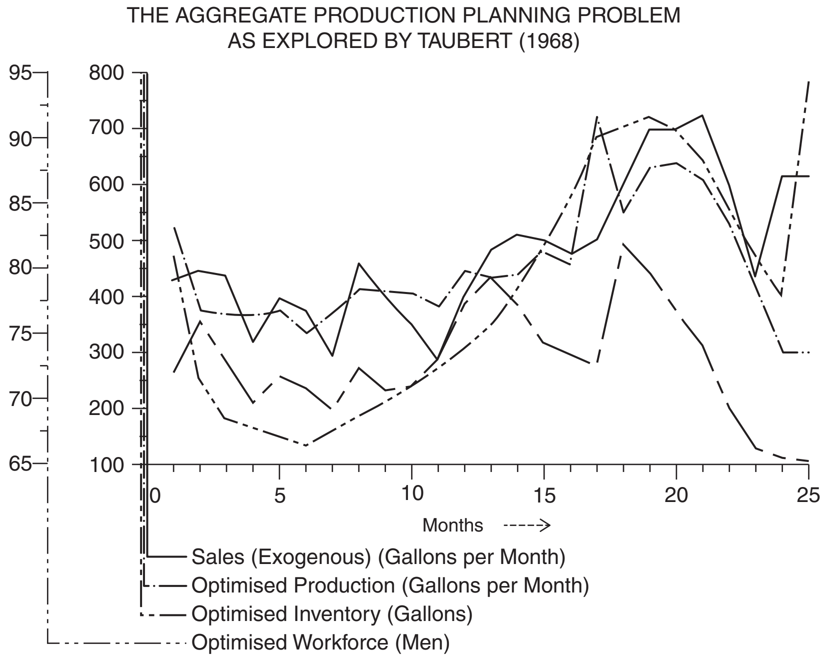


Fig. 6 Three policy variables optimized against the sales data

Figure 6 shows the optimized results for production, workforce and inventories against the exogenously imposed sales data. In general, the optimization has resulted in production (and workforce) tracking sales as closely as possible, whereas (finished goods) inventory is kept reasonably stable, except towards the end of the period. The cumulative performance of the objective function from our system dynamics model of the problem is compared with the cumulative cost data taken from Taubert [22]—see Fig. 7. Again, the degree of equivalence between the two objective functions is very high. This is not unexpected since Taubert had used the self-same heuristic search algorithm as is in DYSMOD, but this approach had not involved embracing the problem with a system dynamics model.

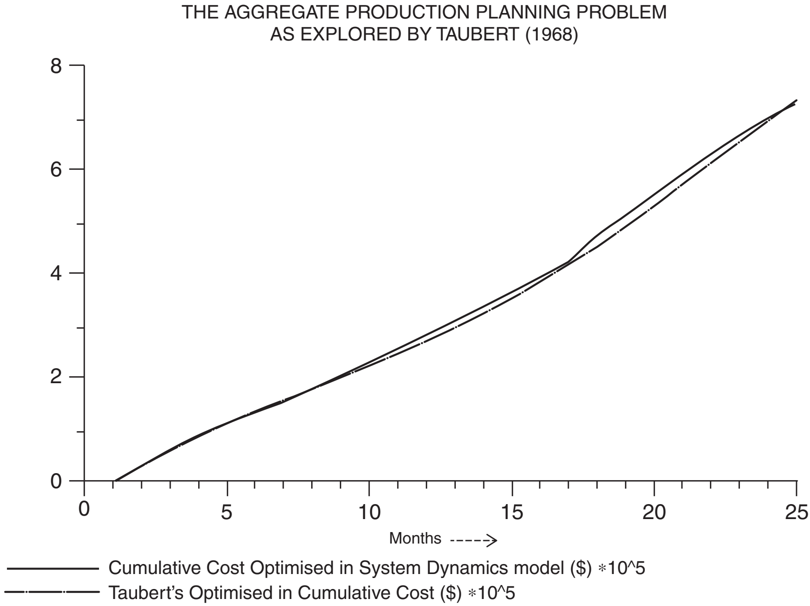


Fig. 7 Objective functions arising from optimization of the aggregate production planning model using separate modelling methods

A Problem with Spatial and Dynamic Elements

System dynamics optimization is obviously concerned with time-varying problems. Yet problems of that type and which also involve a spatial element can additionally be tackled. One such problem was postulated by Zermelo in 1931 and is recounted by Vincent and Grantham [48]. A boat situated at an arbitrary 'origin' at time $t = 0$ has to move with a velocity of unit magnitude relative to a stream of constant speed, say, $V = 2$. The problem here is to discover the constant steering angle (Θ) which will minimize the time taken to reach a target, say the disc:

$$T = \left\{ [x_1 x_2] : (x_1 - 5)^2 + (x_2 - 1)^2 \leq 1 \right\}.$$

The dynamics of this system are described by the following differential equations:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = 2 + \cos(\Theta)$$
$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = \sin(\Theta).$$

The system dynamics formulation of this problem, together with a transcript of the optimization using DYSMOD, is given in the Appendix. From the results given there, and converting radians to degrees, it can be seen that the required angle is 24.29° and the minimum time is slightly less than 1.40625 time units.

Fitting an AIDS Model to Data

A final example of optimization is devoted to our ongoing research into modelling the epidemiology of AIDS. Current work involves optimizing the parameters of a model of AIDS spread in the homosexual population by fitting it to quarterly data on reported cases of AIDS in this particular population. A reporting delay is estimated, discarding those cases reported as diagnosed in the last 18 months in order to overcome some of the problems relating to right-censored data. The lags between diagnosis and report are fitted to a negative exponential distribution and the resultant value is adopted as a parameter that is excluded from the set to be optimized. This set consists of:

1. the size of the susceptible population;
2. the mean number of different partners per unit time (refined by a factor that mimics changing behaviour);
3. the probability of passing on HIV infection, specified as three separate probabilities, which relate to distinct phases in the natural history of HIV infection over the long incubation period;
4. the duration of the second and third phases of the assumed three-stage incubation distribution.

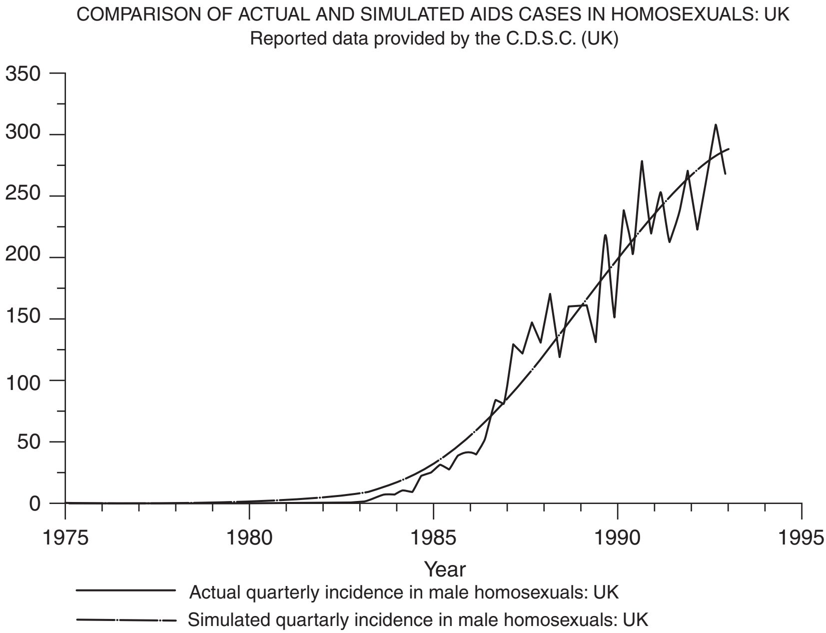


Fig. 8 Model of AIDS spread optimized to the United Kingdom data

The fit to the data for the United Kingdom is shown in Fig. 8. If the model is allowed to run on further than the limit of the data it shows that peak incidence of AIDS in this exposure group will occur sometime in 1995. Finally, a graph is shown of the optimized probabilities of passing on HIV, and which relate to the three stages of HIV infection over the long incubation period, as well as this period's total duration. The model was fitted to data from five European countries, each reporting in excess of 1000 cases of AIDS in the homosexual exposure group, see Fig. 9. Besides the UK, these were France, Germany, The Netherlands and Switzerland. The overall results are not inconsistent with the independent evidence, which supports the contention that infectivity follows a U-shaped pattern with peaks just after initial invasion of HIV and then, much later, just before the onset of clinical AIDS. Taken together with estimates of the overall duration of the incubation period averaging ten years, this offers confirmatory evidence recovered from the time series data. This particular result is illustrated in Fig. 9.

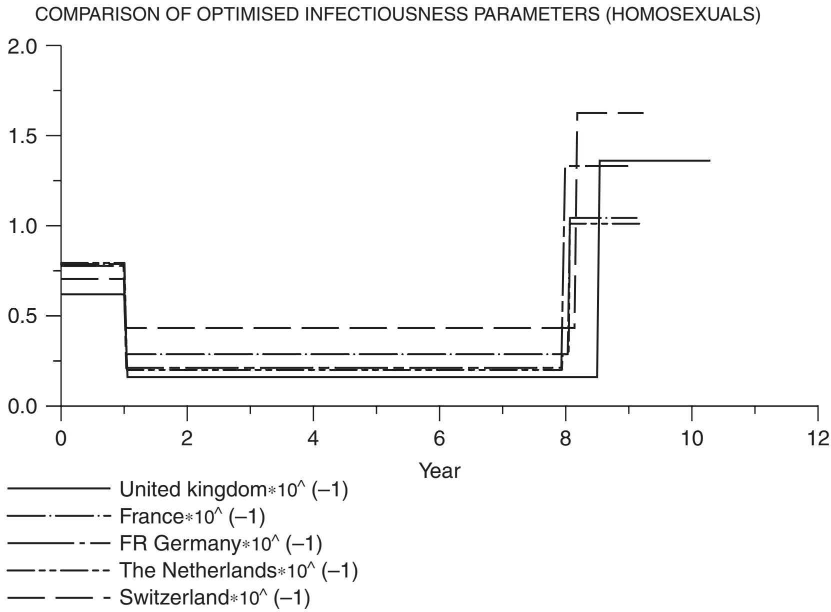


Fig. 9 Support for the hypothesis of a U-shaped infectivity profile in the natural history of HIV

Conclusions

The work reported above shows that the method of pattern search optimization of system dynamics models has considerable utility. More use of optimization methods in system dynamics is needed so that a bank of experience is built up. However, this experience should arise from applications to real-life modelling studies as opposed to text-book problems. Further calibration of the pattern search heuristic in DYSMOD is suggested and this needs to be undertaken using suitable dynamic models for which maximum likelihood parameter estimates are already known. Recent testing and use has shown this software to be very adaptable and highly accurate. It is to be hoped that additional tests succeed in confirming its very considerable promise.

Appendix: Listing of DYSMOD Run for the Navigation Problem

```
* ZERMELO'S NAVIGATION PROBLEM
NOTE
NOTE  HOLDTIM will stop dummy time (DUMTIME) incrementing when
NOTE  the target is reached. Equation for T describes location
NOTE  of the target.
NOTE
R HOLDTIM.KL=CLIP(1,0,1,T.K)
A T.K=(X1.K-5)*(X1.K-5)+(X2.K-1)*(X2.K-1)
L DUMTIME.K=DUMTIME.J+DT*(ADDTIM.JK-HOLDTIM.JK)
N DUMTIME=0
R ADDTIM.KL=1
NOTE
NOTE
NOTE  Compute position of boat using co-ordinates X1 and X2
NOTE
L X1.K=X1.J+DT*RCHX1.JK
N X1=0
L X2.K=X2.J+DT*RCHX2.JK
N X2=0
NOTE
NOTE  Differential equations describing movement of boat and
NOTE  movement stops when the target is reached
NOTE
R RCHX1.KL=CLIP(0,2+COS(THETA),1,T.K)
R RCHX2.KL=CLIP(0,SIN(THETA),1,T.K)
NOTE
NOTE  Arbitrary starting value for theta
NOTE
C THETA=0
A DEG.K=THETA*180/3.141592654
A OBJ.K=DUMTIME.K
NOTE
NOTE  Use very fine time step so as to compute time required to
NOTE  reach the target as accurately as possible (1/128 time unit).
NOTE  Also, choose LENGTH arbitrarily such that it is likely to exceed
NOTE  the required time.
NOTE
C DT=0.0078125
C LENGTH=5
C PRTPER=0.0078125
PRINT OBJ,T,DEG
RUN DETERMINE CONSTANT THETA
+
```

```
 NO SYNTAX ERRORS DETECTED
 =========================


         MODEL SYNTAX OK


 NO MODEL ERRORS DETECTED
 SORTING


 DYSMOD COMPILATION COMPLETED.
 =============================


         MODEL COMPILED INTO FORTRAN OK


 NAME OF OBJECTIVE FUNCTION ?
OBJ
 DO YOU WANT MIN OR MAX ?
MIN
 NAME OF PARAMETER  1 ?
THETA
 MIN,MAX OF THE PARAMETER THETA   ?
 MIN =          0.000E+00  MAX =           3.14
 NAME OF PARAMETER  2 ?

 TYPE 1 IF YOU WANT PARAMETER VALUES

 LENGTH OF SIMULATION ?
     5.00000
 NUMBER OF ITERATIONS ?
         500
 SIZE OF STEP MULTIPLIER ?
      0.3000
 NUMBER OF OUTPUT LINES ?
          20
 TYPE 1 IF YOU USE BASE-VECTOR

 TYPE 1 IF SIMPLIFIER

 TYPE 1 IF PLANNING HORIZON
```

```
ITER   VALUE OF OBJ

  0     5.0000
481     1.4063
482     1.4063
483     1.4063
484     1.4063
485     1.4063
486     1.4063
487     1.4063
488     1.4063
489     1.4063
490     1.4063
491     1.4063
492     1.4063
493     1.4063
494     1.4063
495     1.4063
496     1.4063
497     1.4063
498     1.4063
499     1.4063
500     1.4063


FINAL SOLUTION

THETA    =          0.4239000

NO OF OBJ      EVALUATIONS   501
INITIAL VALUE OF OBJ                 5.00000C
FINAL   VALUE OF OBJ                 1.40625C


   >
  RERUN

 * RERUN-MODE *

 >>
RUN
 RUNLENGTH?
      5.00000
```

```
                DYSMOD/486

PAGE    1

  ZERMELO'S NAVIGATION PROBLEM        DETERMINE CONSTANT THETA
=================================================================
  TIME          OBJ          T            DEG
- - - - - -  - - - - - -  - - - - - -  - - - - - -
  0.0000E+00  0.00000E+00   26.000       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
  78.125E-04  78.1250E-04   25.767       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
  15.625E-03  15.6250E-03   25.534       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
      .            .            .             .
      .            .            .             .
      .            .            .             .
      .            .            .             .
   1.391        1.3906       1.0880       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
   1.398        1.3984       1.0425       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
   1.406        1.4063      0.99804       24.288  < Target located
- - - - - -  - - - - - -  - - - - - -  - - - - - -
   1.414        1.4063      0.99804       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
   1.422        1.4063      0.99804       24.288
- - - - - -  - - - - - -  - - - - - -  - - - - - -
      .            .            .             .
      .            .            .             .
      .            .            .             .
      .            .            .             .
* SDR-MODE *
>
QUIT
```

References

1. R. I. Hall (1976) A system pathology of an organisation: the rise and fall of the old Saturday Evening Post. *Administrative Set Q.* **21**, 185–211.

2. C. W. Nelson and H. M. Krisbergh (1974) A search procedure for policy oriented simulations: applications to urban dynamics. *Mgmt Sci*. **20**, Bl164–B1174.
3. J. W. Forrester (1969) *Urban Dynamics.* MIT Press, Cambridge, MA.
4. J. W. Forrester (1961) *Industrial Dynamics.* MIT Press, Cambridge, MA.
5. A. Tustin (1953) *The Mechanism of Economic Systems.* Harvard University Press, Cambridge, MA.
6. A. Bensoussan, E. G. Hurst and B. Näslund (1974) *Management Applications of Modern Control Theory.* North Holland, Amsterdam.
7. M. I. Kamien and N. L. Schwartz (1981) *Dynamic Optimisation: The Calculus of Variations and Optimal Control in Economics and Management,* North Holland, Amsterdam.
8. S. P. Sethi and G. L. Thompson (1981) *Optimal Control Theory: Application to Management Science.* Martinuss Nijhoff, Boston.
9. J. R. Burns and D. W. Malone (1974) Computational techniques for analysis of system dynamics models of social systems. *J Socioeconomic Planning Sciences* **8**, 215-223.
10. J. A. Sharp (1978) Optimal control theory as a framework for the interpretation of system dynamics. *Dynamica* **4**(3), 138–152.
11. P. K. J. Mohapatra (1980) Structural equivalence between control systems theory and system dynamics. *Dynamica* **6**(1), 28–35.
12. H. Kivijärvi and M. Tuominen (1986) Solving economic optimal control problems with system dynamics. *Syst Dynamics Rev* **2**(2), 138–149.
13. P. K. J. Mohapatra (1979) Modal control theory and industrial dynamics: a dual approach to the design of policy decisions in a multistage production-inventory system. *Mathematics and Computers in Simulation* **21**, 141–149.
14. P. K. J. Mohapatra and S. K. Sharma (1985) Synthetic design of policy decisions in system dynamics models: a modal control theoretical approach. *Syst. Dynamics Rev.* **1**(1), 63–80.
15. J. J. Talavage (1980) Modal analysis to aid system dynamics simulation. In *System Dynamics.* (A. A. Legasto, J. W. Forrester and J. M. Lyneis, Eds.), pp. 229–240. North-Holland, Amsterdam.
16. D. W. Peterson (1980) Statistical tools for system dynamics. In *Elements of the System Dynamics Method* (J. Randers, Ed.), pp. 226–245. MIT Press, Cambridge, MA.
17. F. C. Schweppe (1973) *Uncertain Dynamic Systems.* Prentice Hall, Englewood Cliffs, NJ.
18. VENSIM (1994) Ventana Systems Inc., 149 Waverley Street, Belmont MA 02178, USA.

19. R. Keloharju (1983) *Relativity Dynamics*. Helsinki School of Economics, Helsinki.
20. R. Hooke and T. A. Jeeves (1961) 'Direct search' solution of numerical and statistical problems. *J. Ass. Computing Machinery* **8**, 212–229.
21. E. S. Buffa and W. H. Taubert (1972) *Production-Inventory Systems: Planning and Control*. R. D. Irwin, New York.
22. W. H. Taubert (1968) A search decision rule for the aggregate scheduling problem. *Mgmt Sci.* **14**, 343–359.
23. D. J. Wilde (1964) *Optimum Seeking Methods.* Prentice Hall, Englewood Cliffs, NJ.
24. A. L. Pugh III (1976) *DYNAMO* II_F *User's Manual.* MIT Press, Cambridge, MA.
25. R. Keloharju (1977) Multi-objective decision models in system dynamics. *Dynamica* **3**(1), 3–13 and **3**(2), 45–55.
26. H. Krallmann (1976) The integration of alternative modelling with DYNAMO. In *The System Dynamics Method* (J. Randers and L. K. Ervik (eds.), pp. 653–684. Oslo. Productivity Press: Portland, Oregon.
27. A. Luostarinen (1982) *DYSMOD User Manual*. Helsinki School of Economics, Helsinki.
28. R. Keloharju and E. F. Wolstenholme (1988) The basic concepts of system dynamics optimisation. *Systems Practice* **1**(1), 65–86.
29. R. Keloharju and E. F. Wolstenholme (1989) A case study in system dynamics optimization. *J. Opl Res. Soc.* **40**, 221–230.
30. COSMOS (1992) The COSMIC Holding Company, 8 Cleycourt Road, Shrivenham, Swindon, Wiltshire SN6 8BN, UK.
31. C. A. Roberts and B. C. Dangerfield (1992) Estimating the parameters of an AIDS spread model using optimisation software: results for two countries compared. In *System Dynamics 1992*, (J. A. M. Vennix, J. Faber, W. J. Scheper, and C. A. Th. Takkenberg, Eds.), pp. 605–617. System Dynamics Society, Cambridge, MA.
32. J. D. Sterman (1984) Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* **10**(2), 51–66.
33. B. C. Dangerfield and C. A. Roberts (1994) Fitting a model of the spread of AIDS to data from five European countries. In *O.R. Work in HIV/AIDS*, 2nd edn. (B. C. Dangerfield and C. A. Roberts, Eds.), pp. 7–13. Operational Research Society, Birmingham.
34. J. A. Nelder and R. W. M. Wedderburn (1972) Generalised linear models. *J. Roy. Stat. Soc. A* **135**, 370–384.

35. P. Kotler (1965) Competitive strategies for new product marketing over the life cycle. *Mgmt Sci.* **9**, B104–119.
36. T. Ansio and E. Mattila (1979) Marketing strategy formulation—pure versus mixed strategies. *J. Opl Res. Soc.* **30**, 1097–1101.
37. T. Ansio and E. Mattila (1978) Three strategies of pollution control: a heuristic optimisation model. *Dynamica* **5**, 35–54.
38. R. G. Coyle (1985) The use of optimisation methods for policy design in a system dynamics model. *Syst. Dynamics Rev.* **1**(1), 81–91.
39. R. G. Coyle (1992) The optimisation of defence expenditure. *Eur. J. Opl Res.* **56**, 304–318.
40. E. F. Wolstenholme and A. S. Al-Alusi (1987) System dynamics and heuristic optimisation in defence analysis. *Syst. Dynamics Rev.* **3**(2), 102–115.
41. E. F. Wolstenholme (1990) *System Enquiry*. Wiley, Chichester.
42. C. A. Roberts and B. C. Dangerfield (1991) System dynamics and statistics: recovering the AIDS incubation time distribution from right-censored data. In *System Dynamics 91* (K. Saeed, D. Andersen, and J. Machuca, Eds.), pp. 744–755. System Dynamics Society, Cambridge, MA.
43. R. G. Coyle (1995) *System Dynamics Modelling*. Chapman and Hall, London.
44. P. H. Franses (1994) Fitting a Gompertz curve. *J. Opl Res. Soc.* **45**, 109–113.
45. B. C. Dangerfield (1994) The system dynamics modelling process and DYSMAP2. In *Modelling for Learning Organisations* (J. D. W. Morecroft, and J. D. Sterman, Eds.), pp. 203-209. Productivity Press, Portland.
46. J. D. W. Morecroft (1984) Strategy support models. *Strategic Mgmt J.* **5**, 215–229.
47. C. C. Holt, F. Modigliani, J. F. Muth and H. A. Simon (1960) *Production Planning, Inventories and Workforce*. Prentice-Hall, Englewood Cliffs, NJ.
48. T. L. Vincent and W. J. Grantham (1981) *Optimality in Parametric Systems*. Wiley, New York.

Simulation by Repeated Optimisation

R.G. Coyle

Introduction

It is a cardinal point in system dynamics that the behavior of a model can be improved, often quite remarkably, by experimental changes to the parameters representing the system's policies. Usually, even more significant enhancements in behaviour stem from changes to the model's structure. It should be stressed, therefore, that, throughout this paper, 'parameter' includes those switches which can be used to activate or suppress components of the model's structure as well as those which represent its policies.

The only drawback is, however, that there is always a nagging doubt that, had one tried only one more experiment, something even better

Journal of the Operational Research Society (1999) **50**(4)

Professor Geoff Coyle has taken early retirement from his academic post. He continues to be active in system dynamics and was the recipient of the System Dynamics Society's first Lifetime Achievement Award.

R.G. Coyle (✉)
Shrivenham, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_7

would have been found. Unfortunately, there is *always* yet one more experiment, so the doubts can never be assuaged.

The problem of an experimental approach to policy design to improve dynamic behaviour is more difficult still. It has long been a tenet in system dynamics that feedback loops are the unit of analysis in planning and interpreting simulation experiments. That is probably true for very small models but is impractical for a model of even modest size. For example, the well-known world model is only a hundred or so equations yet it contains nearly 2000 feedback loops, most of which are practically impossible to detect visually, and it is clearly not possible to assert *a priori* that some which are easy to see are more important than the rest.

It would, therefore, be highly desirable to have some automated way of performing parameter variations and reporting to the analyst the best result. However the number of possible combinations and conceivable numerical values of the parameters is usually colossal so testing all parameter combinations is impractical. One needs, therefore, some sort *of guided* search of the parameters to be considered, and the numerical value each might have, so as to seek out the result which is most rewarding in terms of enhancing the system's performance, without pursuing blind alleys. Unfortunately, there is no perfect way of achieving that, but the principle of dynamic optimisation comes very close to providing this subtle searching of the design possibilities of the system.

This concept has been used in previous system dynamics work. Winch [1], for example, linked a FORTRAN version of a system dynamics model to optimisation software. This was laborious so, in collaboration with the Helsinki School of Economics, the DYSMAP package was linked to optimisation software [2]. Coyle [3] demonstrated the method for a production problem. Dangerfield and Roberts [4] describe some aspects of the approach. Coyle [5] has applied optimisation to a complex problem of defence planning, and Wolstenholme [6] similarly studied tactical choices by an attacking force.

There is, however, no compact account of the technique and this paper will discuss the theory of dynamic optimisation, which is not widely used within the usual disciplines of OR, illustrated by examples [7]. An important aspect of dynamic optimisation is the development of suitable objective functions which is also considered.

The System Dynamics products which originally supported optimisation in the late 1970s were DYSMAP2, and COSMOS, followed within the last few years by VENSIM and Powersim. All of them automatically link an 'ordinary' system dynamics model to the optimisation software, though the details and simplicity of doing so depend on the package. DYSMAP2 and COSMOS are very similar and use an efficient hill-climbing algorithm. VENSIM appears to use a grid search approach. Powersim's optimiser uses not more than five parameters and operates as an automated sensitivity tester. COSMOS is used to illustrate the ideas in this paper. For a fuller discussion of system dynamics software see Reference [7].

The System Response and the Parameter Plane

Figure 1 shows a two-dimensional picture of a three-dimensional object. The two dimensions in the horizontal plane are labelled for two of the parameters in a model. It is essential to realise that these may be ordinary policy parameters, or they may be structural parameters, or they may be 'pressure points', at which investments of resources could be made in a system. Each parameter has a range within which it may lie, shown as, for example, $P1_{UPPER}$ and PILOWER- The ordinate is a measure of the quality of dynamic behaviour which the model produces for any given combination of parameters. For the moment, we shall simply label that scale as running from 'Bad' to 'Good'. The quantification of performance is discussed later.

The initial, or base case, values are labelled Pl_1 and Pl_2 on the parameter axes. When these two values are projected into the 'parameter plane', following the dotted lines, they intersect at Point A. When the model is run with those values, the response is, we shall suppose, rather poor, so a short line is drawn in the vertical direction to indicate that. This idea of the model's response at a point in the parameter plane is valid regardless of the number of parameters being used. In fact, in earlier work for a commercial client COSMOS was used to optimise a model of some hundreds of equations with 35 parameters, the runs took a few seconds, once the search parameters had been loaded.

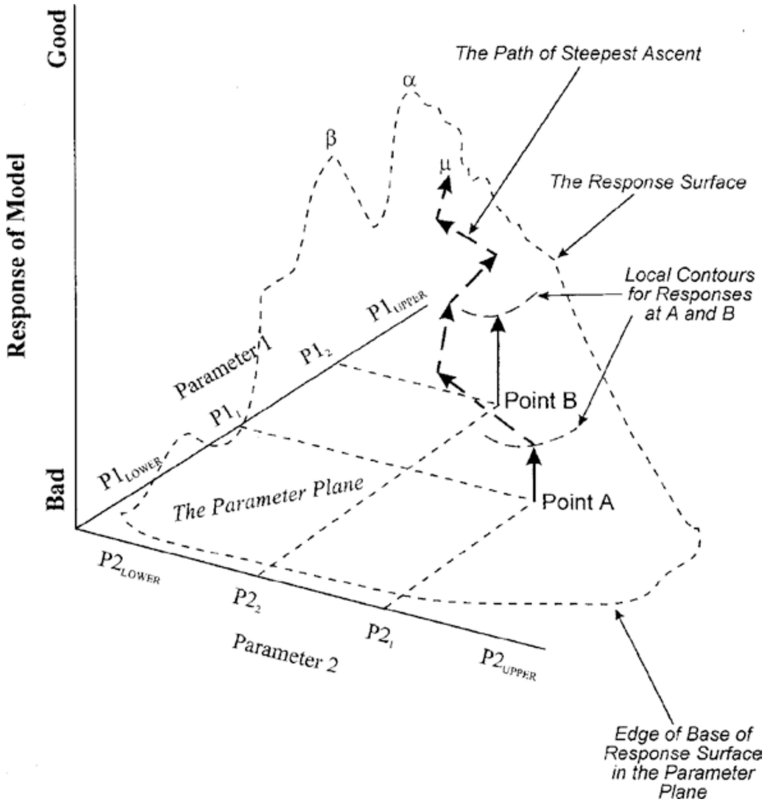


Fig. 1 The concept of hill-climbing optimisation

If the two parameters are now changed to Pl_1 and Pl_2, a new point B is defined which, we shall imagine, produces better performance and hence qualifies for a rather longer arrow in the vertical direction.

Figure 1 develops the idea of a response surface as the three-dimensional locus of the responses at all possible combinations of PI and P2. The response surface could well be a very rugged mountain, with several peaks. Two of the peaks, α and β, are of rather different heights and α is clearly 'better' than β. Figure 1 also suggests that there are all sorts of irregularities and gullies and that the mountain slope is much steeper in some places than in others; it could well resemble a photograph of the Alps. The reason for the extreme irregularities lies in the complexity of system dynamics models and the non-linearities which they often contain.

It was argued earlier that one cannot test all possible combinations of parameters but, unless that is done, the shape of the response surface will be unknown. The problem of dynamic optimisation is how does one find one's way to the top of a mountain of unknown shape?

Hill-Climbing Optimisation

Hill-climbing optimisation can be understood by analogy with a blind man who is marooned on a mountain and wishes to find his way to the top. His strategy is to feel the shape of the ground around the point where he is sitting (the top of the arrow at Point A). Having detected the direction in which the ground slopes up most steeply in that vicinity, he takes a cautious step or two in that direction and then feels the ground again. In this way, he hopes to find the top of the mountain, as shown by the sequence of arrows moving up the surface of the hill from the point of the arrow at Point A, always pursuing the locally steepest direction. Unfortunately, the blind man's strength eventually fails him and he can go no further than μ. This is not even as high as β, so there is no guarantee that the blind man will reach his goal, α, or even another lower peak.

The analogy is implemented in some system dynamics software packages by using hill-climbing algorithms, developed in the mathematics of numerical analysis, which repeatedly iterate the model for particular sets of values of the collection of parameters being search. The first iteration uses the base case parameter values. After the model has run, the value of the objective function is calculated and retained, with the attendant parameter values, if it is the best result found so far. The sets of parameter values which give *good* results are used to predict how the parameter values should be changed to carry out the guided search for good values without wasting effort examining parameter combinations which lead nowhere. The reader interested in the mathematical technicalities should refer to the literature on numerical analysis; the pragmatist can rely on the evidence that the approach works.

The equivalent of the blind man's strength failing him before he reaches the top of the hill is not commanding sufficient iterations to be performed. If the algorithm is efficient there is no harm in demanding 500 iterations, if one wishes. Experience indicates that, for many models, about 30 iterations are sufficient to find dramatic improvements in performance. Even with a large model, a few hundred iterations take no more than a couple of minutes even on a modest PC.

An alternative to hill-climbing is a grid search in which the algorithm searches the space and gradually homes in on a good solution. This can be much more computationally demanding than hill-climbing. More recent techniques include genetic algorithms.

Overcoming the Limitations of Heuristic Algorithms

Hill-climbing by the method of steepest ascent is clearly heuristic and there is no guarantee of finding the maximum of the response surface. However, numerical hill-climbing algorithms are sophisticated and are capable of searching with something close to the intelligence that the blind man would use.

Consider Fig. 2, the vertical axis is still the model's response but the horizontal axis should be imagined to be a 'cross-section' of the parameter plane, representing varying combinations of the two parameters. The hill-climbing search has started at Point A and followed a steep ridge until it reached B. From there, a valley leads forward into the hills, and there is a ridge to the left, but the locally steepest direction is towards C. On reaching C, it becomes clear that it is a false peak, from which all directions lead downwards. Since the man is not yet exhausted, he would remember C as the best point so far discovered and, reasoning that he can always go back to C, he accepts the loss of height and searches towards D. At that point, he discovers a slowly rising path which emerges from behind peak C, as shown by the faint line, and moves off to E and hence, we hope, to F. As soon as any point is reached which is better than C, it will be remembered as the point to which he can return if no better solution is found.

There is still no guarantee that this course of events will unfold, so another strategy for avoiding failing to find the real peak would be to

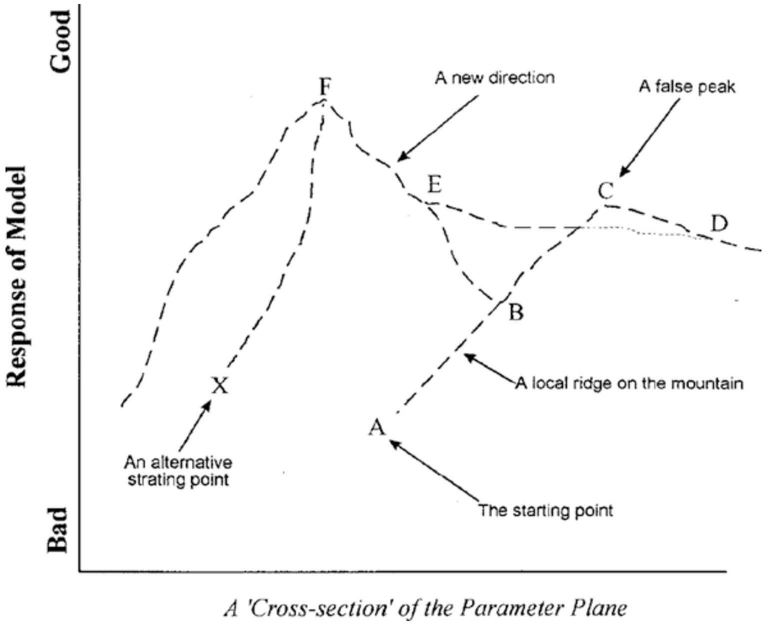


Fig. 2 Continuing to search

recommence the search at a new starting point, X, in the hope that there is a ridge going directly to F. In practice, the improvement in behaviour found in the first optimisation is usually so large that it is hardly worth the effort of repeating searches from new starting points. The main benefit of doing so would be to increase the experimenter's confidence that he has found a good solution, but there would be no end to that process.

The Performance of a Simple Model

Figure 3 is the Influence Diagram for a very simple model of a production system. The firm maintains inventory to meet unpredictable demands for parts. Inventory is replenished by ordering parts to be manufactured and delivered within a 'backlog elimination time'. The firm is hit by a sudden rise in consumption followed by an even larger fall, as shown by the solid curve in Fig. 4. The ensuing behaviour is clearly fairly disastrous. Desired

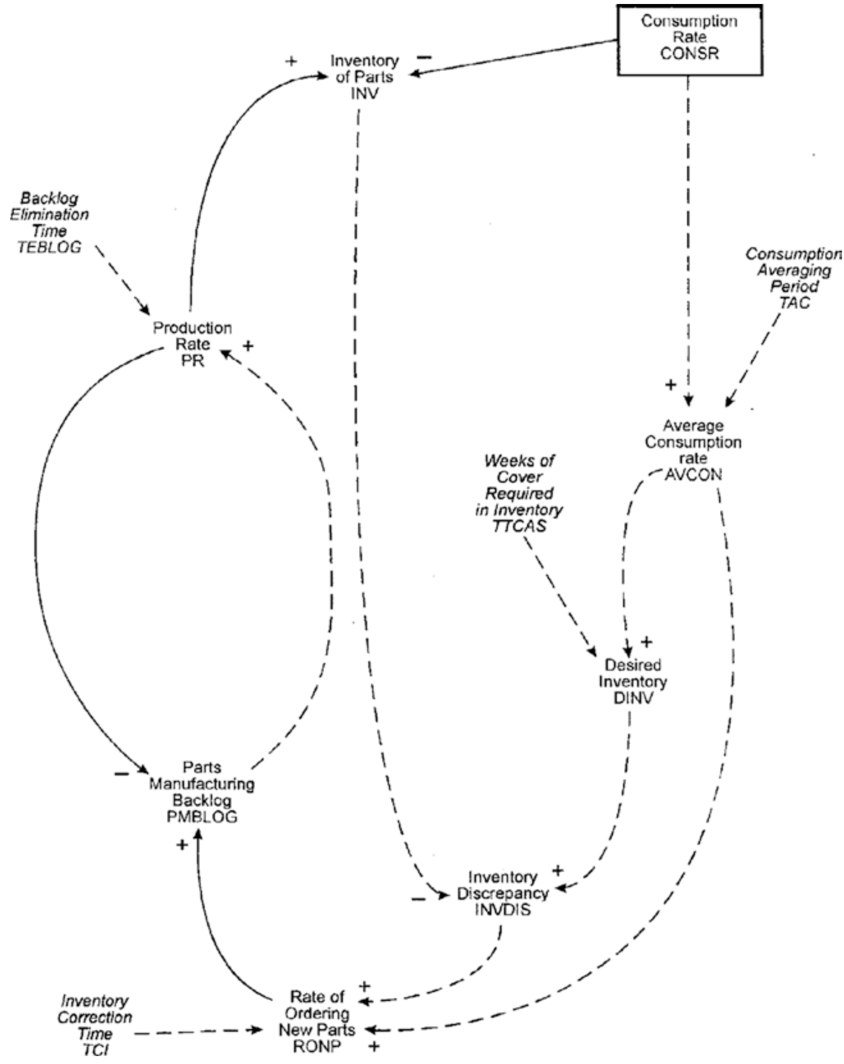


Fig. 3 Influence diagram for optimisation illustration

and Actual Inventory (DINV and INV) do not match until TLME = 140, which is two years after the shocks in consumption.

Furthermore, the parts backlog rises to a peak of about 1100 at LME = 30, but then falls to zero at TLME = 55 and stays there for

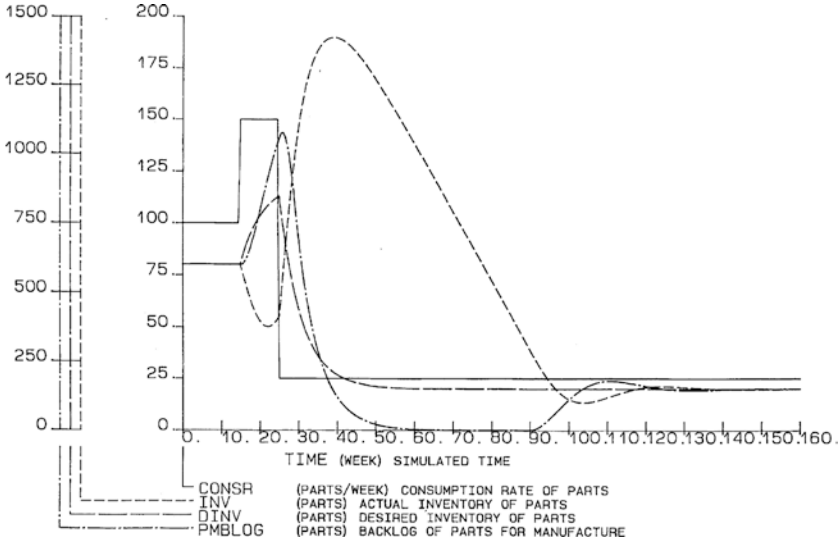


Fig. 4 Basic behaviour of the simple model (base case)

35 weeks. This corresponds to the factory having to be closed for about eight months.

Such a problem is small enough to be tackled by analysing gains and delays in the feedback loops but it is used here to demonstrate optimisation.

Formulating an Objective Function

To optimise, it is first necessary to formulate an objective function as a measure of system performance to guide the optimisation search. It should take account of what the system is trying to achieve and calculate the extent of its success. Developing objective functions for system dynamics models is something of a black art calling for much more research. Simple approaches based on common sense do, however, work well and, in this case, it seems reasonable to assume that the management objective is to make INV match closely to DINV. Much more sophisticated objective functions are used in some of the work cited earlier and that described below.

To write, using the standard DYNAMO format:

$$\text{A OBJFUN.K} = \text{DINV.K} - \text{INV.K}$$

would be misleading as it refers to a single point in time. One must take account of the behavior over the whole of the run, which requires a level variable to act as a memory of what happened during the run. To avoid the positive and negative discrepancies simply cancelling each other one might define an Inventory Penalty, INVPEN, as:

$$\text{L INVPEN.K} = \text{INVPEN.J} + \text{DT}^{*}\left(\text{DINV.J} - \text{INV.J}\right)^{**} 2$$
$$\text{N INVPEN} = 0$$

The numerical value of INVPEN will be some strange numbers so, to make performance comparisons easier, we redefine OBJFUN.K as:

$$\text{O OBJFUN.K} = \text{INVPEN.K / SCALE}$$

where SCALE is chosen to make OBJFUN, the objective function we shall actually use, equal to 100 at the end of the base case run, simply to make it easier to think in percentage terms. In this case, SCALE is equal to 46.7933×10^4 which is found by running the model with SCALE = 1 and the base case parameters, observing the value of INVPEN at the end of the run, and calculating SCALE accordingly. In this case, it is evident that we wish to minimise OBJFUN, which is tantamount to the blind man finding his way down the crater of a volcano (hopefully extinct) by the method of steepest descent.

The Significance of the Objective Function

It is very important to be quite clear about the significance of objective functions. In the first place, they are extra equations added to the model for the analyst's benefit. They are not part of the real system and do not necessarily have physical meaning. They are only there to help the analyst, and the software which serves him, to keep track of how improvements to behaviour can be found.

Secondly, the dimensions of the objective function do not have to have a real-world meaning. The dimensions of INVPEN are [WEEK*PART [2]], which does not correspond to anything in the real system, but that can be ignored. Obviously, there is nothing wrong in having an objective function which is dimensionally sensible, but the objective function, being an artefact of the analyst's thought, does not have to obey the strict requirement for dimensional consistency which applied to all the rest of the system dynamics model [8]. In this instance, SCALE has the same dimensions as INVPEN, so OBJFUN is a dimensionless ratio.

Thirdly, one has to be very careful about choosing objective functions. Minimising average inventory is an attractive though, but the true minimum inventory is zero and a firm which has zero inventory is usually not going to sell very much. As in all of OR, selecting an objective function is not a trivial matter and careful thought is needed about what the firm is really trying to achieve.

Optimisation Experiments

To optimise, one has to specify the parameters to be searched and to state the upper and lower values of each one. In this problem, there are four parameters, TAC, TCI, TTCAS and TEBLOG. TTCAS is a gain, the rest are delays. To start with, we allow all four to be in the parameter plane to be searched. All the base case values are 6, and they are all allowed to lie in the range from 2 to 10. These ranges are chosen purely for illustration but, in practice, much discussion with management would be involved and several ranges might be tried. This is often done most effectively during an intensive study period as optimisation will take only a few seconds on a Pentium PC. Thirty iterations are done and the results are shown in Fig. 5a.

The optimal values of the parameters are reported to TAC = 10, TCI = 2, TTCAS = 2 and TEBLOG = 2. The value of OBJFUN falls from 100 to 4.03. This reduction of practically 96% is by no means

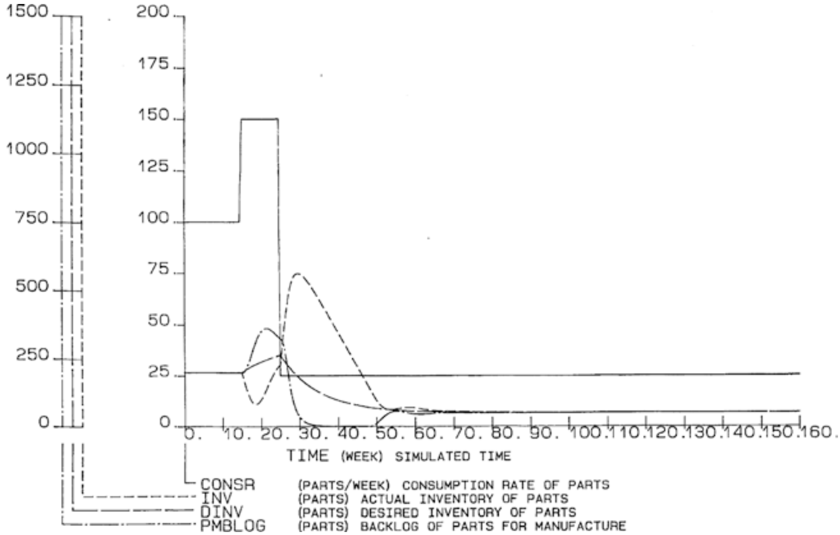


Fig. 5a Optimisation of the simple model (a) Optimisation of four parameters

unusual. All the parameters have been driven to their extreme points. Interestingly, the behaviour of other variables as well as inventory has improved, not surprisingly give the inter-connectedness of policies in a system dynamics model. A particular case is that the factory closure now lasts for only 15 weeks.

The key point is that this result was discovered after a few minutes of work to add OBJFUN to the model and a few seconds of computer time. To have discovered it by traditional loop analysis and experimentation would have taken much longer. The difference in effort would be *much* greater for a large model, even if loop analysis was tractable.

It is interesting to note that the gain, TTCAS, has been reduced, but that only one of the delays, TAC, has been increased while the other two delays have been reduced. It is a rule of thumb in control engineering that reducing gains and increasing delays is likely to increase stability but it has been clearly not worked in this case. Such rules are used in the traditional, loop-based, design approach in SD and, in this instance, optimisation has called them into question.

It is usually not a good idea simply to throw all the parameters into the optimiser and take the results on trust. To do so is to abandon thought and rely on computation. In this case, we might feel that to increase TAC too much could make the system too insensitive to the unpredictable changes in CONSR, so another optimisation is done in which only TCI, TTCAS and TEBLOG are allowed to enter the parameter plane, with the same ranges as before. The results (Fig. 5b) are that all three are driven to their lower value of 2 and OBJFUN is reduced to 5.268. On the face of it, this is nothing like as good as the previous case, because OBJFUN is about 30% larger. In fact, the visible differences are very slight, as shown in Figs. 5a, 5b. The only difference between the two plots is that INV reaches a maximum of 562 with 4 parameters and 581 with three.

The optimisation with three parameters has driven the value of TTCAS to 2, which means that the firm is trying to operate with only two weeks of stock cover. That might be quite insufficient for any noise in the demand pattern and provides little protection against another upsurge in CONSR. One therefore optimises again with TCI and TEBLOG ranging from 2 to 10 as before, but TTCAS in the range from 4 to 10. The optimisation pushes all three parameters to their lower limits and OBJFUN is reduced to 12.84. In management terms, the parameter

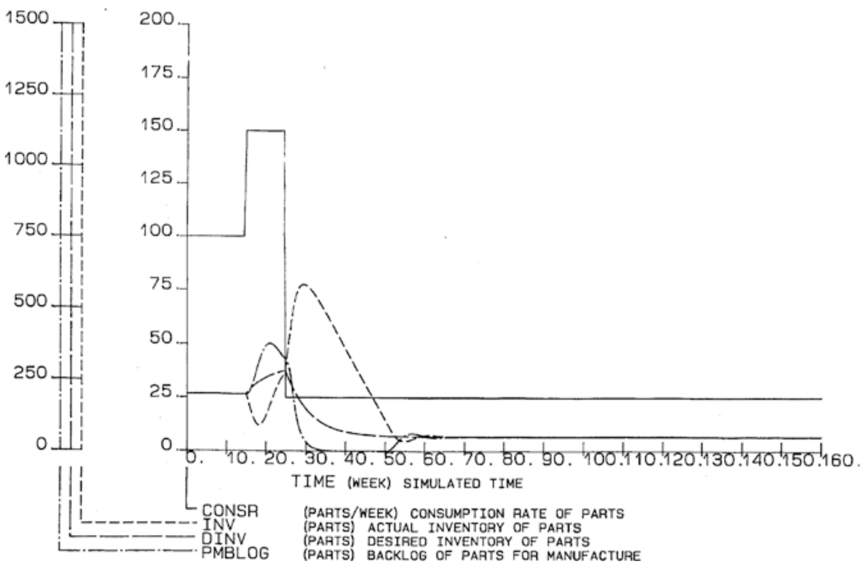


Fig. 5b Optimisation of the simple model (b) Optimisation of three parameters

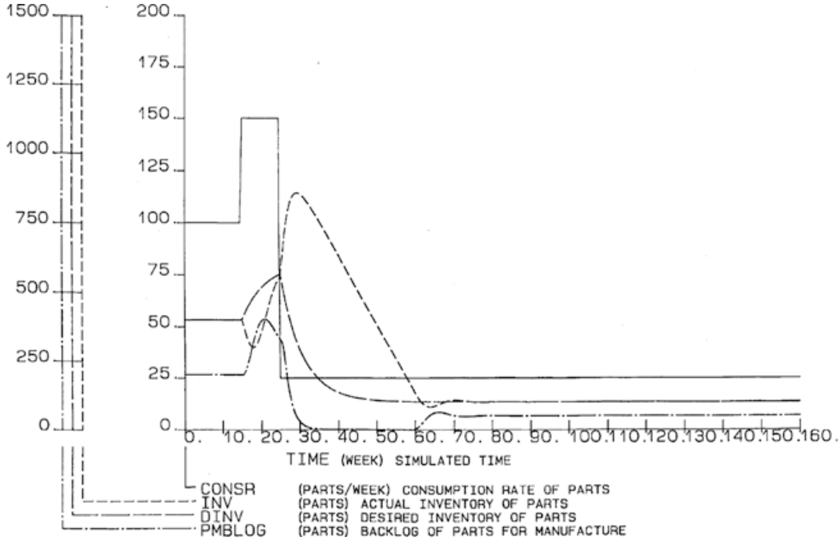


Fig. 5c Optimisation of the simple model (c) Higher stocks

values correspond to holding a prudent level of stock, but reacting quickly to changes. The behaviour is shown in Fig. 5c, and is clearly not as good as the cases in Figs. 5a, 5b. The maximum value of INV is now 857. The difference between 857 and 562 is a quantified indicator of the costs of a qualitative decision to be prudent, though the factory closure is now 25 weeks rather than 15, which might affect the cost of prudence.

This example has been deliberately chosen to be very simple so as not to lose the subtlety of optimisation in the explanation of a complex model. The real power of optimisation comes through with more complex models in which loop analysis can offer no guidance.

Simulation by Repeated Optimisation

The concept that the top of the hill is sought by repeatedly running the model makes it clear that the technique is optimisation by repeated simulation. However, the two other cases, with reduced numbers of parameters and with different ranges, should make it clear that the real underlying theme is *simulation by repeated optimisation*. The model is optimised and something is learned. That leads to further optimisation and more learning, and so on. The value of the optimisation calculation is to provide a much more powerful guided search of the parameter space than could possibly be achieved by ordinary experimentation, but the principal aim is still to experiment and to understand so that a better experiment can be designed.

A notable example of this is the observation of differing periods of factory closure. The model could be optimized again to minimise that or to balance inventory against closure. In a more complex model, there might be many possible objectives, which would often be discovered by studying the effects of previous optimisations. Some of these objectives might, of course, be expressible in financial terms.

Objective Functions for Constrained Optimisation

The optimisation of the simple model was free in the sense that, in theory, no costs are incurred when adopting a new policy. Optimisation of resource investments is not free and the response surface can be visualised roughly as indicated in Fig. 6. The slope is as rough as the mountain but a barrier exists at the point at which all resources have been expended. Since the unit costs of different resources are different, the barrier has to be imagined as being moveable as large amounts of cheap resources may have a different effect from small amounts of expensive ones. Since one is, however, usually optimising several resources of differing unit costs, the barrier can be thought of as being on hinges, as well as on rollers.

Objective functions for constrained optimisation have to allow for resource costs and can be written in terms of OBJFUN as a measure of what is to be achieved and a penalty to be applied if the resource budget is exceeded. Therefore to maximise a performance index, INDEX, subject to a given budget, BUDGET, one could use:

$$\text{A OBJFUN.K} = \text{INDEX.K / SCALE} - 100\text{E}06 \\ \times \text{MAX}\left(0,\ \text{COST} - \text{BUDGET}\right)$$

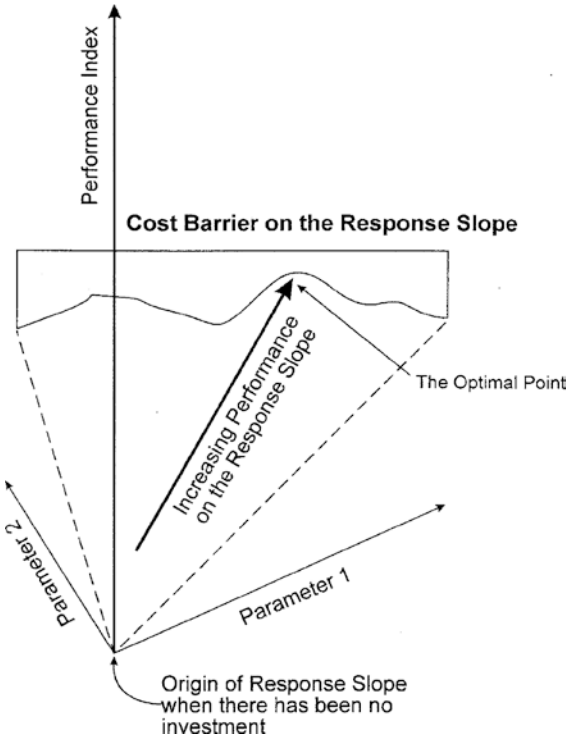


Fig. 6 Response slope for constrained optimisation

COST would be calculated to reflect the various resources expended and, as soon as it exceeds BUDGET, a massive penalty is subtracted from INDEX, making maximisation at that level of COST impossible. Note that, in DYNAMO terms, COST is a computed constant and hence has no K time postscript, the full details are covered in Reference [7]. It is trivial to extend this so that there are two COSTs and two BUDGETS when there are two resources such as manpower and money, and so on for any number of budgets and resources, such that exceeding any one would invoke the penalty.

It is equally trivial to adapt this process to minimising the expenditure required to achieve a given level of performance as opposed to maximising the performance for a given level of expenditure.

Constrained Optimisation—A Case Example

A defence example of constrained optimisation may illustrate the approach. In a model of land operations by a divisional formation, there are 10 parameters which can be freely changed and 20 which involve expenditure. The former represents 'decision rules' or concepts of operations, such as when to commit reserves, when to use forces of a particular type and so forth. The latter group includes such factors as the numbers of attack helicopters, the amount of ammunition, the numbers of combat engineer vehicles and so forth. There is a cost for each of these, such as each attack helicopter costs £X, and there is an overall budget, £Y.

Given a suitable objective function, such as the time for which the division could continue to fight an enemy force of a certain size, three optimisations can be done. In the first, only the 10 free parameters are used. In the second, the 20 costly assets are used with a budget constraint added to the objective function. In the third, all 30 parameters and the constraint are used. The results are in Table 1, in which the base case performance without any changes to any of the parameters is scaled to be 100, and the other performance figures are purely illustrative. F and C, with subscripts 1–3, denote vectors of the optimal values of the 10 free and 20 costly parameters, respectively.

The results are informative. The change from a performance of 100–120 with the free optimisation really means that changing F from F_1 to F_2 is the concept of operations which will get the best from the available forces. Similarly, the change from 100 to 140 as one optimises the costly parameters means that C_2 is the best force composition with the given budget for the existing concept of operations. However, the salient result arises when performance increases to 180 and both F and C change

Table 1 Optimisation of a divisional force

Base case	Free parameters only
Performance 100	Performance 120
Vectors: F_1, C_1	Vectors: F_2, C_1
Costly parameters only	Both costly and free parameters
Performance 140	Performance 180
Vectors: F_1, C_2	Vectors: F_3, C_3

to the new values of F_3 and C_3. This can be interpreted to mean that F_3 is the concept which gets the best out of forces C_3 or that C_3 will allow a better concept, F_3 to be adopted.

Finally, there is the point that the performance differences between 120 with free optimisation and 180 with costly and free optimisation is the marginal return to investment of the budget of £Y. The difference between 120 with costly optimisation and 180 is the marginal benefit of changing concepts to get the best results from investment.

Constrained Optimisation and Marginal Investments

Again, this is a defence example which summarises the work in Reference [5]. The small nation of Heroica faces potential aggression from its powerful neighbour, Nastia. Heroica is allied in its region with two other small nations, North Phalia and East Phalia. All three are threatened by Nastia and are allied with the distance superpower of Columbia. War is not certain but, should Heroica be attacked by Nastia, its forces would fight to maintain land, sea and air control of its territory. If Columbia is able to come to Heroica's aid, air reinforcements can fly directly into Heroica's air bases, if there is room for them, but land force reinforcements will have to travel through dangerous oceans to the region and will then reinforce Heroica, or one of the Phalias, depending on need and on whether they can cross Heroica, or one of the Phalias, depending on need and on whether they can cross Heroican waters in the face of the Nastian navy.

A very simplified influence diagram for the problem appears in Fig. 7 in which some of the feedback loops have been emphasised. The heavy black loop is the snowball loop by which, once Heroica starts to lose control, the loss will accelerate. The two dotted loops are command and control loops which re-allocate forces to prevent the snowball running out of control, or even to turn into Heroica's favour. Again, an objective function can be formulated to maximise Heroica's retention of control subjects to the available budgets for defence improvements over the next few years.

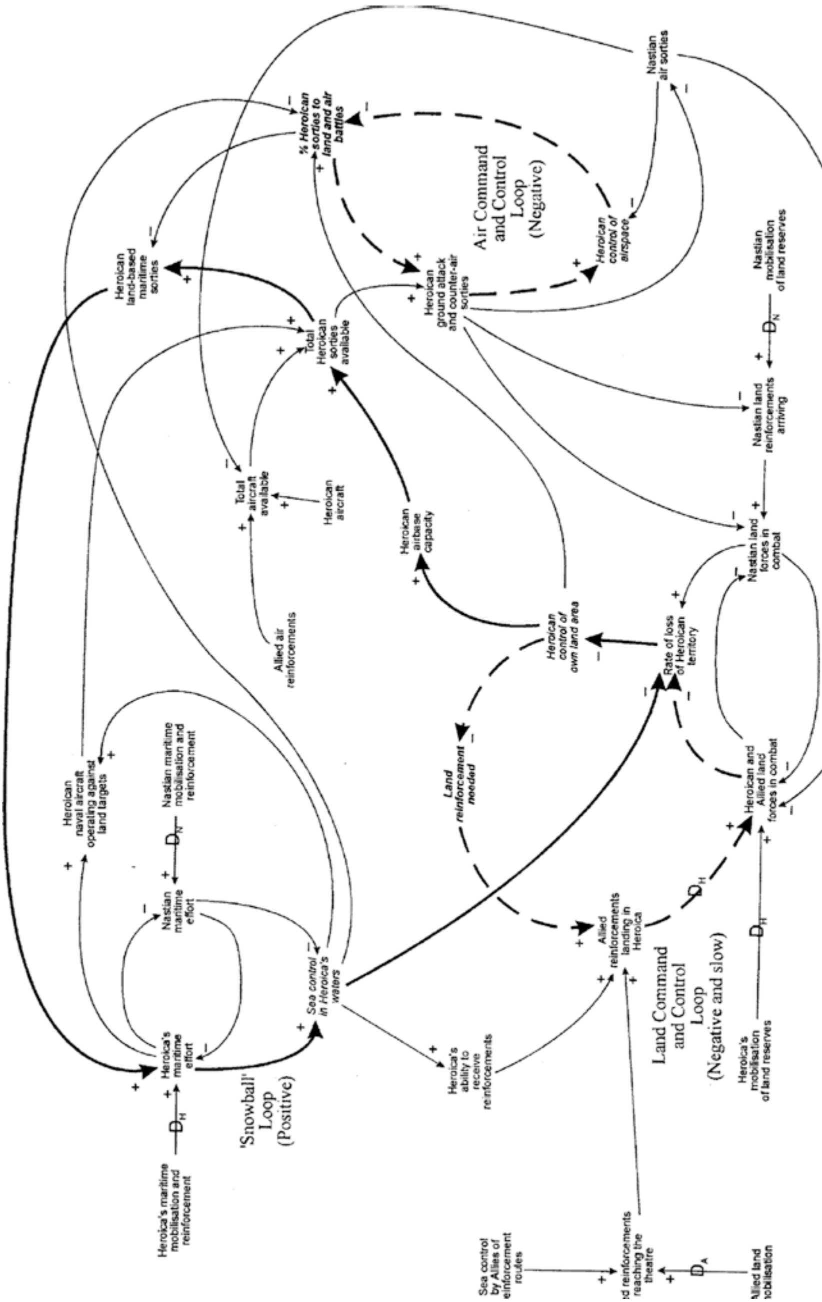


Fig. 7 Simplified influence diagram for Heroica's defence strategic planning model

Table 2 The optimisation results

Spending level	0	1	2	3	4
Overall outcome	Defeat	Time gained	More time gained	Enemy halted	Enemy defeated
Regular land force	2200	2400	3100	3100	3200
Land reserves	3300	3300	3300	3500	3600
Mobilisation delay	4	2	2	2	2
Air force	525	525	525	550	650
Air *base* capacity	4400	4500	4500	4500	4500

The problem is that the budget is unknown, so the optimisation is repeated with four separate levels of expenditure and using a few illustrative parameters, the results appearing in Table 2. It is, however, no use saying to the high command that 'if your defence budget is X your force composition should be such-and-such', as no-one knows what the defence budget is going to be over the next few years. However, scrutiny of Table 2 shows that the mobilisation delay is always driven to its minimum value, the regular army increases somewhat as the budget increases, but then levels off. The reserves and the air force only increase at high levels of expenditure and air base capacity does not change very much. That implies that the first priority should be to reduce mobilisation delay. After that, if there is any money left, the regular army should be increased to the point where it can defend the frontier for a few more days and, finally, if there is still some money left, it should go to the reserves. The air force, as its assets are costly (in this imaginary case), only benefits at very high levels of budget. In this case, optimisation has identified a *programme*, rather than a *single result*, which may, in practice, be the more useful product.

Concluding Remarks

There are numerous potential applications of optimisation, such as fitting models to historical data as part of the process of validation, and in business and government modelling, but space precludes a more extended discussion.

This paper has summarised the underlying theory of optimisation as applied to system dynamics models and shown something of the power of the approach by some examples. The main point to grasp is that optimisation, for all its superficial attractiveness, is not a panacea and it does not guarantee good analysis. There are, indeed, some limitations to the approach.

The first is that the hill-climbing algorithm does not guarantee an optimal solution. In practice, that matters less than might be thought because most managed systems perform so badly that any improvement is welcome and the differences between optima are usually much less than the objective function values might imply, as we saw in Figs. 5a, 5b.

Secondly, the optimisation technique does not, of itself, give any guidance on the development of a good, subtle, objective function. In many cases, the objective function reflects qualitative criteria, such as how well targets are achieved and these can have significant effects as they are linked to management thinking. A simplistic objective function, such as minimising inventory, might be truly disastrous.

The final weakness is that the thought of optimizing something is so seductive that the naive analyst might stop thinking.

In practice, it is only the second and third limitations which are serious and, provided one thinks, optimisation may well be a very powerful development in system dynamics.

Finally, if classical system dynamics involves improvement by successive simulation guided by the analyst's understanding of the feedback loops, optimisation is improvement by successive multiple simulation guided by an objective function which evolves as the analyst's understanding of the ability to meet objectives grows.

References

1. Winch GW (1977). Optimisation experiments with forecast bias. *Dynamica* **2**: 000–000.
2. Coyle RG (1997). System dynamics at Bradford University. *Sys Dynam Rev* **13**:311–321.

3. Coyle RG (1985). The use of optimisation for policy design in a system dynamics model. *Sys Dynam Rev* **1**: 81–91.
4. Dangerfield B and Roberts C (1996). An overview of strategy and tactics in system dynamics optimisation. *J Opl Res Soc* **47**: 405–423.
5. Coyle RG (1992). The optimisation of defence expenditure. *Eur J Opl Res* **56**: 304–318.
6. Wolstenholme EF (1990). *System Enquiry.* Chichester: John Wiley and Sons.
7. Coyle RG (1996). *System Dynamics Modelling: A Practical Approach.* London: Chapman and Hall.
8. See the work cited in Reference 7 for a full discussion of dimensional analysis.

Judgement and Supply Chain Dynamics

A.A. Syntetos, N.C. Georgantzas, J.E. Boylan,
and B. Dangerfield

Introduction

Judgemental adjustments constitute one of the few practical ways for organizations to take account of internal and external performance drivers in their demand forecasts (Fildes et al. 2009). Experts may know that institutions will change, that events have a specific impact that is not included in

Journal of the Operational Research Society (2011) **62**(6), 1138–1158.
https://doi.org/10.1057/jors.2010.56
Published online 9 June 2010.

A.A. Syntetos (✉)
Cardiff Business School, Cardiff, UK

N.C. Georgantzas
Fordham University, Bronx, NY, USA

J.E. Boylan
Buckinghamshire New University, High Wycombe, UK

B. Dangerfield
Department of Management, University of Bristol, Bristol, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_8

the model being used, or that a variable difficult to measure is missing from the model. Such knowledge is often reflected by adjustments to the statistical or model-based forecasts produced by software solutions.

A number of empirical and 'laboratory' studies have been conducted to evaluate the effect of such adjustments on the forecasting task. A few attempts have also been made to model adjustments to offer some insights into their underlying properties (Franses 2007; Fildes et al. 2009). However, no formal modelling has been conducted that could offer insight into the impact such adjustments might have on the entire supply chain performance and the potential development of mechanisms to mitigate any adverse effects. This is because of the associated dynamic complexity that precludes the development of closed-form mathematical models, unless rather simplifying (and thus not particularly realistic) assumptions are employed for such an exercise.

Moreover, while the effect of judgemental adjustments on forecast accuracy has received attention in the academic literature, its inventory implications have been largely ignored. A number of research projects have demonstrated that the efficiency and effectiveness of inventory systems do not relate directly to demand forecasting performance, as measured by standard forecasting accuracy measures. Syntetos et al. (2009b, 2010) examined the stock control implications of judgementally adjusting statistical forecasts through the consideration of relevant accuracy-implication metrics (inventory and service level achieved). They found that the inventory performance could not be explained in terms of the forecast accuracy alone and that the inventory benefits were of a completely different order of magnitude to the forecast accuracy gains. Both results can be attributed, conceptually at least, to the interdependence between forecasting methods and stock control rules in inventory management systems (Syntetos and Boylan 2008). However, presently, the dynamic effects of such interactions are not fully understood. In particular, the reasons why relatively small forecast accuracy changes may translate into significant inventory benefits requires more investigation. It is the contention of this paper that the system dynamics (SD) modelling method can play a crucial role in advancing knowledge in this area, where mathematical modelling cannot accommodate the associated dynamic complexity.

In an inventory setting, demand forecasts are used in conjunction with appropriate stock control rules in order to decide how much to order.

Kolassa et al. (2008) examined the inventory systems of three major German retail corporations and reported that replenishment orders may also be subject to judgemental intervention. (This may, or may not, occur in conjunction with judgementally adjusting sales forecasts.) Our study constitutes part of a UK government (EPSRC) project and all our industrial partnering organizations in this project rely, to a certain extent, on adjusting forecasts and/or replenishment orders. This is also true for a number of other companies we have examined. The effects of judgementally adjusting replenishment orders have never been studied before or explored in the academic literature.

Judgemental adjustments should have a considerable effect on supply chain performance and the creation of bullwhip- type phenomena through, for example, stock amplification as we move upstream in the supply chain. The purpose of our work is to study this effect in detail. Our investigation entails the application of SD models on a wide-range of scenarios, to evaluate the effects of forecasting and/or ordering adjustments on supply chain performance. A three-stage supply chain is considered (*Retailer, Wholesaler* (Home Office/Co-ordinating Unit) and *Factory*) and results are presented for three well-known stock control policies: (i) the linear Anchor and Adjust (*AaA*); (ii) the re-order point *s*, order- up-to-level *S* (*s*, *S*); and (iii) the order-up-to-level *S* (*utS*). Other control parameters include the point of intervention (the stage at which the supply chain managers intervene to make the adjustments, including all possible combinations), and the nature of the adjustments (purely optimistic, purely pessimistic and mixed downward and upward optimistic and pessimistic). The performance metric we employ is the factory stock *amplification ratio* (Sterman 2000, p. 673), which relates the maximum change in stocks at the factory level to the maximum change in forecasts or orders, as a consequence of judgemental adjustment. The results enable insight to be gained into the effects of judgemental interventions on bullwhip amplification in the supply chain.

The remainder of our paper is structured as follows: in the next section the research background is presented, followed in Section "Experimental Structure" by the experimental structure employed for the purposes of our research. Details related to SD modelling are presented in Section "Model Description". Section "Analysis of Results" contains the results of our investigation and their analysis and interpretation. The paper concludes with an agenda for further research in Section “Conclusions and Extensions”.

Research Background

In this section the literature related to our research is reviewed under three main sub-sections. First, we discuss studies that have considered the effects of judgemental interventions in supply chain management. Second, we review contributions in the area of the bullwhip effect; its contributory factors are synthesized in a diagram that illustrates the potential effects of judgemental interventions. Third, we consider a number of SD projects that focus on supply chain modelling and management.

Judgemental Interventions in Supply Chain Management

Franses (2007) investigated judgemental adjustments to model-based forecasts and explored whether they contributed to forecasting accuracy, that is, is there any value being added? He also examined whether the contribution of the model forecast and that of the expert added value (as opposed to expert forecast) are of equal importance, that is, does the 50% model—50% manager rule (advocated by Blattberg and Hoch 1990) hold?

In the above study, the null hypothesis that the root mean squared error prediction (RMSPE) of the expert is equal to that of the model (against the alternative that the expert is better) was tested based on recommendations proposed by Clark and McCracken, (2001, 2005—one-sided tests). The empirical dataset consisted of monthly observations (demand realizations, model and expert forecasts) for a wide range of pharmaceutical products belonging to different categories and sold across many countries.

The main conclusions are that the expert’s added value frequently matters and when it matters it also frequently occurs on a 50–50 basis, but the experts put too much emphasis on their own added contribution. The implications of this research project are viewed as very important: (i) the way the statistical model works should be better communicated to the experts; (ii) experts should start documenting what they effectively do when they adjust model-based forecasts; (iii) the experts should become aware that they may be putting too much weight on their expertise. When expert judgement is useful, there is no problem, but when it is not, forecasts can become dramatically bad.

A further attempt to investigate the properties of expert adjustments on model-based SKU level forecasts was made by Franses and Legerstee (2009). They analysed a database containing one-step-ahead model-based forecasts adjusted by many experts located in 37 countries, who make forecasts for pharmaceutical products within seven distinct categories. Their results were consistent with earlier findings that the experts make frequent adjustments (managers were found to adjust model-based forecasts in 90% of the cases) and that these tend to be upward. In addition, they found that expert adjustment itself is largely predictable, where the weight of a forecaster's own earlier adjustment is about three times as large as the weight of past model-based forecast errors. Finally, they showed that expert adjustment is not independent of the model-based forecasts. They argued that this interaction should be taken into account in any evaluation of the effect of the contribution of expert adjustment to the overall forecast quality.

Fildes et al. (2009) collected data from more than 60,000 forecasts from four supply chain companies. In three of the companies, on average, judgemental adjustments were found to increase accuracy. However, detailed analysis revealed that while the relatively larger adjustments tended to lead to greater average improvements in accuracy, the smaller adjustments often damaged accuracy. In addition, positive adjustments, which involved adjusting the forecast upwards, were much less likely to improve accuracy than negative adjustments. They were also made in the wrong direction more frequently, suggesting a general bias towards optimism. Models were then developed to eradicate such bias.

Small adjustments have been found not to be very effective. This is confirmed by analysis of both intermittent and faster- moving demand data by Syntetos et al. (2009b) and Fildes et al. (2009), respectively. It seems that in this case the experts merely add 'noise' to the forecasts resulting in higher errors however the errors are measured. Consequently, it may be useful modelling the adjustments (under those conditions) for the purpose of further developing our understanding of their implications. We note that the analysis under concern reflects the unfavourable setting where adjustments do not perform well, that is they do not convey an important piece of information that would be lost otherwise.

Eroglu (2006) explored the variables that affect a forecaster's performance when making judgmental adjustments to statistical forecasts (see also Eroglu and Croxton 2010). He also looked at the conditions under which judgmental adjustments were beneficial or detrimental to forecasting performance. Eroglu (op. cit.) examined two other important issues both of which have received insufficient attention in this context in the academic literature: learning and biases.

The data set used in the study came from the forecasting records of a company that judgmentally adjusted statistical forecasts to improve the forecast accuracy. The original (statistical) forecast, the adjusted forecast and the actual sales in the corresponding time period were retrieved from the research company's data warehouse. Forecasting performance (accuracy improvement, learning and biases) was calculated using these data. The study covered a period of 12 months from the beginning of June 2004 to end of May 2005. Most interestingly, the data set was augmented with independent variables that were measured with a survey instrument, administered in 390 company stores located in several midwestern and southeastern states in the USA. Pertinent constructs included personality, motivational and situational variables. The survey responses were matched with corresponding judgmental adjustments made by the respondent. The main conclusions can be summarized as follows: (i) the analysis of the data provided evidence for accuracy improvement; (ii) the data analysis provided no evidence of learning; (iii) in addition to accuracy improvement, data analysis detected evidence of bias (optimism bias and overreaction bias).

Factors Influencing the Bullwhip Effect

The bullwhip effect is a term promoted by Lee et al. (1997a) but was observed and modelled decades earlier by Forrester (1958). It occurs whenever demand becomes more variable as it proceeds through the supply chain, away from the consumer and towards the supplier. Recent research on the bullwhip effect (e.g., Lee et al. 2000; Zhang 2004) has tended to focus on mathematical modelling and has treated one cause of bull- whip independently of the others. The potential to use SD to understand the interactions between causes of the bullwhip effect is partly addressed by the research reported in this paper.

Lee et al. (2000) discussed four causes of the bullwhip effect, namely: demand signal processing, rationing/shortage gaming, order batching and price fluctuations.

Demand signal processing refers to the magnification in variance that occurs through the interaction between forecasting procedures and inventory rules at each stage of the supply chain (see also Chen et al. 2000a, 2000b; Wong et al. 2007; for a review of studies in this area please refer to Syntetos et al. 2009a).

Lee et al. (1997b) argued that *rationing and shortage gaming* is a major cause of the bullwhip effect and occurs in situations where the demand exceeds the production capacity. In these situations, the manufacturer may ration or allocate supplies to the retailers. On recognising the rationing criteria, the retailer may place orders exceeding the required quantity, to secure a greater share of the supplies from the manufacturer. This gives the manufacturer a false impression of the true demand and they in turn place large orders on their suppliers. This particular cause of the bullwhip effect has also been discussed, amongst others, by Cachon and Lariviere (1999), Cheung and Zhang (1999) and Paik and Bagchi (2007).

A common practice in industry is not to place orders on the upstream link as soon as demand arises. Instead, the individual demands are batched or accumulated before placing the orders (*order batching*) and thus, instead of frequent orders, weekly, biweekly or monthly orders are placed. This is done for various reasons including economies of scale and distribution efficiencies, or similar factors (see, for example, Pujawan 2004; Potter and Disney 2006).

Price fluctuations refer to the practice of offering products at reduced prices in order to stimulate demand (e.g., Gavirneni 2006; Reiner and Fichtinger 2009). For price-elastic products, when the price of an item changes, the customer demand will also change. Customers buy in bulk quantities when the price of the product is low. Then, customers stop

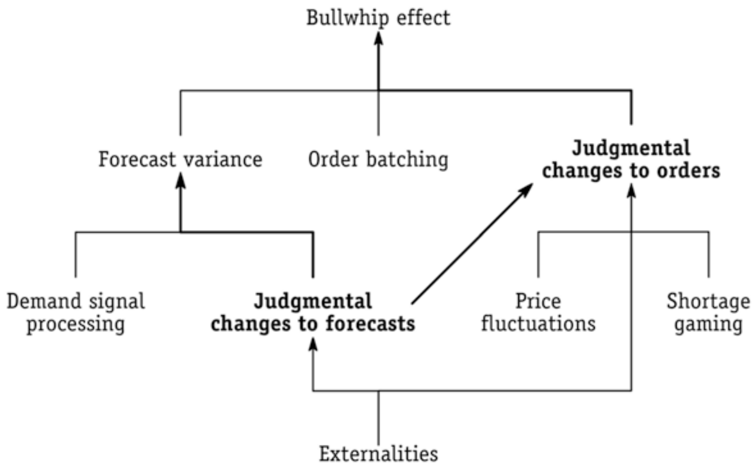


Fig. 1 Factors impinging on the bullwhip effect

buying when the price returns to normal, until they have exhausted their inventory. Thus, the actual customer sales do not match the true demand for the product when there are price variations. This results in the bullwhip effect, as the variance of the order quantities amplifies upstream because of the temporary price reductions.

There have been numerous papers investigating these causes individually, but little attention has been paid to their interaction. Such interactions may be adequately captured with the help of the SD modelling method. A critical element in any bullwhip effect model is the variance of upstream orders. Figure 1 shows a basic map of the most important factors that influence the bullwhip effect.

This example of a generic or archetypal causal map (Fig. 1) highlights two behavioural factors: (i) judgemental changes to forecasts and (ii) judgemental changes to orders. The second may be distinct from the first, as it is not necessarily based on any expectation of a changed demand pattern, but rather on a subjective reaction to external stimuli (e.g., change in price, shortage of supply). However, the former may indeed affect the latter if a single person, for example, performs both adjustments. Similarly, in the context of a small organization, the judgemental adjustments of the orders may reflect a certain reaction mechanism on the part of the stock controller to known adjustment behaviours on the forecasting side. The distinction between forecasting and ordering adjustments has been neglected in much of the academic forecasting literature, but is very important in practice (Kolassa et al. 2008).

System Dynamics Modelling of Supply Chains

The initial exposition of this application topic in SD was made by Forrester (1958, 1961). He considered a three-stage supply chain, demonstrating the now well-known amplification of orders as they are transmitted upstream. He explicitly considered the inventory policy as a cause of this. Orders were determined through a policy which considered: (desired − actual inventory) + (desired − actual orders in the pipeline) + (actual − normal unfilled orders). To preserve the dimensional consistency, each of these terms has to be divided by an adjustment time which Forrester emphasized was a critical explanatory factor in the determination of the overall system behaviour. The 'desired' values were obtained by smoothing an actual demand flow and then multiplying by a number of units of coverage.

Barlas and Ozevin (2004) investigated two different yet related research questions about stock management in feedback environments. The first purpose was to analyse the effects of selected experimental factors on the performances of subjects (players) in a stock management simulation game. In the light of these results, the second objective was to evaluate the adequacy of standard decision rules typically used in dynamic stock management models and to seek improved formulations. In the first part, gaming experiments were designed to test the effects of three factors on decision-making behaviour: different patterns of customer demand, minimum possible order decision (review) interval and, finally, the type of receiving delay. These factors were analysed with regards to their effect on 10 different measures of behaviour (such as max–min range of orders, inventory amplitudes, periods of oscillations and backlog durations).

In the second phase of research, the performances of subjects were compared against some selected ordering heuristics (formulations). These included the linear 'anchoring and adjustment rule' (commonly used in

SD studies), several alternative non-linear rules, and some standard discrete inventory control rules common in the inventory management literature. The non-linear and/or discrete rules, compared with the linear stock adjustment rule, were found to be more representative of the subjects' ordering behaviour in many cases, in the sense that these rules could generate nonlinear and/or discrete ordering behaviours. Another important finding was that the well-documented oscillatory dynamic behaviour of the inventory is a quite general result, not just an artefact of the linear anchor and adjust rule.

Saeed (2009) examined the use of trend forecasting in driving ordering policies in supply chains by comparing it with derivative control in classical control theory. He found that although both processes involve the use of trend to determine policies for achieving reliable performance, the former often worsens instability while the latter can improve stability with certainty. The similarities and differences between the two processes were discussed and a framework was suggested for improving the efficacy of trend forecasting in ordering policies.

Yasarcan and Barlas (2005) proposed a generalized SD stock control formulation for stock management problems involving information delays and delays implicit in controlling a primary stock indirectly via a secondary stock. It is well accepted that the behaviour of a standard SD stock management structure can be highly oscillatory if the stock control formulation (typically the linear anchoring and adjustment rule) does not take into account the supply line (material) delays. However, such delays do not constitute the only type of delay in stock management problems; there are other types such as information delays in decision processing, delays caused in trying to control a stock indirectly via a secondary stock and combinations of these. Yasarcan and Barlas (op. cit.) investigated the implications of ignoring such composite and indirect delays in the stock control formulation. They showed that the consequences of ignoring information delay in the decision stream or ignoring the delay implicit in secondary stock control are both equivalent to ignoring the supply line delay in the standard case: large, possibly unstable, oscillations. Subsequently, they proposed a generalized stock control heuristic that does take into account these more advanced types of delays and showed that the result is stabilized dynamic behaviour. In this research, they introduced the notion of a 'virtual supply line' (VSL), a conceptual generalization of the standard notion of supply line delay to structures involving information delays and 'secondary stock control-induced-delays'. They implemented their generalized decision heuristic on a complex example involving all three types of delays, demonstrating the usefulness of the proposed formulation whilst illuminating some implementation issues.

Croson and Donohue (2005) examined whether giving supply chain partners access to downstream inventory information is more effective at reducing bullwhip behaviour, and its associated costs, than similar access to upstream inventory information. They used a controlled version of the *Beer Distribution Game* as the setting for their experiment, and varied the amount and location of inventory information shared across treatments. They first independently tested whether sharing upstream or downstream inventory information helps reduce bullwhip behaviour, and found that only downstream information sharing leads to significantly lower order oscillations throughout the supply chain. Subsequently, they compared the reduction in order oscillations experienced by supply chain level and found that upstream supply chain members benefit the most from downstream information sharing. A very important observation of this study is that it is not the information *per se* but the interaction between the information and the decision setting that has the potential to improve performance in dynamic tasks.

For some very interesting discussions on the role of system dynamics in supply chain management and its usefulness for performance measurement in such a context, the reader is referred to Akkermans and Dellaert (2005), Kleijnen and Smits (2003), and Otto and Kotzab (2003).

The brief literature review on SD in supply chains presented in this subsection shows that only a small part of the generic map shown in Fig. 1 has been addressed using a system dynamics approach. In this paper, we go further and incorporate judgemental changes to orders and to forecasts into an SD model of a three-stage supply chain. We will not examine shortage gaming or price fluctuations. Order batching will not be analysed directly, although batching effects will be observed in one of the inventory management rules. It is hoped that the work reported here will provide a springboard for more comprehensive SD models of bullwhip effects.

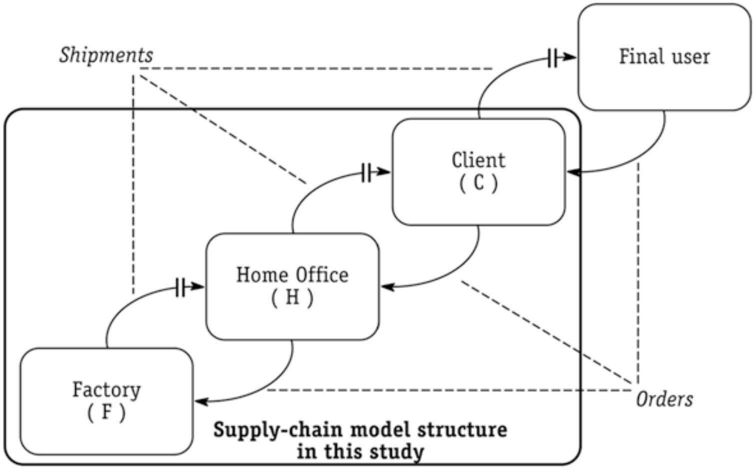


Fig. 2 Supply-chain model structure in this study

Experimental Structure

Excluding the final user stage, a three-stage supply chain has been considered for the purposes of this investigation (Fig. 2). It consists of a factory (*F*) or supplier stage, a home office (*H*) stage, where *H* takes the role of wholesaler, and a client (*C*) stage that acts as a retailer, serving the final end consumers. Each stage contains three sub-stages (called Work-In-Progress, WIP), which may represent such processes as booking in and inspection.

Our SD models show the middle stage *H* (wholesaler) as the home office or Head-Quarters of a vertically integrated supply chain or, correspondingly, the 'co-ordinating' unit (in terms of material and information flows) of any supply chain that consists of separate business entities. Consequently, the retailer may be viewed as the wholesaler's client (*C*) that in turn serves customers in the final user stage.

The demand experienced at the client (retailing) stage is assumed to be deterministic and constant at 100 items per week. The time unit chosen for all experimental results equals 1 (one) week. Other time buckets could have been chosen but weekly periods constitute a realistic reflection of various industries and their inventory control systems.

With regards to the demand, we considered the possibility of introducing some random variations to the underlying pattern, but that would necessitate conducting multiple runs (30 at least) for each experimental scenario for the purpose of averaging the results, leading to a totally unrealistic size of the simulation output. Likewise, a deterministic sine wave could have been introduced. However, the judgemental adjustment effects would not then be distinguishable from the deterministic sine wave effects on the observed bullwhip. On the other hand, the flat deterministic demand pattern renders the application of any forecasting method, such as single exponential smoothing (SES), redundant and thus forecasting adjustments cannot be evaluated. To overcome this problem, demand is indeed assumed to be forecasted using SES, but the forecast adjustment intervention occurs on the input of the SES procedure rather than on its output, that is, actual demand data are adjusted in lieu of the forecast output. We return to this issue later in this section.

Results have been generated for three stock control policies: (i) the linear Anchor and Adjust (*AaA*); (ii) the re-order point *s*, order-up-to-level *S* (*s*, *S*); and (iii) the order-up-to- level *S* (*utS*). Details governing the implementation of these three stock control policies follow in the next section. We assume that unfilled demand is backlogged, that is, no lost sales occur. All stages in the supply chain are assumed to employ the same stock control policy. We recognize that this is a rather restrictive assumption, but computational considerations dictate that we implement it in our SD supply-chain models. In addition, such a formulation would indeed reflect reality for a vertically integrated supply chain (or a supply chain connected through an ERP-type solution that dictates the employment of the same stock control policy through its functionality; in the case of SAP this would be the re-order point *s*, order-up-to-level *S* policy). The lead time has been set to three time periods (weeks). This is a convenient assumption based on the real-life organizations we have worked with. Lastly, the SES smoothing constant (α) has been set equal to 0.2, again something that conforms to generally accepted practices among forecasting practitioners.

The point(s) of intervention is also introduced as a control parameter in our modelling exercise. There are three stages in the system where adjustments can take place (Retailer, Wholesaler and Factory). We explore all possible single interventions (in only one of the three stages), dual interventions (in any two of the three stages) and finally a triple intervention (adjustments take place in every stage of the supply chain). This is in conjunction with the scenarios of: (i) adjusting only the forecasts; (ii) adjusting only the orders; and (iii) adjusting both forecast and orders.

The type of adjustment is the last control parameter. Each simulation runs for 100 periods and adjustments are introduced periodically (in periods 8, 24 and 40) with a magnitude of 25%. Fildes et al. (2009) analysed more than 60,000 forecasts (along with their adjustments) from four supply chain companies. The median adjustments reported in this study, across the four datasets, in terms of their magnitude in relation to the forecasts, varied between 13 and 33% (depending also on their direction). Subsets of these datasets were also analysed by Syntetos et al. (2009b, 2010). Both studies indicate that 25% constitutes a reasonable descriptive summarization of the relative magnitude of the adjustments. Finally, the selection of this control parameter value has been also confirmed as a realistic one by the organizations partnering with us in this project. In one case (Electronics Manufacturer) it was disclosed that when human intervention is exercised on the size of the replenishment orders, the relative magnitude of the adjustments is always (as a rule of thumb) 25%, regardless of the direction of the adjustment (plus or minus).

The optimism bias discussed in the previous section is reflected by consistent positive (upwards) adjustments introduced in our experiment. Inconsistent behaviour on the part of the managers with no sound justification as to why they are adjusting is reflected by a structure of alternate positive and negative adjustments. For completeness, consistent negative adjustments are considered as well. In all cases, the adjustments are unwarranted, as there is no change in the underlying demand pattern, allowing an assessment of the effect of un-necessary forecast adjustments on the supply chain.

Returning to the mechanism employed for forecast adjustment purposes, a realistic representation of this intervention would be to adjust the SES forecast by 25%. Instead, we adjust the demand input to the SES forecast by 25%. Although this may seem unnatural, there is little difference, for simulation purposes, between adjusting forecasts in the proposed way and intervening in a manner compatible with real world practices.

As discussed above, the SD models run for 100 periods and the interventions are introduced periodically in periods 8, 24 and 40. The different adjustment mechanisms introduce pulses, at the designated intervention points to the actual demand, with no change to the forecast. Subsequently, the actual demand remains unchanged (with no pulse), and the forecast increases upwards as a pulse, and then declines exponentially. This closely resembles the behaviour of a judgementally adjusted forecast which subsequently declines to adjust to the true level of demand. Most importantly, in this research we are concerned with the effects of such interventions on stock amplification. Consider the first intervention point as an example: Fig. 3 shows a typical response of the SES procedure (under the *AaA* stock policy) to the judgemental adjustment introduced in period (week) 8 along with the response of the stock held at the factory. The maximum amplification at the factory occurs in week 15 which is time distanced enough to justify that had we adjusted the SES forecast output itself, the difference would have been imperceptible.

In addition, the adjustment mechanism used for the purposes of this research introduces pulses that are always of lower magnitude that than those resulting from adjusting the forecasts directly, because the change in actual demand is immediately smoothed. Thus, we are probably

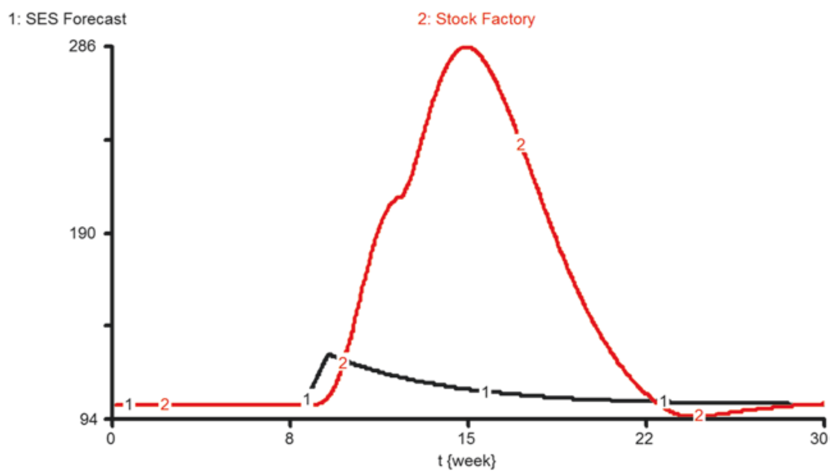


Fig. 3 SES and factory stock response to adjustments

conservative with regards to the potential consequences we simulate for the whole supply chain. Finally, had the adjustment been of a lower percentage (say 5%), the difference between the adjustment mechanisms would have been minor even in the intervention points. With regards to the replenishment orders, these are directly adjusted (by 25%) upon their generation by any of the policies considered and in any of the SC stages.

Taking into account all these concerns, our final experimental design enables us to explore various combinations of forecast and order adjustments in terms of their plus or minus direction, thereby accounting for multiple scenarios of optimistic and pessimistic adjustments. It is likely, in large organizations, that the number of process stages between the forecast intervention point and the order intervention point is greater than in a small firm. Thus, the current structure allows simulating varying combinations of forecast and order adjustments, that is, in terms of individuals' optimism and pessimism.

The experimental conditions considered for the purposes of our investigation are summarized in the Table 1. Owing to the very high number of experimental conditions, the gradual effect of adjusting forecasts and

Table 1 Control parameters and experimental scenarios

Control parameter	Experimental scenarios	#
Demand pattern	Deterministic, Constant	1
Forecasting	Single Exponential Smoothing (*SES*)	1
Inventory policies	Anchor and Adjust (*AaA*)	
	(*s*, *S*) up to S (*utS*)	3
Intervention	Adjust forecasts	
	Adjust orders	3
	Adjust forecasts and orders	
Points of intervention	Stage 1: client (*C*)	
	Stage 2: home (*H*)office	
	Stage 3: factory (*F*)	
	Stages 1 and 2: *C* and *H*	7
	Stages 1 and 3: *C* and *F*	
	Stages 2 and 3: *H* and *F*	
	Stages 1, 2 and 3: *C*, *H* and *F*	
Optimism (*O*) and Pessimism (*P*)	Persistent pessimism: $P \cdot P \cdot P$	
	Persistent optimism: $O \cdot O \cdot O$	
	Mixed: $P \cdot O \cdot P$	4
	Mixed: $O \cdot P \cdot O$	

orders will only be reported for the Factory stock, as this is typically the stage exhibiting the greatest amplification in the whole supply chain. Also, due to the considerable amount of the output of our investigation it is natural that only some results may be presented here. However, an electronic companion has been introduced to our paper that may be accessed at: http://www.business.salford.ac.uk/research/ommss/projects/SD/. This contains a more comprehensive selection of simulation outputs. The entire exercise has been performed using the *iThink® Software* (Richmond 2009).

Model Description

The subsystem diagram of Fig. 2 shows a typical supply- chain (SC), with the SD model structures in this article dynamically interconnected through material flows (i.e., shipments) and bundled information connectors (i.e., orders). Even small SD modelling examples, such as the one in Fig. 2, show the interdependencies among variables connected through multiple feedback loops, which typically contain time lags and delays.

In order to show the different causal structure behind each stock policy used in this study, this section presents only the middle, home office (*H*) stage (Fig. 2), model sector under each stock policy. In all three cases, the causal structure behind the client (*C,* Sector 1: not shown) and factory (*F,* Sector 3: not shown) supply-chain stages is almost identical to the one of the home office (*H*) stage (Sectors 2a–2c, Fig. 4). The only difference is in the factory (*F*) stage, which has to account for its own backlog only, thereby excluding a prior-stage backlog, simply because, under each stock control policy considered, the factory stage is assumed to be the very beginning of the entire chain.

Figure 4 shows the stock and flow diagram of the home (*H*) supply-chain model sector, under the anchor and adjust (*AaA*) stock policy. There is a one-to-one correspondence between the model diagram in Fig. 4 and its equations (Table 2, see Appendix). Building the model entailed first drawing the model structure and then specifying simple algebraic equations and parameter values. The *iThink® Software* enforces consistency between diagrams and equations, while its built-in functions help quantify parameters and variables pertinent to the causal structure of each stock control policy. As discussed in the previous section, we assume that all stages use the same stock policy consistently within their supply chain.

In the SD modelling method, rectangles represent stocks or level variables that can accumulate, such as the Work- In-Progress (*WIP*) *A Home* and *Stock Home* stocks (top of Fig. 4 and Eqs. (1) and (2), Table 2). Emanating from cloud-like sources and ebbing into cloud-like sinks, the double-line pipe-and-valve-like icons that fill and deplete the stocks represent flows or rate variables that cause the stocks to change. The *to stock home* flow (Eq. (7)), for example, at once feeds the *Stock Home* stock and also depletes the *WIP C Home* stock, modulated by the *WIP C Home* level (Eq. (4)) and the *lead time home* (Eq. (15)). Singleline arrows represent information connectors, while circular or plain text icons depict auxiliary converters where behavioural relations, constants or decision points convert information into decisions. The *coverage home* auxiliary converter, for example, depends on both the *lead time home* converter and the *Est D* (estimated demand) *Home* stock (Eq. (23)).

Supply chains always entail a stock and flow structure like the one on the top panel of Fig. 4 for the acquisition, storage and conversion of inputs into outputs, and decision rules (middle of Fig. 4) governing the flows.

The causal structure of auxiliary variables and parameters in the middle of Fig. 4 shows the decision rules pertinent to the *AaA* stock control policy. This set of rules first 'anchors' the *order from factory* decision (Eq. (19)) on the *adjust stock home* converter (Eq. (11)) and then adjusts according to the *adjust SC home* (Eq. (10)). The order decision also takes into account the *Est D Home* stock (Eq. (23)), which essentially is the output of a SES forecasting procedure with $\alpha = 0.2$ (Eqs. (25)–(28)).

Figure 5 shows the stock and flow diagram and Table 3 (see Appendix) the corresponding equations of the home (*H*) supply-chain model sectors, under the (*s*, *S*) stock control policy. Both the basic stock and flow structure in the top panel of Fig. 5 and the corresponding equations in Table 3 are identical with the *AaA* stock policy model (Fig. 4). So are the causal structure and corresponding equations of the *home backlog* and *home forecast* sectors in Fig. 5 and Table 3. The only difference between

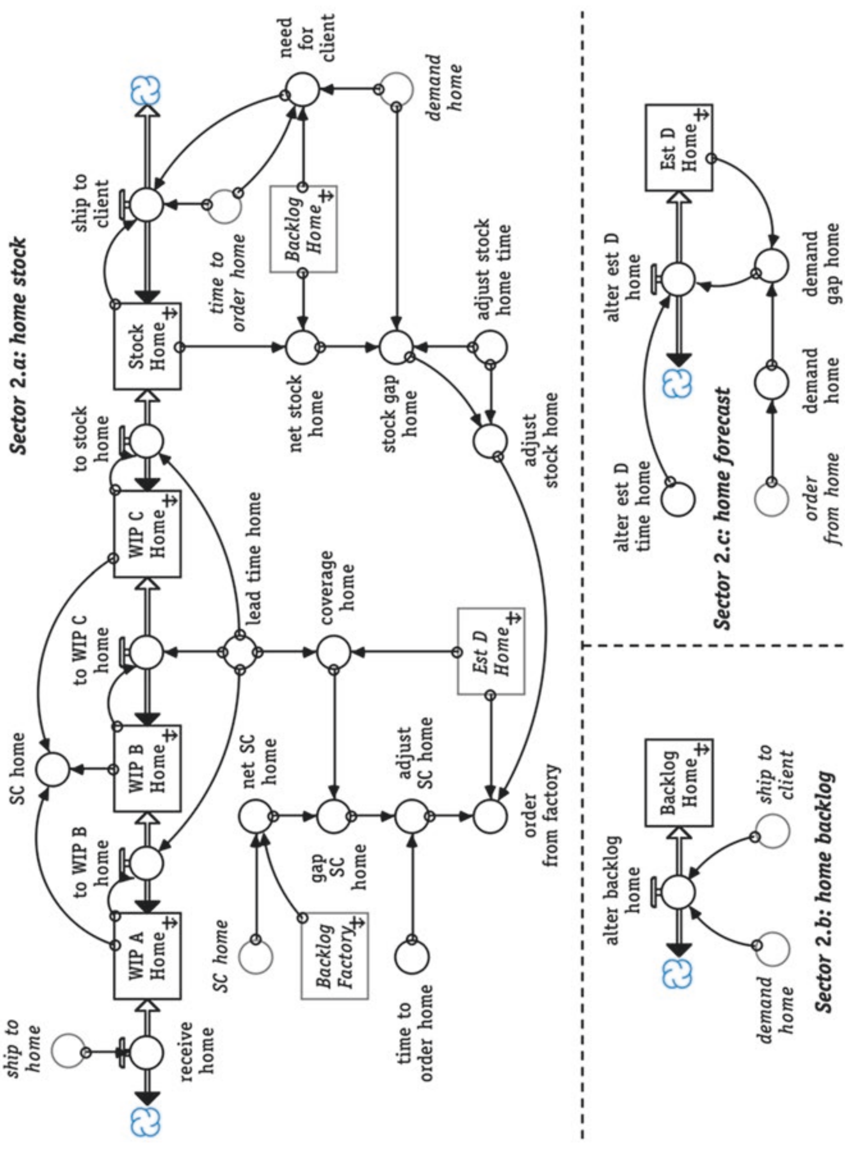


Fig. 4 Stock and flow diagram of the home (*H*) supply-chain model sectors, AaA

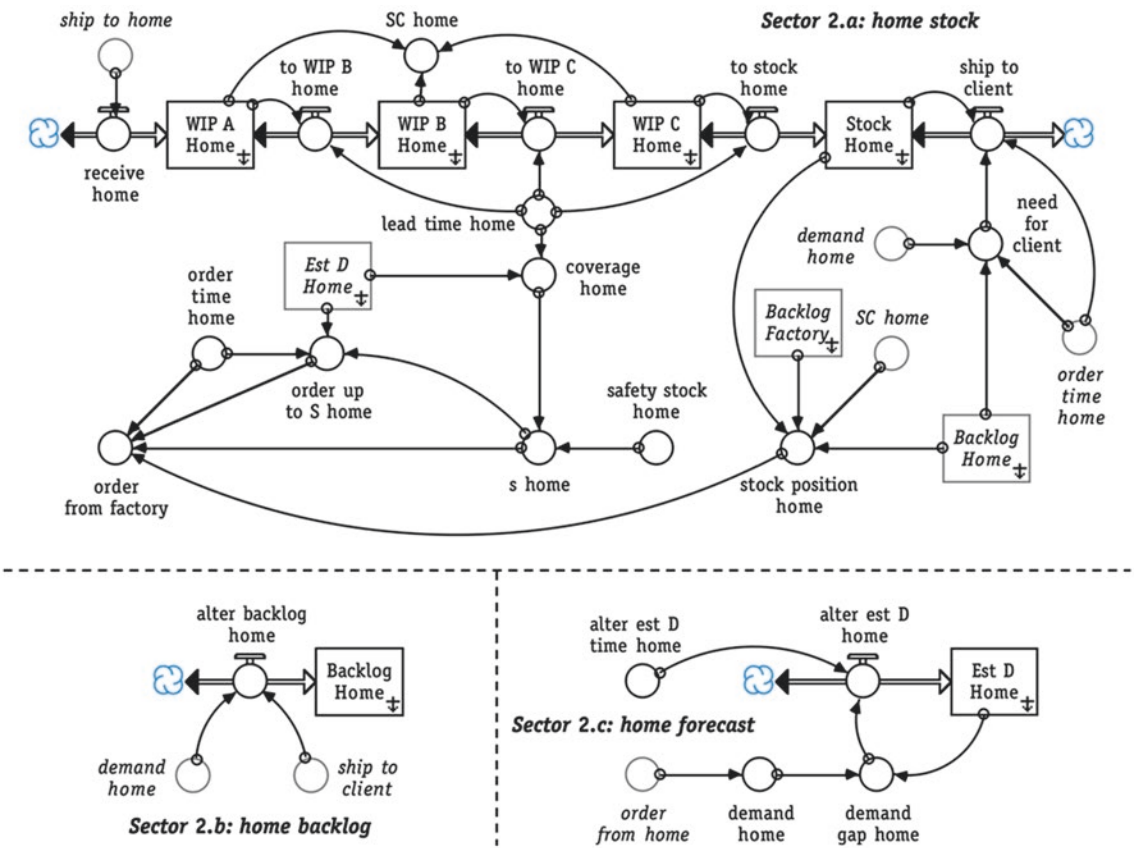


Fig. 5 Stock and flow diagram of the home (*H*) supply-chain model sectors, (*s*, *S*)

the two models concerns the (*s*, *S*) decision rules governing the stocks and flows of the supply chain in the top panel of Fig. 5.

The middle right of Fig. 5 now shows the *stock position home* converter (Eq. (39)), which takes into account the *Stock Home* level, plus the *SC home* and its supplier's *Backlog Factory*, but it must subtract its own *Backlog Home* level of orders already placed but not yet satisfied to the downstream client (*C*) stage. The middle and left of Fig. 5 shows the variables and auxiliary constants that drive the (*s*, *S*) based *order from factory* decision converter (Eq. (33)). In addition to the *stock position home*, the decision also depends on the (re-order point) *s home* converter (Eq. (36)) as well as the *order up to S home* converter (Eq. (35)) and the *order time home* (Eq. (34)).

As the middle left of Fig. 5 shows, the *Est D Home* SES forecast output and the *order from factory* decision point are closely positioned. The large *safety stock home* (Eq. (37)) guards against large backlogs and renders the (*s*, *S*) model robust to large bullwhip effects. Also, it helps to easily initialize the model at steady state, a condition crucial if the computed simulation results are to make sense. Unless a SD model is initialized at steady state, then one might simply observe artefact dynamics attributable to initial transient conditions.

Figure 6 shows the stock and flow diagram and Table 4 (see Appendix) the corresponding equations of the home (*H*) supply-chain model sectors, under the order up to S (*utS*) stock policy. Both the basic stock and flow structure on the top panel of Fig. 6 and the corresponding equations in Table 4 are again identical with the *AaA* stock policy model (Fig. 4). So are the causal structure and related equations of the *home backlog* and *home forecast* sectors in Fig. 6 and Table 4. Once more, the only difference between the two models lies in the middle of Fig. 6, where the *utS* decision rules govern the stocks and flows of the supply chain.

As under the (*s*, *S*) policy, in the supply-chain rules on the middle of Fig. 6, the *stock position home* converter (Eq. (47), Table 4) takes into account the *Stock Home* level, plus the *SC home* and its supplier's *Backlog Factory*, but it must once more subtract its own *Backlog Home* level of orders already placed but not yet satisfied to the client (*C*) downstream. The middle of Fig. 6 shows the variables and auxiliary constants that drive the *utS*-based *order from factory* decision converter (Eq. (44)). In

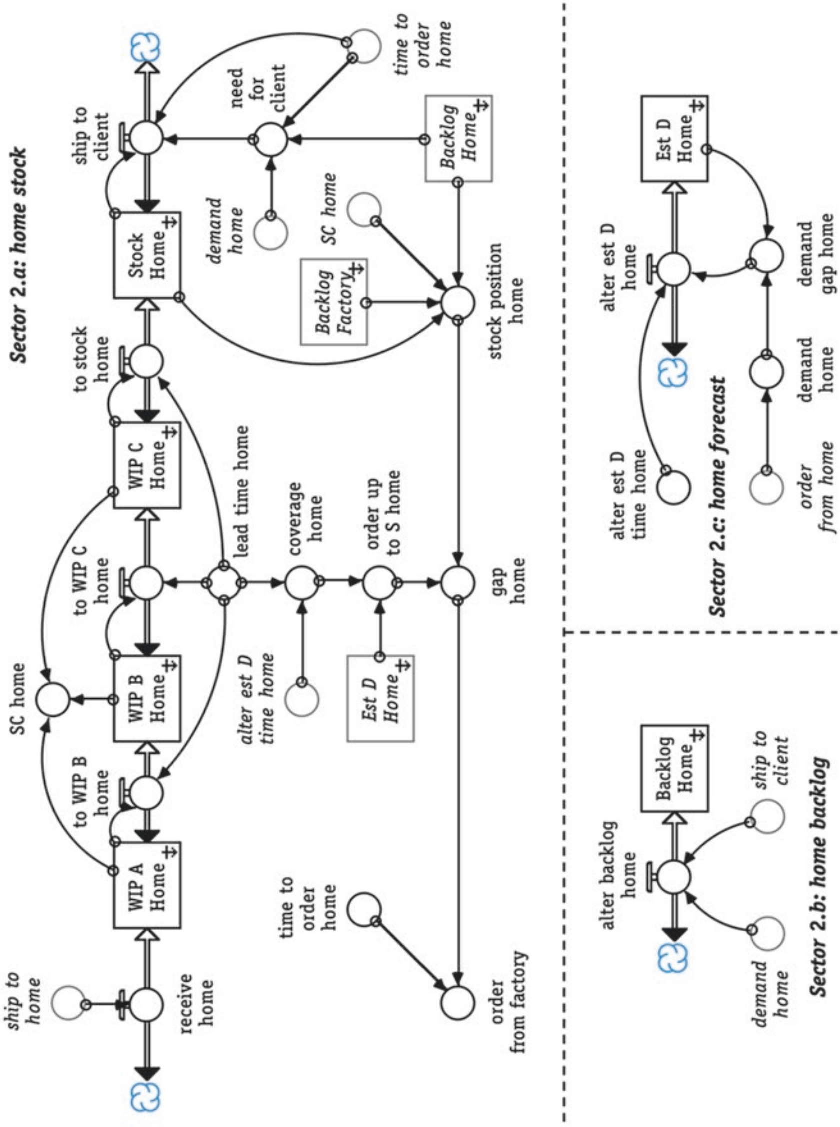


Fig. 6 Stock and flow diagram of the home (*H*) supply-chain model sectors, *utS*

addition to the *stock position home*, the ordering decision also depends on the *order up to S home* converter (Eq. (45)) as well as the *time to order home* auxiliary constant (Eq. (48)).

An important difference between the *utS* and the other two stock control policies used in this study has to do with the unit specification of the *coverage home* converter Eqs. ((13), (30) and (40)). Under the first two stock policies, *AaA* and (*s*, *S*), Eqs. (13) and (30) have volume units, while for the *utS* policy Eq. (40) has time units. Amplifying the time inventory coverage under the *utS* stock control policy helps to easily initialize the corresponding SD model at steady state; again a point utterly crucial to the validity of the simulation results.

Performance Measurement

The assessment of supply chain performance takes place through the factory stock amplification ratio (*FSAR*). Figure 7 shows the stock and flow diagram and Table 5 (see Appendix) the corresponding equations of the SD model sector that computes the *FSAR*, in order to assess the independent and joint effects of judgemental forecast and order adjustments on supply chain performance.

Sterman (2000) defines the amplification ratio as '*the ratio of the maximum change in the output to the maximum change in the input* (p. 673)'. Accordingly, the *Stock Factory* Δ stock on Fig. 7 (Eq. (49), Table 5) accumulates the maximum change in the *Stock Factory* via the *add* Δ flow (Eq. (50)). This flow feeds the *Stock Factory* Δ stock, incrementally, only when judgmental forecast and/or order adjustments cause *Stock Factory* to change and to reach a level higher than its previous one. The *factory stock* % Δ converter (Eq. (52)) becomes the numerator of the *FSAR* (Eq. (53)). Its denominator is the auxiliary constant parameter *input % Δ* = 0.25 (Eq. (55)) because all the judgmental forecast and/or order adjustment interventions in this study, equally spaced in time, entail either a downward or an upward adjustment of a 25% change in the input, always independent from any previous change.

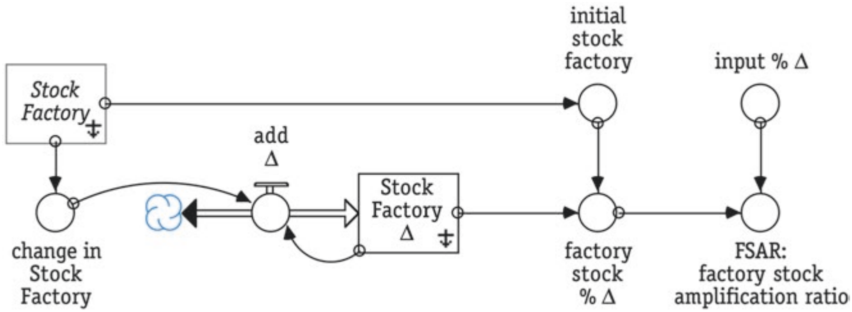


Fig. 7 Computing the factory stock amplification ratio (*FSAR*)

Analysis of Results

Our results have been summarized as follows: each *FSAR* figure shows five graphs that reflect different experimental conditions, also summarized within each figure. Within each figure, graphs *a* and *b* summarize the independent (within only one stage at a time: *Client*, *Home* and *Factory*) intervention effects on *FSAR*. Each line represents results for a different stock control policy, and each point represents a run with a particular type of intervention, details of which are presented on the right hand side of each graph. *Switch 1* indicates an alternate plus–minus adjustment pattern or an only positive adjustment pattern. *Switch −1* indicates an alternate minus–plus adjustment pattern or an only negative adjustment pattern. Similarly, graphs *c* and *d* show the 2-way interaction effects of judgemental interventions at: *Client & Home (C H*), *Client & Factory (C F), Home & Factory (H F).* On the *x-axis* the stages at which we intervene are preceded by a plus (+) or minus (−) sign, for the *switch 1* and *switch −1,* respectively, to indicate the type of adjustment that has been considered for the stage under concern. Finally, graph *e* shows the effects of intervening in all stages of the SC (with a similar notation used on the *x-axis*).

Figures on time domain results indicate, for the specified control parameter combinations, the inventory level (in units) at each of the supply chain stages considered (*Client*, *Home* and *Factory*) for all three stock control policies investigated in our experiment.

Overall, the results indicate that judgemental interventions have a considerable effect on supply chain performance. This is true for both forecast and order adjustments. The impact varies according to the intervention point in the supply chain.

Relatively speaking, the judgemental forecast adjustments seem to have the most prominent effects, while the judgemental order adjustments appear to have the least prominent effects on the *Factory Stock Amplification Ratio*. This implies that it would be difficult for managers to compensate for the effects of judgemental forecast adjustments through judgementally adjusting replenishment orders.

Figure 8a, b show the independent effects of judgemental forecast adjustments at the *C*, *H* and *F* stages. Irrespective of the mixed intervention pattern assumed, adjustments that take place at the *C* stage result in the highest *FSAR*. This is true for all stock control policies considered. Conversely, reducing the number of stages between the forecast adjustment intervention point and the *Factory* stock causes its amplification to decline. That is, the impact of the forecast adjustments is less prominent as the intervention point moves upstream in the supply chain, from *Client* to *Home* to *Factory*.

Figure 8c, d assess the effects of adjustments applied concurrently at two intervention points. The results are consistent with those discussed above: adjustments that take place at the *Client* stage have the greatest impact. The *AaA* policy appears to be the least sensitive to the location of the intervention points considered. This issue is further discussed later in this section.

The results shown in Fig. 8e, where forecast adjustments are performed at all three stages of the supply chain, indicate a rather variable behaviour of the stock control policies depending on the combination of optimistic and pessimistic adjustments at the various stages. However, they do demonstrate that, for each specific combination, the stock control policy affects the *FSAR*. We also further elaborate on this issue later on in this section.

The results presented in Fig. 9 confirm overall those discussed above with the only difference related to the response of the (*s*, *S*) stock policy, which is highly sensitive to pessimistic adjustments (especially at the *Client* stage). In Fig. 9e, as we move from persistent pessimism to persistent optimism at the *Client* stage, the *FSAR* drops dramatically. The implicit order batching design of this policy introduces a delay in the system's upwards response, so if the Factory stock has been depleted due to pessimistic forecast adjustments then it takes longer for this policy (as compared to the other two) to rebuild inventory.

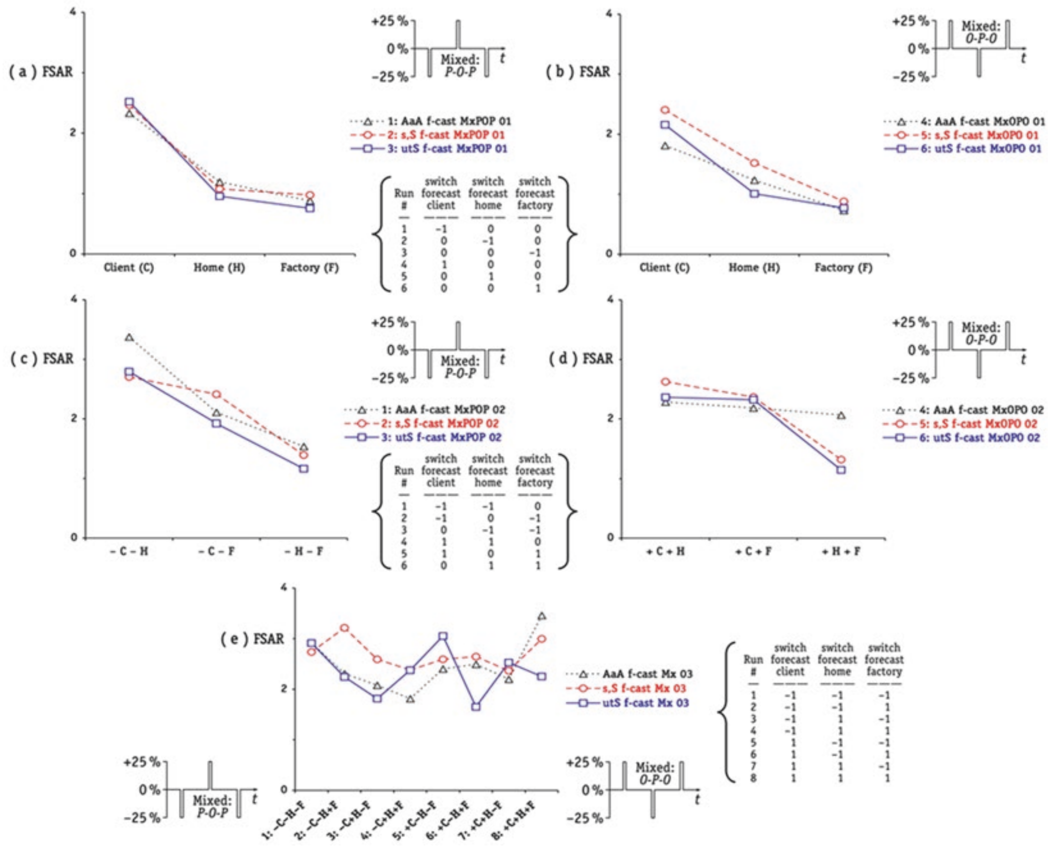


Fig. 8 *FSAR* results of client, home and factory sector forecast adjustments, with *mixed* optimism (*O*) and pessimism (*P*), under the *AaA*, (*s*, *S*) and *utS* stock policies

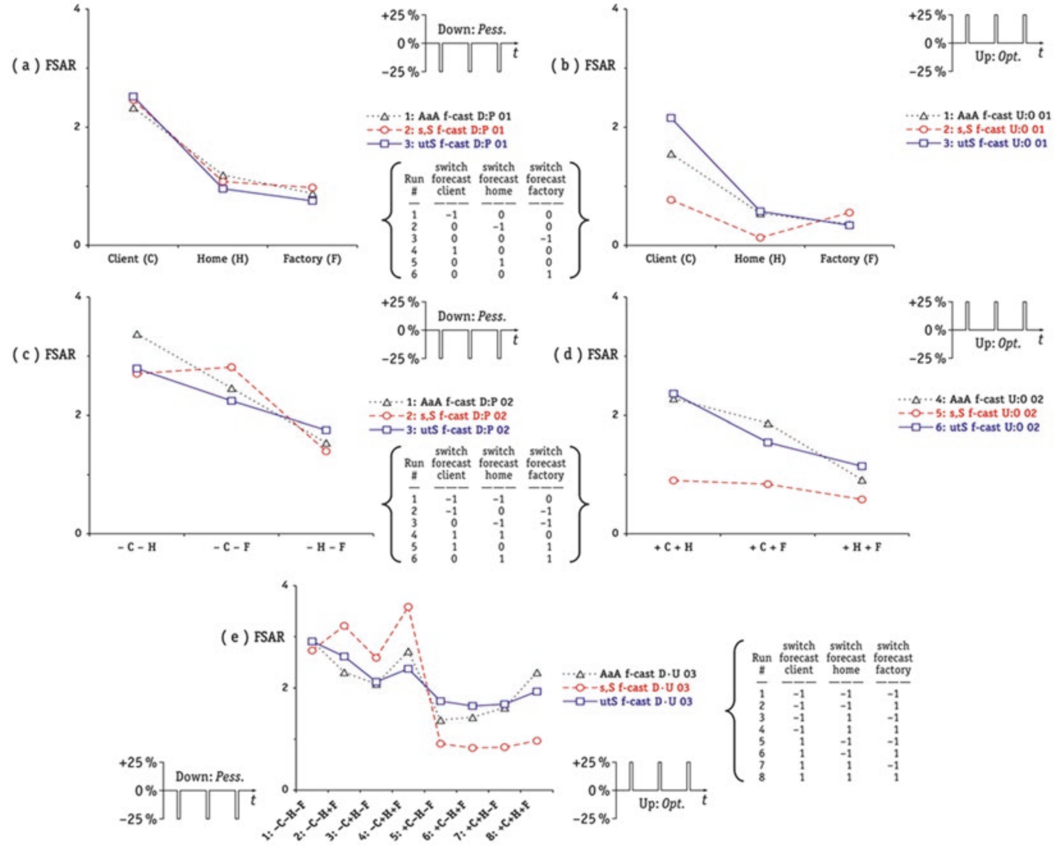


Fig. 9 *FSAR* results of client, home and factory sector forecast adjustments, with *persistent* optimism (*O*) and pessimism (*P*), under the *AaA*, (*s*, *S*) and *utS* stock policies

Figure 10 shows the results of the last run of Fig. 9e (persistent optimism reigns across all stages of the supply chain) for all three stock control policies. This figure demonstrates the reason we chose the *FSAR* as a proxy to assess the overall supply chain performance. In all three cases, the *Factory* stock clearly shows the highest amplification, irrespective of the underlying stock policy. The *Home* stock rates second in terms of amplitude while the *Client* stock appears as the most resistant to these persistent optimistic adjustments. In case of the *Home* stock (line no. 2) there are indications of phase-doubling particularly in the *AaA* and *utS* policies but this is somewhat subdued in the case of the (*s*, *S*) stock policy.

Figure 10b shows that the initial equilibrium of the (s, S) stock policy driven system is an unstable one. This is very easy to understand if one looks at Eq. (33) (Table 3). The *if- then-else* structure of the *Order from Factory* converter creates sharp discontinuities that do not allow the system to return to its initial unstable equilibrium. In addition, Fig. 10a, c clearly shows that the highest *FSAR* emanates from the initial 'shock' that the first forecast adjustment introduces into the system. Subsequent judgemental forecast adjustments continue to create 'shocks' but their amplification amplitude is not as great as the initial one. This downward sloping trend that the *FSAR* shows in the time domain is not always followed, as demonstrated by other experimental results not shown in this paper. Our electronic companion present cases where subsequent adjustments may increase the factory stock amplification progressively, thereby creating an upward trend in *FSAR*.

Figure 11 shows the effects of judgemental order adjustments under a mixed (Optimistic and Pessimistic) intervention mechanism. Graph 11e shows the resistance of the (*s*, *S*) policy to order adjustments across all mixed optimistic and pessimistic conditions. Conversely, the other two stock policies *(AaA* and *utS*) appear sensitive to all optimistic and to all pessimistic mixed order adjustments and less sensitive to the scenarios that deviate from those extremes.

Figure 12 shows the interaction between optimistic forecast adjustments and persistent, either optimistic or pessimistic, order adjustments. The most important results are as follows: under conditions of optimistic forecasts in all sectors and persistent optimistic and pessimistic order adjustments the (*s*, *S*) inventory policy seems to be the least sensitive to

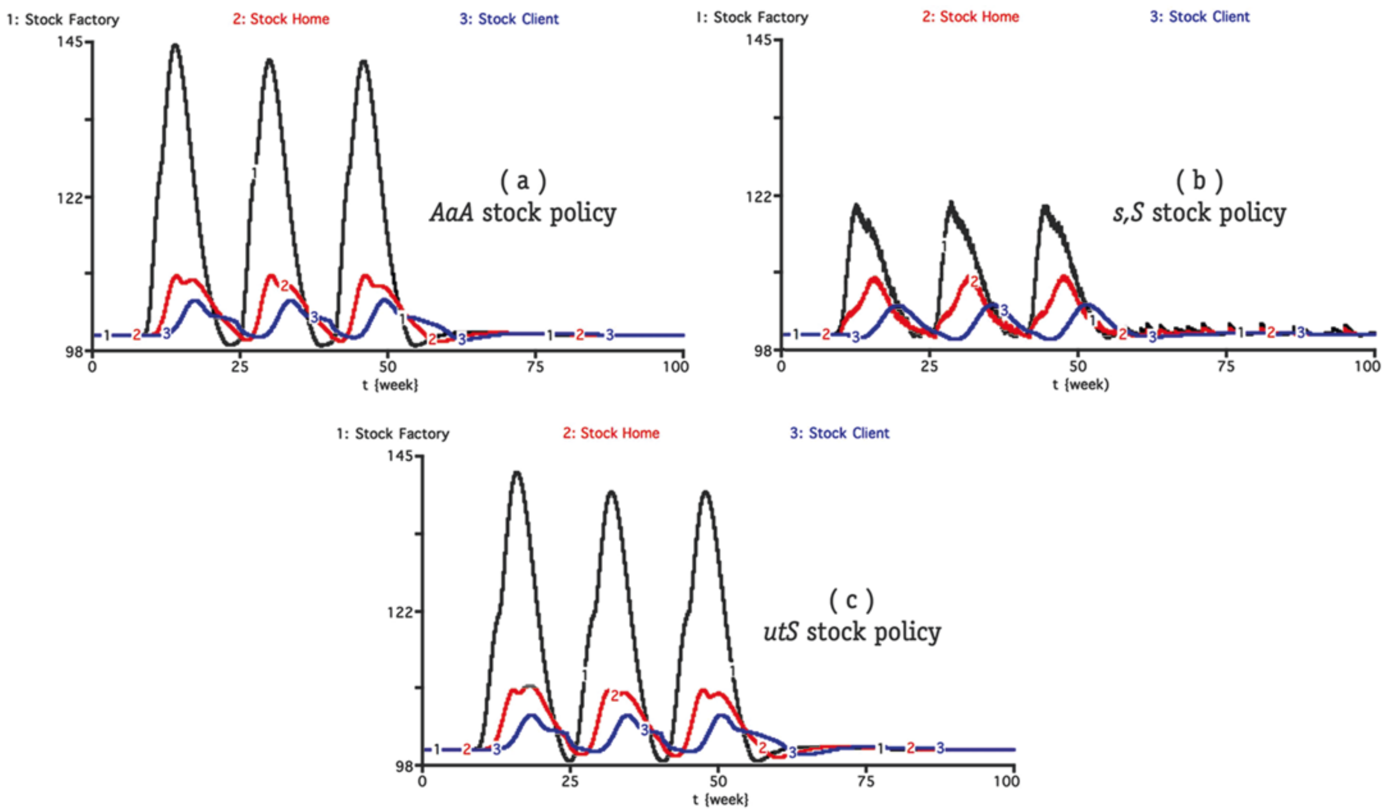


Fig. 10 The time-domain results of run #8 on Fig. 9e

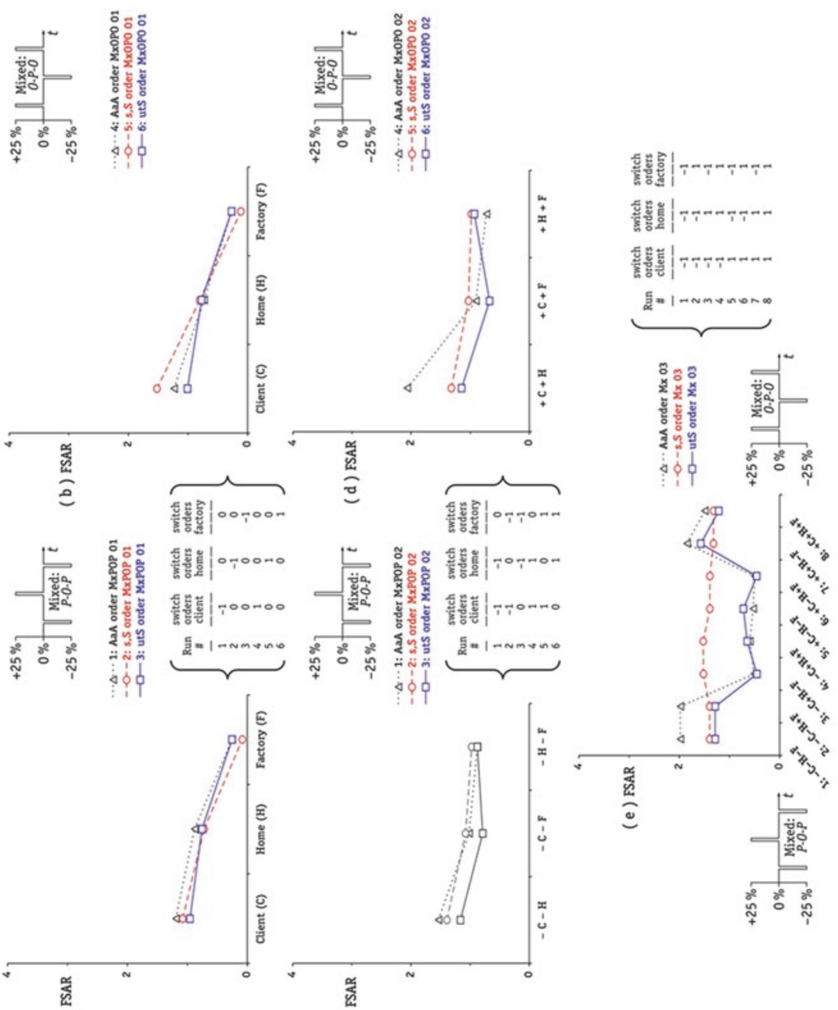


Fig. 11 *FSAR* results of client, home and factory sector *order* adjustments, with *mixed* optimism (*O*) and pessimism (*P*), under the *AaA*, (*s*, *S*) and *utS* stock policies

them. On all graphs of Fig. 12, the *FSAR* is the lowest under the (*s*, *S*) policy. Perhaps this is good news for SAP users since the inventory modules of the ERP package under concern are explicitly based on (*s*, *S*) policies.

In the time domain, the phase plot of Fig. 13d shows the relation between the *Factory Stock* and its autocorrelation (*r*), sampled once every 5 weeks. On the horizontal axis, *y* represents the *Factory Stock*. Although the axis is bounded on the lower side at 0, the scale has been extended only for presentational purposes. Phase plots hide the time dimension, but the little arrows on Fig. 13d show how the relation between *Factory Stock* and *r* evolves through time. The wildly oscillating *r* suggests that the highly interdependent variables in supply chains may be too ill-behaved to assess with mathematical models.

It is worth noting that, with regards to the phase-doubling component of the bullwhip effect, contrary to the results presented in Fig. 10, where the *Home* stock shows signs of phase doubling under the *AaA* and *utS* stock control policies, on Fig. 13 the Home stock amplification shows clear signs of phase doubling under the (*s*, *S*) stock policy.

Conclusions and Extensions

Pioneered by Forrester (1958, 1961) and influenced by engineering control theory, SD calls for formal simulation modelling that provides a rigorous understanding of system behaviour. Simulation modelling has become an essential research tool in social science because '*people's intuitive predictions about the dynamics of complex systems are systematically flawed* (Sterman 1994, p. 178)', mostly because of our bounded rationality. SD is a modelling method with high descriptive ability and theory building potential (Georgantzas 2001; Davis et al. 2007; Lane and Schwaninger 2008).

Adapted predominantly from Sterman (2000, Chaps. 17 and 18) and other colleagues, who model supply chains with SD (Georgantzas 2003, 2009; Barlas and Ozevin 2004; Yasarcan and Barlas 2005), the SD model sectors in this study explain the sources of oscillation, amplification and phase lag generally seen in client–supplier chains; phenomena which executives at 3M, Bristol-Myers Squibb, Hewlett-Packard and P&G

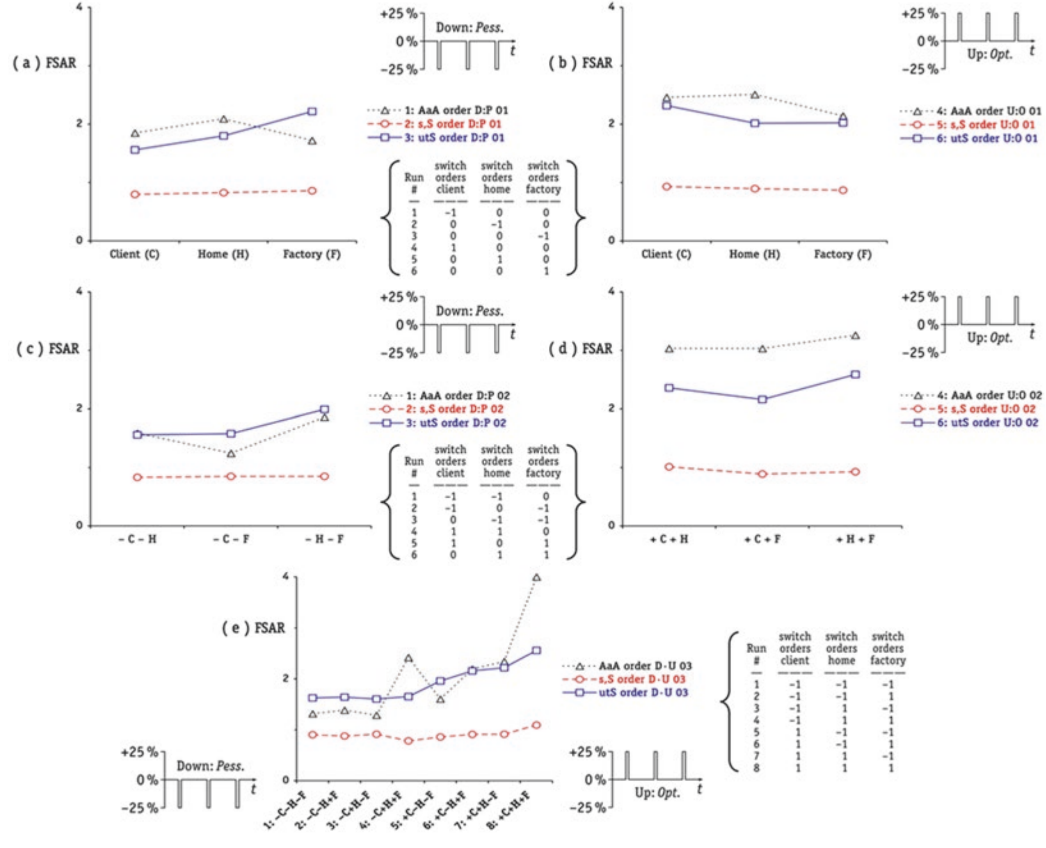


Fig. 12 *FSAR* results of the interaction among all-sector optimistic forecast and all-sector order adjustments with persistent optimism (O) and pessimism (P), under the AaA, (s, S) and utS stock policies

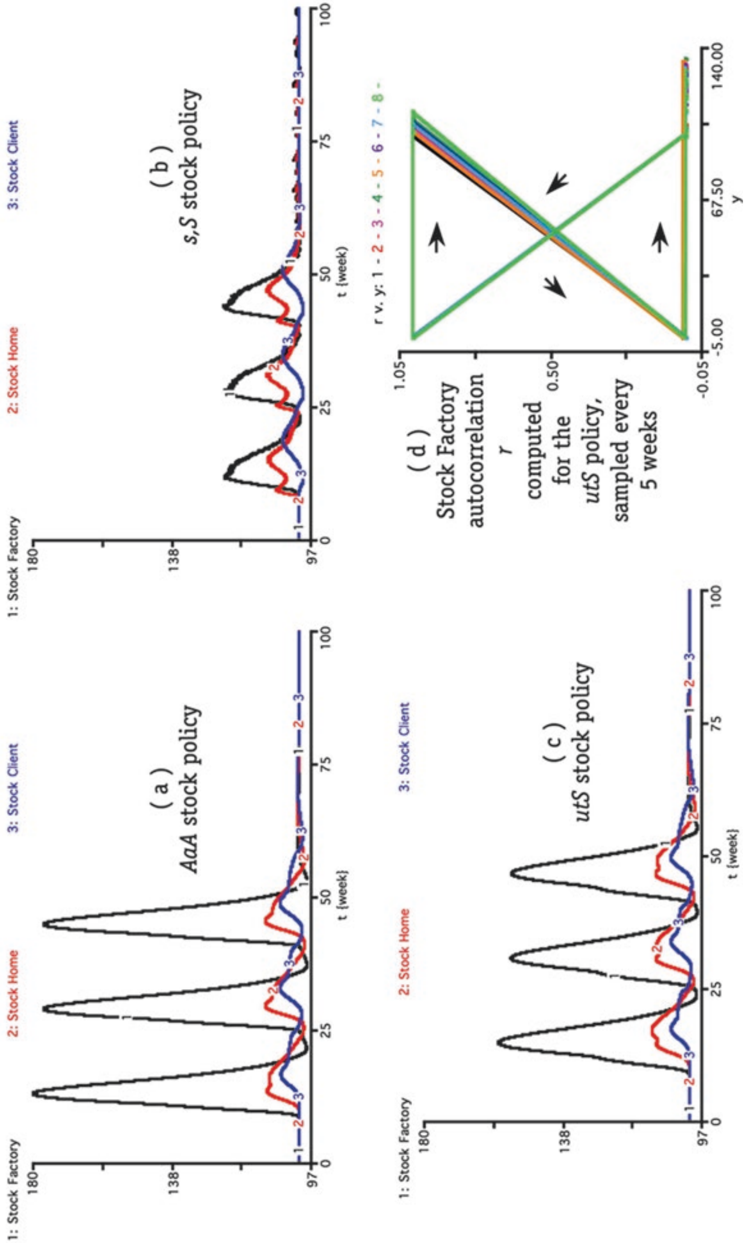


Fig. 13 The time-domain results of run #8 on Fig. 12e

collectively call the bullwhip effect (Lee et al. 1997b). Locally rational policies that create smooth and stable adjustment of individual business units can, through their interaction with other functions and firms, cause oscillation and instability, that is bullwhip-type dynamics. The models incorporate policy parameters pertinent to decision-making and timing that allow testing the sensitivity of client–supplier value chains to exogenous judgmental forecast and order adjustments. The results reveal policies that managers and their suppliers can use to improve performance.

Our study indicates that judgemental interventions may have a substantially adverse effect on supply chain performance if undertaken unnecessarily. Judgemental forecast adjustments have more prominent effects than judgemental order adjustments on the *Factory Stock Amplification Ratio*. To the best of our knowledge, this differential effect has not been reported previously in the literature. As discussed in the previous section, this finding implies that it may be more difficult for managers to compensate for the effects of judgemental forecast adjustments through judgementally adjusting replenishment orders.

Our investigation also shows that the impact of the forecast and order adjustments is less prominent as the intervention point moves upstream in the supply chain, from *Client* to *Home* to *Factory.* For the Anchor-and-Adjust (*AaA*) policy, this finding is consistent with results from system dynamics studies of multi-echelon inventory systems as far back as Forrester (1958). For the order-up-to-level S (*utS*) policy, the result is consistent with the amplification of variance through the supply chain, quantified by Lee et al. (2000), although his paper did not address the issue of judgemental adjustments to forecasts or orders. For the re-order point, order-up-to-level (*s*, *S*) policy, we have provided new results on the effect of adjustments on the *Factory Stock Amplification Ratio*.

Most previous papers on the bullwhip effect have focused on the order-up-to-level *S* policy. In this paper, we have shown that the (*s*, *S*) policy is less sensitive to order and forecast adjustments than the order-up-to *S* policy (*utS*). The only exception to this result is that (*s*, *S*) systems are more adversely affected by pessimistic forecast adjustments. In this case, the implicit order batching means that it takes longer for this system to rebuild inventory. The findings on the (*s*, *S*) system are important, as this policy is frequently adopted in practice.

Further Research

The research described in this paper constitutes an initial attempt to explore the effects of judgemental interventions on supply chain performance. Naturally, there are many avenues for further contributions in this area. In particular, the standardization of the magnitude of the adjustments to 25% is viewed as rather restrictive. We have attempted to capture a wide range of possible interventions in terms of the direction of the forecast and order adjustments and the point of intervention in the supply chain. However, the magnitude of such adjustments has not been introduced as a control parameter. Although some empirical justification has been offered to support our choice, further research should look at the effect that the size of the adjustments may have on supply chain dynamics. In addition, demand has been assumed to be deterministic and constant for the purposes of our study. Experimentation with variable demand and/or deterministic demand patterns that may be associated with ramp or step changes over time should offer valuable insights on the performance of the system. Moreover, we have assumed that each stage in the supply chain employs the same stock control policy. Perhaps under some circumstances a more realistic representation of the problem would involve a combination of such policies. Finally, other stock control policies could have been introduced as well. The re-order point s, order quantity Q (s, Q) policy for example would enable a more thorough investigation of the effects of order batching in conjunction with judgemental adjustments.

Future research must necessarily render both the judgemental forecast and the judgemental order adjustments endogenous. Depending on managers' own mood disposedness towards pessimism and/or optimism as well as that of the organization they are affiliated with or manage, they may respond differentially to an initial system 'shock'. Consequently, they may alter their subjective interventions on forecast and replenishment orders according to how initial 'shocks' shape their personalized organizational goals. System dynamics can play a crucial role in evaluating the impact of any learning effects (i.e., the adjustments get better over time) in the process of intervening with forecasts and/or orders. No evidence of such effects has been found through empirical studies and this constitutes a very promising area for further research.

Most importantly, our research focused on the implications of people making an adjustment when such adjustment is not needed. That is to say, in all the experimental conditions considered the adjustments are unwarranted, as there is no change in the underlying demand pattern. This allows for an assessment of the effect of un-necessary forecast adjustments on the supply chain. Although such a scenario is realistic and supported by empirical evidence, further research should also look at the effect of fully or partially warranted adjustments to forecasts and/or orders, in response to changes in the demand patterns or to other organizational factors.

Acknowledgements The research described in this paper has been funded by the Engineering and Physical Sciences Research Council (EPSRC, UK) Grant no. EP/G070369/1 (a project entitled: Cognitive Mapping, System Dynamics and the Bullwhip Effect). More information on this project may be obtained at: http://www.business.salford.ac.uk/research/ommss/projects/SD/.

Appendix: Model Equations

Supplementary information: Additional information pertinent to the modelling and results presented in this paper is available at: http://www.business.salford.ac.uk/research/ommss/projects/SD/.

Table 2 The home (*H*) supply-chain model sector equations, under the *AaA* stock policy

Sector 2.a: Home stock level (state) or stock variables {unit}	Eq. #
Stock Home(t) = Stock Home(t – dt) + (to stock home – ship to client) × dt	(1)
INIT Stock Home = demand home × time to order home {SKU}	(1.1)
WIP A Home(t) = WIP A Home(t – dt) + (receive home – to WIP B home) × dt	(2)
INIT WIP A Home = Stock Home {SKU}	(2.1)
WIP B Home(t) = WIP B Home(t – dt) + (to WIP B home – to WIP C home) × dt	(3)
INIT WIP B Home = Stock Home {SKU}	(3.1)
WIP C Home(t) = WIP C Home(t – dt) + (to WIP C home – to stock home) × dt	(4)

(*continued*)

Table 2 (continued)

Sector 2.a: Home stock level (state) or stock variables {unit}	Eq. #
INIT WIP C Home = Stock Home {SKU}	(4.1)
Sector 2.a: Home stock flow or rate variables {unit}	
Receive home = MAX (0, ship to home) {SKU/week}	(5)
Ship to client = MIN (need for client, Stock Home/time to order home) {SKU/week}	(6)
To stock home = MAX (0, WIP C Home/(lead time home/3)) {SKU/week}	(7)
To WIP B home = MAX (0, WIP A Home/(lead time home/3)) {SKU/week}	(8)
To WIP C home = MAX (0, WIP B Home/(lead time home/3)) {SKU/week}	(9)
Sector 2.a: *Home stock auxiliary or converter variables and constants* {unit}	
Adjust SC home = gap SC home/time to order home {SKU/week}	(10)
Adjust stock home = stock gap home/adjust stock home time {SKU/week}	(11)
Adjust stock home time = 1 {week}	(12)
Coverage home = lead time home × Est D Home {SKU}	(13)
Gap SC home = coverage home − net SC home {SKU}	(14)
Lead time home = 3 {week}	(15)
Need for client = demand home + (Backlog Home/time to order home) {SKU/week}	(16)
Net SC home = Backlog Factory + SC home {SKU}	(17)
Net stock home = Stock Home − Backlog Home {SKU}	(18)
Order from factory = MAX (0, (Est D Home + adjust stock home + adjust SC home)) {SKU/week}	(19)
SC home = WIP A Home + WIP B Home + WIP C Home {SKU}	(20)
Stock gap home = demand home-net stock home/adjust stock home time {SKU/week}	(21)
Time to order home = 1 {week}	(22)
Sector 2.b: *Home backlog level (state) or stock variables* {unit}	
Backlog Home(t) = Backlog Home(t − dt) + (alter backlog home) × dt	(13)
INIT Backlog Home = 0 {SKU}	(13.1)
Sector 2.b: *Home backlog flow or rate variables* {unit}	
Alter backloghome = demand home-ship to client {SKU/week}	(24)
Sector 2.c: *Home forecast level (state) or stock variables* {unit}	
Est D Home(t) = Est D Home(t − dt) + (alter est D home) × dt	(25)
INIT Est D Home = demand home {SKU/week}	(25.1)
Sector 2.c: *Home forecast flow or rate variables* {unit}	
Alter est D home = demand gap home/alter est D time home {SKU/week/week}	(26)
Sector 2.c: *Home forecast auxiliary or converter variables and constants* {unit}	
Alter est D time home = 5 {week}	(27)
Demand gap home = demand home − Est D Home {SKU/week}	(28)
demand home = order from home {SKU/week}	(29)

Table 3 The home (*H*) supply-chain model sector equations, under the (*s*, *S*) stock policy

Sector 2.a: Home stock level (state) or stock variables *{unit}*	Eq. #
Identical with Eqs. 1 through 4.1 on Table 2	
Sector 2.a: Home stock flow or rate variables {unit}	
Identical with Eqs. 5 through 9 on Table 2	
Sector 2.a: Home stock auxiliary or converter variables and constants {unit}	
Coverage home = lead time home × Est D Home {SKU}	(30)
Lead time home = 3 {week}	(31)
Need for client = demand home + (Backlog Home/order time home) {SKU/week}	(32)
Order from factory = IF (stock position home ⩽ s home) THEN (MAX (0, (order up to S home − stock position (33) home)/order time home)) ELSE (0) {SKU/week}	(33)
Order time home = 1 {week}	(34)
Order up to S home = s home + Est D Home × order time home {SKU}	(35)
s home = coverage home + safety stock home {SKU}	(36)
Safety stock home = 100 {SKU}	(37)
SC home = WIP A Home + WIP B Home + WIP C Home	(38)
Stock position home = Stock Home + SC home + Backlog Factory − Backlog Home {SKU}	(39)
Sector 2.b: Home backlog level (state) or stock variables {unit}	
Identical with Eqs. 23 through 23.1 on Table 2	
Sector 2.b: Home backlog flow or rate variables {unit}	
Identical with Eq. 24 on Table 2	
Sector 2.c: Home forecast level (state) or stock variables {unit}	
Identical with Eqs. 25 through 25.1 on Table 2	
Sector 2.c: Home forecast flow or rate variables {unit}	
Identical with Eq. 26 on Table 2	
Sector 2.c: Home forecast auxiliary or converter variables and constants {unit}	
Identical with Eqs. 27 through 29 on Table 2	

Table 4 The home (*H*) supply-chain model sector equations, under the *utS* stock policy

Sector 2.a: Home stock level (state) or stock variables {unit}	Eq. #
Identical with Eqs. 1 through 4.1 on Table 2	
Sector 2.a: Home stock flow or rate variables {unit}	
Identical with Eqs. 5 through 9 on Table 2	
Sector 2.a: Home stock auxiliary or converter variables and constants {unit}	
Coverage home = lead time home + (alter est D time home − lead time home) {week}	(40)
Gap home = order up to S home − stock position home {SKU}	(41)
Lead time home = 3 {week}	(42)

(*continued*)

Sector 2.a: Home stock level (state) or stock variables {unit}	Eq. #
Need for client = demand home + (Backlog Home/time to order home) {SKU/week}	(43)
Order from factory = MAX (0, gap home/time to order home) {SKU/week}	(44)
Order up to S home = coverage home × Est D Home {SKU}	(45)
SC home = WIP A Home + WIP B Home + WIP C Home {SKU}	(46)
Stock position home = Stock Home + SC home + Backlog Factory − Backlog Home {SKU}	(47)
Time to order home = 1 {week}	(48)
Sector 2.b: *Home backlog level (state) or stock variables* {unit}	
Identical with Eqs. 23 through 23.1 on Table 2	
Sector 2.b: *Home backlog flow or rate variables* {unit}	
Identical with Eq. 24 on Table 2	
Sector 2.c: *Home forecast level (state) or stock variables* {unit}	
Identical with Eqs. 25 through 25.1 on Table 2	
Sector 2.c: *Home forecast flow or rate variables* {unit}	
Identical with Eq. 26 on Table 2	
Sector 2.c: *Home forecast auxiliary or converter variables and constants* {unit}	
Identical with Eqs. 27 through 29 on Table 2	

Table 5 Computing the factory stock amplification ratio *(FSAR)*

Sector 4: FSAR stock level (state) or stock variables {unit}	Eq. #
Stock Factory Δ(t) = Stock Factory Δ(t − dt) + (addΔ) × dt	(49)
INIT Stock Factory Δ = 0 {SKU}	(49.1)
Sector 4: FSAR flow or rate variables {unit}	
add Δ = IF (change in Stock Factory > Stock Factory Δ) THEN (change in Stock Factory/DT) ELSE (0) {SKU/week}	(50)
Sector 4: FSAR auxiliary or converter variables and constants {unit}	
Change in Stock Factory = ABS (Stock Factory) − INIT (Stock Factory) {SKU}	(51)
Factory stock % A = Stock FactoryΔ/initial stock factory {unitless}	(52)
FSAR: factory stock amplification ratio = factory stock %Δ/input %Δ {unitless}	(53)
Initial stock factory = INIT (Stock Factory) {SKU}	(54)
Input % Δ = 0.25 {unitless}	(55)

References

Akkermans H and Dellaert N (2005). The rediscovery of industrial dynamics: The contribution of system dynamics to supply chain management in a dynamic and fragmented world. *Syst Dynam Rev* **21**: 173–186.

Barlas Y and Ozevin MG (2004). Analysis of stock management gaming experiments and alternative ordering formulations. *Syst Res Behav Sci* **21**: 439–470.

Blattberg RC and Hoch SJ (1990). Database models and managerial intuition: 50% model + 50% manager. *Mngt Sci* **36**: 887–899.

Cachon GP and Lariviere M (1999). Capacity allocation using past sales: When to turn-and-earn. *Mngt Sci* **45**: 685–703.

Chen F, Drezner Z, Ryan JK and Simchi-Levi D (2000a). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Mngt Sci* **46**: 436–443.

Chen F, Ryan JK and Simchi-Levi D (2000b). The impact of exponential smoothing forecasts on the bullwhip effect. *Nav Res Logist* **47**: 269–286.

Cheung KL and Zhang AX (1999). The impact of inventory information distortion due to customer order cancellations. *Nav Res Logist* **46**: 213–231.

Clark TE and McCracken MW (2001). Tests of equal forecast accuracy and encompassing for nested models. *J Econom* **105**: 85–110.

Clark TE and McCracken MW (2005). Evaluating direct multi-step forecasts. *Econom Rev* **24**: 369–404.

Croson R and Donohue K (2005). Upstream versus downstream information and its impact on the bullwhip effect. *Syst Dynam Rev* **21**: 249–260.

Davis JP, Eisenhardt KM and Bingham C (2007). Developing theory through simulation methods. *Acad Mngt Rev* **32**: 480–499.

Eroglu C (2006). *An investigation of accuracy, learning and biases in judgmental adjustments of statistical forecasts*. PhD thesis, Ohio State University, USA.

Eroglu C and Croxton KL (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *Int J Forecast* **26**: 116–133.

Fildes R, Goodwin P, Lawrence M and Nikolopoulos K (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *Int J Forecast* **25**: 3–23.

Forrester JW (1958). Industrial dynamics—A major breakthrough for decision makers. *Harvard Bus Rev* **36**: 37–66.

Forrester JW (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press; currently available from Pegasus Communications: Waltham, MA.

Franses PH (2007). *Does experts' adjustment to model-based forecasts contribute to forecast quality?* Econometric Institute Report 2007-37, Erasmus University Rotterdam, The Netherlands.

Franses PH and Legerstee R (2009). Properties of expert adjustments on model-based SKU-level forecasts. *Int J Forecast* **25**: 35–47.

Gavirneni S (2006). Price fluctuations, information sharing and supply chain performance. *Eur J Opl Res* **174**: 1651–1663.

Georgantzas NC (2001). Simulation modelling. In: Warner M (ed). *International Encyclopedia of Business and Management*, 2nd edn. London, UK: Thomson Learning, pp 5861–5872.

Georgantzas NC (2003). Tourism dynamics: Cyprus' hotel value chain and profitability. *Syst Dynam Rev* **19**: 175–212.

Georgantzas NC (2009). Scenario-driven planning with system dynamics, In: Meyers B (ed). *Encyclopedia of Complexity and System Science.* Springer, NY, Entry # 573, p. 23.

Kleijnen JPC and Smits MT (2003). Performance metrics in supply chain management. *J Opl Res Soc* **54**: 507–514.

Kolassa S, Shutz W, Boylan JE, and Syntetos AA (2008). Judgemental changes to forecasts: Higher inventories, unchanged out-of-stocks. *28th International Symposium on Forecasting*, Nice, France.

Lane DC and Schwaninger M (2008). Theory building with system dynamics: Topic and research contributions. *Syst Res Behav Sci* **25**(4): 439–445.

Lee HL, Padmanabhan V and Whang S (1997a). Information distortion in a supply chain: The Bullwhip effect. *Mngt Sci* **43**: 546–558.

Lee HL, Padmanabhan V and Whang S (1997b). The Bullwhip Effect in supply chains. *Sloan Mngt Rev* **38**: 93–102.

Lee HL, So KC and Tang CS (2000). The value of information sharing in a two-level supply chain. *Mngt Sci* **46**: 626–643.

Otto A and Kotzab H (2003). Does supply chain management really pay? Six perspectives to measure the performance of managing a supply chain. *Eur J Opl Res* **144**: 306–320.

Paik S-K and Bagchi PK (2007). Understanding the causes of the bullwhip effect in a supply chain. *Int J Retail Distrib Mngt* **35**: 308–324.

Potter A and Disney SM (2006). Bullwhip and batching: An exploration. *Int J Prod Econ* **104**: 408–418.

Pujawan IN (2004). The effect of lot sizing rules on order variability. *Eur J Opl Res* **159**: 617–635.

Reiner G and Fichtinger J (2009). Demand forecasting for supply processes in consideration of pricing and market information. *Int J Prod Econ* **118**: 55–62.

Richmond B (2009). *Think® Software* (*version* 9.1.3). iSee Systems™: Lebanon, NH.

Saeed K (2009). Can trend forecasting improve stability in supply chains? A response to Forrester's challenge in Appendix L of Industrial Dynamics. *Syst Dynam Rev* **25**: 63–78.

Sterman JD (1994). Learning in and about complex systems. *Syst Dynam Rev* **10**: 291–330.

Sterman JD (2000). *System Dynamics: Systems Thinking and Modelling for a Complex World.* Boston, MA: Irwin McGraw-Hill.

Syntetos AA and Boylan JE (2008). Demand forecasting adjustments for service-level achievement. *IMA J Mngt Math* **19**: 175–192.

Syntetos AA, Boylan JE and Disney SM (2009b). Forecasting for inventory planning: A 50-year review. *J Opl Res Soc* **60**(S1): 149–160.

Syntetos AA, Nikolopoulos K, Boylan JE, Fildes R and Goodwin P (2009a). The effects of integrating management judgement into intermittent demand forecasts. *Int J Prod Econ* **118**: 72–81.

Syntetos AA, Nikolopoulos K and Boylan JE (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *Int J Forecast* **26**: 134–143.

Wong CY, El-Beheiry MM, Johansen J and Hvolby HH (2007). The implications of information sharing on bullwhip effects in a toy supply chain. *Int J Risk Assess Mngt* **7**: 4–18.

Yasarcan H and Barlas Y (2005). A generalized stock control formulation for stock management problems involving composite delays and secondary stocks. *Syst Dynam Rev* **21**: 33–68.

Zhang X (2004). Evolution of ARMA demand in supply chains. *Manuf Serv Opns Mngt* **6**: 195–198.

Comparing Discrete-Event Simulation and System Dynamics: Users' Perceptions

A.A. Tako and S. Robinson

Introduction

Discrete-event simulation (DES) and system dynamics (SD) are two popular simulation approaches used in operational research (Pidd 2004). DES models a system as a set of individual entities moving through a series of queues and activities in discrete time. SD models a system as a set of stocks and flows that are adjusted in pseudo-continuous time. It is clear that both approaches can be used to support management learning and decision-making (Robinson 2004). While some argue that DES and SD are quite separate simulation approaches (Brailsford and Hilton 2001), others see them as complementary to one another (Morecroft and Robinson 2005).

Journal of the Operational Research Society (2009) **60**(3), 296–312.
https://doi.org/10.1057/palgrave.jors.2602566
Published online 13 February 2008.

A.A. Tako (✉) • S. Robinson
School of Business and Economics, Loughborough University,
Loughborough, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_9

Work on the comparison of the two simulation approaches is limited, consisting mainly of some conference papers. The few comparative studies that can be found in the literature are mostly based on the authors' personal opinions (Brailsford and Hilton 2001; Morecroft and Robinson 2005). To date there has been no empirical study reported that provides an evidence base for a comparison of the two approaches.

The current study aims to address this dearth of evidence by identifying the significant differences and similarities between the two simulation approaches empirically. In this paper, the focus is on model use and more specifically we look at users' opinions about two simulation models, one in DES and the other in SD. A separate study is being undertaken on the model development process for DES and SD. The approach taken is to present managers (executive MBA students) with two models of the same problem situation and to evaluate their perceptions on: model understanding, complexity, validity, learning and model results. For the models a public sector problem, the UK prison population, was chosen. The key contribution of the study is to provide empirical evidence on the similarities and differences between DES and SD from a user's perspective.

The paper starts by reviewing existing work on the comparison of DES and SD model use. The empirical study is then presented in which the case study, the simulation models, the subjects of the study (executive MBA students), the model use sessions and the questionnaire are described. The results across a range of factors are given, before discussing the implications of the findings from the point of view of the modeller and the manager. Some limitations of the study are discussed and potential further work is identified.

Existing Work on the Comparison of DES and SD Model Use

This section reviews the existing literature on the comparison of two simulation approaches DES and SD, focusing on comments made regarding the aspect of model use and understanding. The comparisons are summarized under the following headings: understanding, complexity, model validity, model usefulness and model results.

The literature on the comparison of DES and SD is scarce. As a general comment, work on the comparison of the two simulation techniques consists mostly of generally accepted statements (Brailsford and Hilton 2001; Morecroft and Robinson 2005). Furthermore, comparisons tend to be biased towards either the DES or SD approach (Brailsford and Hilton 2001; Morecroft and Robinson 2005). We have not yet identified any empirical evidence certifying statements made about the use of DES and SD models.

Understanding

Both simulation approaches can be used to understand how systems behave over time (Sweetser 1999). However, contradictory statements are made regarding the level of understanding that users can gain from using these models. According to Brailsford and Hilton (2001), DES models are transparent to the clients. Animation and on-screen displays can provide useful insights about the model's structure. Lane (2000) argues that while DES models are convincing to the client, users do not necessarily understand the underlying mechanics of the model. On the other hand, Lane (2000) states that SD models are transparent and compelling to the client. Randers (1980), in a comparison of SD and modelling used for prediction (applicable to DES models), rated SD as having a higher capacity to increase clients' (users') understanding and also their learning, calling it 'insight generation capacity'. However, a disadvantage related to SD models is that there is no animation and the user has to rely on graphs and numerical displays (Sweetser 1999).

Complexity

Looking at both simulation approaches in terms of model complexity, DES is more concerned with detailed complexity, while SD with dynamic complexity (Lane 2000). This is due to their inherent features, where DES can model great complexity and detail, while SD represents the aggregate picture of the system. In SD, a model's behaviour is determined by the feedback structure and dynamic complexity arising from the influences among endogenous variables. In DES, complexity is the result of multiple random processes and the endogenous structure of the system (Lane 2000; Morecroft and Robinson 2005). It is generally claimed that DES follows an open-loop structure and feedback is not modelled (Coyle 1985). It has been argued, however, that feedback is involved in DES models but that it is not made explicit to the users (Sweetser 1999; Lane 2000; Morecroft and Robinson 2005).

Model Validity

We consider model validity in terms of model use as a measure of the user's confidence in the model (credibility). Credibility is seen in terms of representativeness (Robinson 2004, p. 231), confidence in the results and confidence in using the model for decision-making (Robinson 2004, p. 214). Randers (1980) rates SD models as highly representative, compared to a predictive model (including DES). It is generally accepted that both simulation approaches are concerned with building models that are representative of reality, providing confidence in the results and in decision-making. Indeed, Akkermans (1995) argues that in most cases DES and SD can represent the real world with equal validity.

Model Usefulness

Another important factor mentioned in the literature when comparing DES and SD is model usefulness. The concept of learning from using simulation models is widely mentioned in the SD literature (Forrester 1961; Morecroft and Sterman 1994). Business flight simulators are considered to be appropriate 'learning laboratories' that can help managers gain insights about their business operations. On the contrary, DES models are seen mostly as the domain of simulation experts and are used less as learning tools by non-technical managers (Sweetser 1999). However, these are statements made by modellers without considering users' opinions about specific models.

Another facet of model use is the nature of problems modelled by each simulation technique, 'strategic' *versus* 'tactical/operational'. It is generally accepted that while DES takes an analytic view, SD takes a holistic view of a system's performance. SD focuses mainly on strategic policy analysis, while DES tends to study operational, tactical problems (Sweetser 1999; Lane 2000). DES models generally have a narrow focus (Sweetser 1999) and are usually applied at an operational, tactical level (Brailsford and Hilton 2001).

Both simulation approaches can be used as tools to facilitate the communication of ideas in group discussions (Robinson 2004). Brailsford and Hilton (2001) state that DES software provides animation and graphics facilities, features 'very useful for communication with clients'.

Model Results

Mak (1993) points out that SD models provide a full picture of a system in the simulated period. In SD, point predictions are rarely made (Sweetser 1999). In DES modelling emphasis is given to point prediction, with outputs providing statistically valid estimates of the system's performance measures (Sweetser 1999; Brailsford and Hilton 2001; Law 2007). DES models provide a wide range of outputs, principally of a quantitative nature. Additionally, the interpretation of DES model results requires some statistical analysis. The outputs of one simulation run represent only one possible outcome due to the randomness in the model. For this reason, a practice used often in DES is running many iterations of the model with the use of different random number seeds (Pidd 2004; Robinson 2004). In order to make a proper analysis of the DES output, the model user should have some statistical background (Sweetser 1999; Brailsford and Hilton 2001).

When looking at model results, due to the inherent features of the two modelling techniques, different aspects of the model can be picked up by the users. DES models contain random variables and are stochastic in nature, while SD systems generally depict deterministic behaviour. Therefore, SD model results are considered as a source of understanding the reasons that cause changes in the system's performance, resulting from counterintuitive effects of the system's structural behaviour (Morecroft and Robinson 2005). Meanwhile DES modellers and model users are less interested in the events that actually cause these changes and focus more on the numerical results (Sweetser 1999).

Summary of Previous Comparison Work

The opinions stated in the literature comparing DES and SD are summarized in Table 1. There appears to be a general level of agreement on the nature of the differences. It should be noted, however, that a contrary view has been expressed by Akkermans (1995). He considered different types of modelling in business (DES, SD and spreadsheets) for real case scenarios. He claimed that as part of the model building process, the choice of modelling approach is not highly important. He also adds that the clients are usually not concerned about the choice of the simulation software used in a modelling project. If correct, this suggests that the differences may not be as clear cut as indicated in Table 1.

Methods and Research Design

The current empirical study aims to confirm/refute the statements found in the literature. The aim here is to empirically identify how different the two simulation techniques are from the users' point of view. More specifically our objective is to assess and compare the two simulation techniques in respect to the following criteria:

1. Understanding derived from using equivalent DES and SD simulation models.
2. Perceived complexity of equivalent DES and SD simulation models.
3. Credibility in using equivalent DES and SD simulation models.
4. Perceived usefulness of equivalent DES and SD simulation models in terms of learning, strategic thinking and communication of ideas.
5. Result interpretation of equivalent DES and SD models outputs.

In terms of this study an 'equivalent model' is a typical DES or SD model of the same problem situation.

Table 1 Summary of literature comparison of DES and SD model use

Model use	DES	SD
Model understanding		
Understanding (parts of) the model	The client does not understand the underlying mechanics	Models (links and flows) are transparent to the client
Animation	Animation and graphic tools help model understanding	No animation. Visual display of model aids model understanding
Complexity		
Level of detail	Emphasis on detail complexity	Emphasis on dynamic complexity
Feedback	Feedback is not explicit	Feedback effects are clear to the client
Model validity		
Credibility	Both models are perceived as representative, provide realistic outputs and create confidence in decision-making	
Model usefulness		
Learning tool	DES models are less used as learning tools	SD models, so-called ‘learning laboratories’, enhance users’ learning
Strategic thinking	DES models are mostly used in solving operational/tactical issues	SD models aid strategic thinking
Communication tool	Both DES and SD models are seen as good communication tools and facilitate communication with the client	
Model results		
Nature of results	DES provides statistically valid estimates of system’s performance. Results aid instrumental learning	SD model results provide a full picture of the system. Results aid conceptual learning
Interpretation of results	More difficult, requires users to have statistical background	Outputs are easily interpreted, little or no statistical analysis is required
Results observation	Randomness/variation of results is explicit	Generally deterministic results, which convey causal relationships between variables

Based on the literature discussed in Section "Existing Work on the Comparison of DES and SD Model Use" above, we expect users to find SD and DES models equally credible for giving answers to a problem situation and equally helpful as communication tools. However, we expect to find differences in users' opinions about model understanding, model complexity, interpretation of model results and the models' role in learning and strategic thinking.

The empirical study took the form of two separate sessions delivered to two different groups of executive MBA students at Warwick Business School as part of the 'Modelling and Analysis for Management' core module (Robinson et al. 2003). The experimental factor was the simulation model used. One group used a DES model and the other group a SD model. We asked the participants to use these models working in groups with the view to giving them hands-on experience. Their task was to provide answers as to how to solve the problem presented in the case study. At the end of the sessions, the participants evaluated the simulation models by completing a questionnaire survey.

The Case Study

Choosing an appropriate case study was considered important for the purposes of this work. The simulation models based on the case study need to be simple enough so that managers, who usually have little or no prior experience of simulation modelling, can be in a position to understand and use them for decision-making. The case study and models were designed to fit with the MBA course curriculum and requirements for a 1.5-h session. In addition, a suitable case study needs to accommodate models from both simulation techniques, so that the specific features of each technique (randomness in DES *versus* deterministic models in SD, the aggregated presentation of entities in SD *versus* the individual representation of entities in DES, etc) are present in the models built.

After thoughtful consideration, a case study on the UK prison population based on Grove et al. (1998) was chosen. The prison population is a topical subject both in the UK and elsewhere (e.g., Korporaal et al., 2000). The inherent feedback that exists in the system, with prisoners entering and returning back to prison due to re-offending (recidivism), can be uniquely represented by each simulation approach, DES and SD. DES and SD have both been used to model the prison population. DES models of the prison population have been developed by Kwak et al. (1984), Cox et al. (1978) and Korporaal et al. (2000). An SD model has been developed by Bard (1978). Therefore, we consider the UK prison population as a suitable case study to use for this research.

The case study starts with a brief introduction to the prison population problem and draws particular attention to the issue of overcrowded prisons. An overview of the system is shown in Fig. 1. Two types of offenders are considered, petty and serious offenders. There are initially 76,000 prisoners in the system, of which 50,000 are petty and 26,000 serious offenders. Offenders enter the system as first-time offenders and receive a sentence depending on the type of offence. Petty offenders enter the system at a higher rate, due to a higher rate of offending (on average 3000 people/year *versus* 650 people/year for serious offenders), but receive a shorter sentence length (on average 5 years *versus* 20 years for serious offenders). After serving time in prison, the offenders are released. A proportion of the released prisoners re-offend and go back to jail (recidivists) after, on average, 2 years. Petty prisoners are more likely to re-offend, 70% re-commit petty crimes and go back to jail and another 3% commit even more serious crimes and are re-convicted as serious offenders. Serious offenders represent a small percentage of the total offender population and have lower rates of recidivism. Only 30% of serious offenders re-offend and go back to jail as serious offenders after 2 years.

The case study presents the reasons for, and impacts of, the problem, followed by a set of possible alternative policies, which can be implemented in order to solve the existing problem. The problem is presented as a typical public sector resource allocation problem with an objective to improve the capacity of the criminal justice system in preventing crime and deterring its repetition, by taking into consideration the specified budget allocation. Based on the above, the case study's question is: 'taking the role of a government consulting service, suggest possible policy changes to government authorities'. Since both DES and SD have been used to address this issue in practice, we believe that the two techniques have a rough equivalence with respect to this objective.

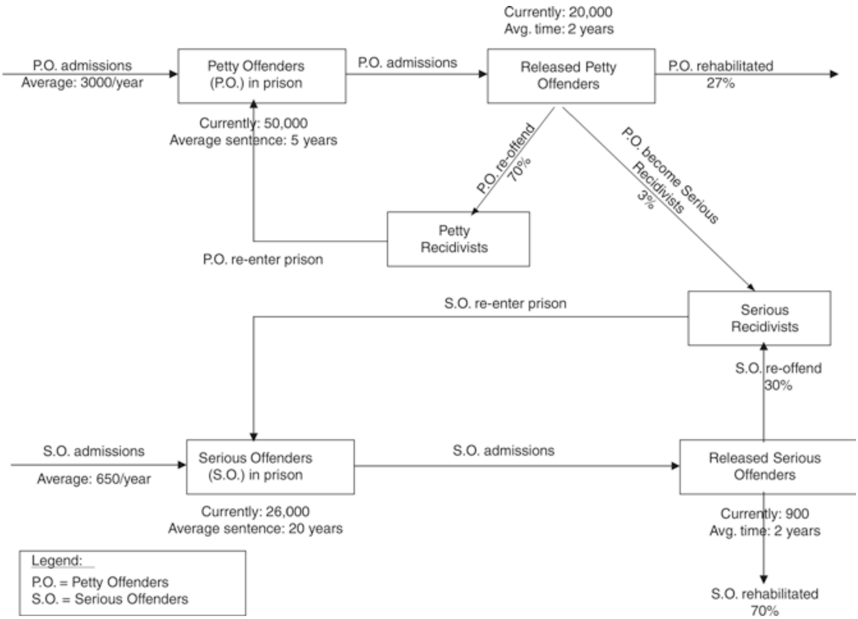


Fig. 1 Overview of the prison population problem

The Simulation Models

Based on the case study described above, a DES and an SD simulation model were built. Both simulation models are a simple representation of a prison overcrowding problem showing how the prison population evolves over time. Our main objective was to build two simple models, which enable experimentation with different scenarios/policies, with the intention of using them as tools for decision-making. The DES model was developed using WITNESS (www.lanner.com, accessed November 2007), a powerful and versatile DES simulation package. For the SD model, Powersim Studio 2005 (www.powersim.com, accessed November 2007) was used. This is a package used widely in the field of SD. Both models incorporate a user interface that enables inputs to be set and altered. Witness and Powersim are typical of the simulation software in their respective fields. Although there is some variation in the facilities in alternative packages, there is no specific reason to believe that the choice of package would have much influence on the representation of a simple model such as the prison population case used in this research (Robinson 2008).

The simulation environment of the DES model is presented in Fig. 2. The model environment includes a number of different windows that consist of the model (the box on the left in Fig. 2), the input data (on the top, right-hand corner) and the model outputs (at the bottom, right-hand corner). Control buttons (i.e., run, stop, reset, etc.) are included and also a window that reports the time. On the click of the run button, a window appears asking the users to choose the relevant input data according to the policy or policies chosen. The user can also access a series of graphical results by selecting the graphs button. These include plots of the prison population, plots of rehabilitated prisoners and plots of the recidivists over time, and also bar charts with the distribution of sentence lengths, for both petty and serious offenders.

In the DES model, entities enter the system and two attributes are set, 'sentence length' and 'time to re-offend'. The entities then go straight to queues that represent the prison population, either as 'PettyinPrison' or as 'Serious-inPrison'. In these queues, they serve time according to the attribute 'time in prison'. Prisoners then go into the release activity ('ReleasePetty' or 'ReleaseSerious') from where they are either rehabilitated and exit the system ('Ship') or go to the recidivist queues, ('PettyRecidivists' or 'SeriousRecidivists') according to the crime they have committed. In the recidivist queues, the entities stay according to the attribute 'time to re-offend' and then go to the re-offend activities ('PettyReoffend' or 'SeriousReoffend'), where the attribute 'sentence length' is reset. From there the entities re-enter the prison population.

The SD model consists of four different pages: introduction, control panel, prison population diagram and the main model. The pages are linked via hyperlinks so that users can easily navigate from one page to the other. The SD model representation is shown in Fig. 3. Two separate flows, petty and serious admissions enter the system and go straight into the prison population stocks ('Petty criminals in prison' and 'Serious criminals in prison'). Prisoners flow out of prison through the outflows ('Petty Release rate' and 'Serious release rate') to the stocks 'Released petty' and 'Released serious'. Prisoners leave the released prisoner stocks

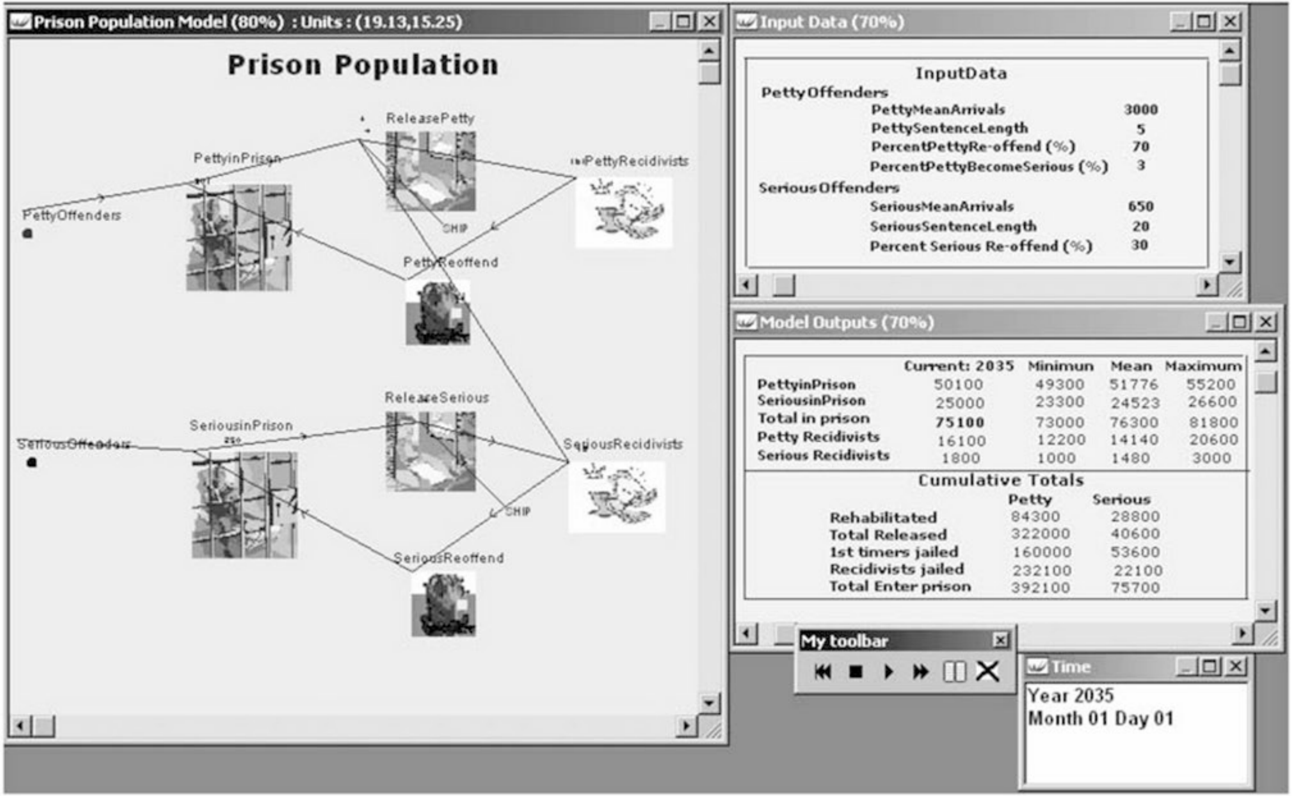


Fig. 2 DES model representation in WITNESS, with the model on the left-hand side, input criteria in the top box on the right and in the box below model outputs

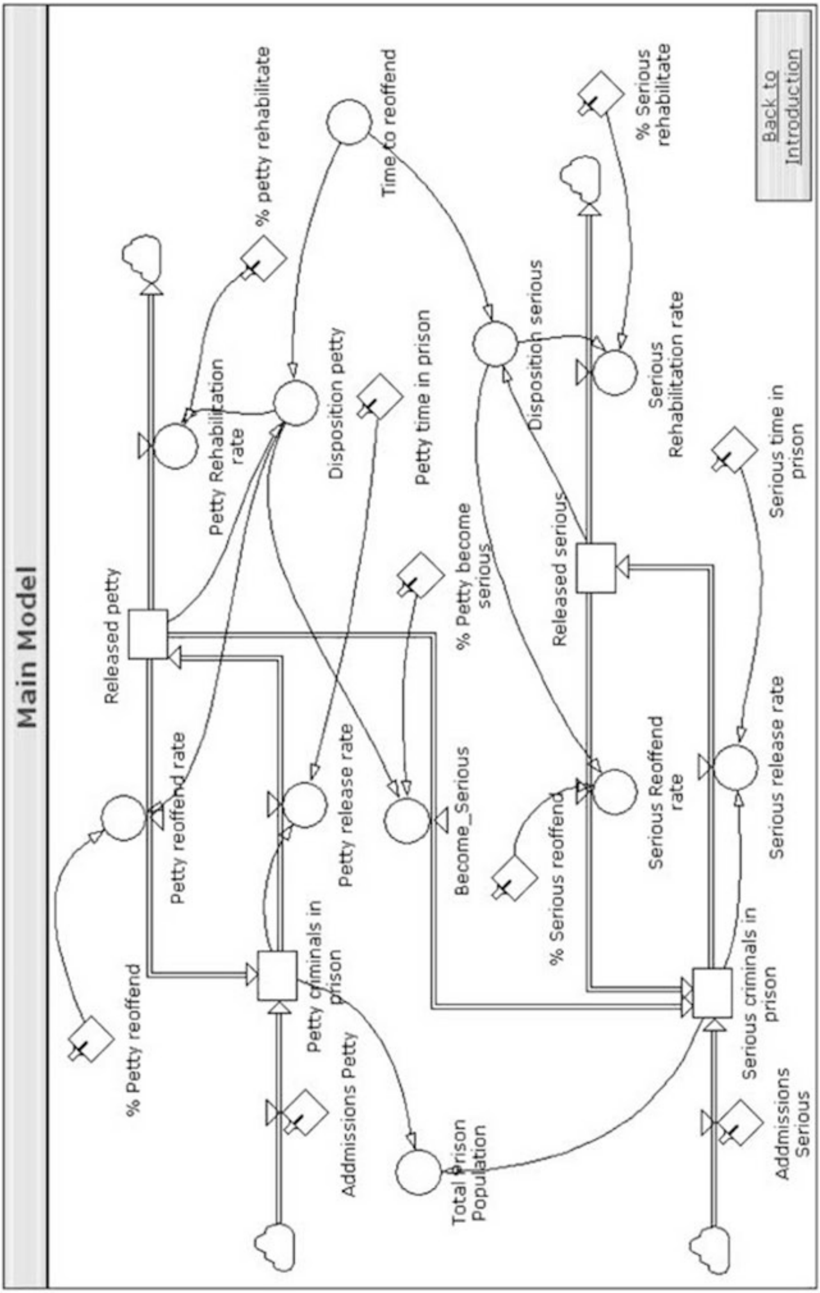


Fig. 3 SD model representation in Powersim

either as rehabilitated prisoners or re-offenders, the latter creating a feedback loop to the prison population stocks. The stock 'Released petty' has an additional outflow, 'Become Serious', which takes a small part of the released prisoners straight to the stock 'Serious criminals in prison'. The feedback structure, typical of SD models, is evident in the flows of released and re-offending prisoners.

The control panel (Fig. 4) is the main working environment where users can interact with the model and enter inputs according to their choice of policy or policies and observe relevant outcomes. The control panel consists of two parts. The user interface includes a set of sliders for the prison admission rates and the sentence time, and three combo boxes that provide choices for the percentage of re-offending. Next to the user interface are the model results, which consist of a set of graphs and tables of key outputs that are simultaneously updated.

Some key differences can be observed in the DES and SD models presented above. In the DES model the entities are individually represented and specific attributes assigned to them, that is, sentence length, offender type, number of times incarcerated etc. Due to the large number of entities,

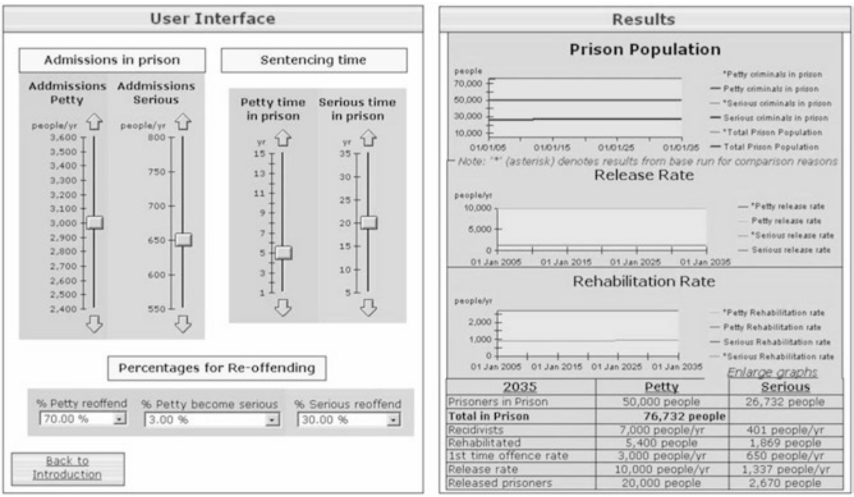


Fig. 4 SD model control panel, including the user interface and model results page—the main working environment

the run speed of the model becomes very slow and so the numbers have been scaled down to a fraction of 1/100, where one entity represents 100 offenders. Grouping entities is a well-known practice in DES modelling (Robinson 2004). Therefore, it can be claimed that there is some level of aggregation involved in the DES model. However, the main feature of DES, which enables the tracking of entities (in this model the group of 100 prisoners) and their attributes, is still present. After all, one of the main reasons DES is chosen in practice is its capacity to track individuals in the system.

On the other hand, in the SD model the entities are presented as a continuous quantity, where state changes happen continuously at small segments (Δt) of time. In the SD model, the time-step (Δt) used is 1 year. Specific entities cannot be followed through the system. Therefore, it can be claimed that there is a higher level of aggregation in the SD model than in the DES model. Modelling the large number of people in the system did not require any specific handling in SD, which is naturally suited to dealing with large populations.

Key variables in the DES model are sampled using the exponential or Erlang distributions, for example admissions to prison, time to re-offend and sentence length. In this way randomness is incorporated into the model. On the contrary, in the SD model these same variables are represented as deterministic average values.

Another fundamental difference in the two model representations is related to how the initial number of people already in the system is set up. Powersim and all SD software packages have a facility for setting up the initial stocks at the beginning of the simulation run. In Witness, however, there is no such facility for queues. Two options were available. The model could be run for a warm-up period to allow the system to fill up to the desired level. The other option was to create dummy entities that enter the model at the start of the simulation run and are assigned to the various queues. The latter option was considered more appropriate, as a warm-up period would have added significantly to the run time of the model and it would have been less intuitive for the users. Because the DES model collects results on the individual entities, each dummy entity had to be given a history of when it had entered the model, otherwise the results would have been skewed. This was achieved by sampling negative times of entry.

A conceptual difference between the two models is the way that released prisoners were dealt with. In the DES model, released prisoners who do not re-offend leave the system straight away after being released. In contrast, in the SD model all released prisoners are kept for 2 years in the released stocks and after that a proportion of the stock flows out of the model. The difference arose because in the SD model it is necessary to accumulate all released prisoners into a stock before determining what happened to them next.

Regarding the data requirements, both models required almost the same data inputs. The DES model is a close representation of the existing real-life system, with variables set (approximately) to the values as described in the case by Grove et al. (1998). However, in the SD model variables that do not exist in real life were created in order to represent intended behaviours. For instance, some variables were created such as disposition ('Disposition Petty' and 'Disposition Serious'), in order to obtain the correct proportions of re-offending and rehabilitation. Disposition calculates the release rate for all prisoners at liberty, who stay for 2 years in the stocks of released prisoners, before calculating the rates of re-offending and rehabilitation. In this respect, it seems that SD has more flexible structures.

Despite the differences discussed here, both models depict almost similar behaviour and the key outputs are quite similar (Table 2). There are probably some differences in variable definitions from one model to the other, and thus some differences in the results. For example in the DES model the cumulative number of released prisoners (petty and serious) is displayed in the outputs, while in the SD model the number of released prisoners (petty and serious) at liberty at a specific point of time is displayed. In addition, in the DES model the number of recidivists represents the number of released prisoners at liberty in the community who will re-offend at some point in the future. Whereas in the SD model this number represents the rate of re-offending, that is, the number of prisoners who re-offend annually. Despite these differences, the two models can still be considered equivalent.

Table 2 Comparison of DES and SD models outputs

	DES outputs	SD outputs
Petty in prison	50,100 people	50,000 people
Serious in prison	25,000 people	26,732 people
Total in prison	75,100 people	76,732 people
Petty recidivists	16,100 people	7000 people/year
Serious recidivists	1800 people	401 people/year
Released petty	–	20,000 people
Released serious	–	2670 people
Total released petty	322,000 people	–
Total released serious	22,100 people	–

The Subjects

In any organization, it is the managers who are the ultimate users of a simulation model, whether it be directly experimenting with the model or as recipients of the results. In the latter case, the manager would normally interact with the model to, at least, gain some confidence in the results. Managers, therefore, were considered the most relevant participants for the purposes of this study. Since we had ready access to executive MBA students at Warwick Business School, these were chosen as the subjects of the study.

The executive MBA students at Warwick are highly representative of managers working in the public and private sector. They have on average 12 years of work experience (www.wbs.ac.uk/students/mba/learn/class-profile-mod.cfm, accessed November 2007) and at the same time studying or holding managerial positions in their organizations. During the first year of their studies they take a core module, Modelling and Analysis for Management (MAM) (Robinson et al. 2003), on which one of the authors (Robinson) teaches.

The study was implemented with two different groups of MBA students who took the MAM module at two different times, the first in June 2006 and the second in February 2007. The first group consisted of 57 participants, this group used the DES model. The second group was made up of 37 participants and evaluated the SD model.

The Sessions

Before the sessions, the subjects were given the case study description to read in advance. The sessions started with a brief presentation introducing the case study, the basics of the simulation models and how they work. Two further sets of hand-outs were given, of which one was the model description and the other included guidance as to how to use each model. The participants were then divided into syndicated groups and were asked to work on the task for 30–40 min. During this time, they were asked to take the role of the government consulting service and to identify solutions to the problem. The groups consisted of 4–6 participants. All group members were involved in group discussions. During the group session, the researchers (the authors) were roaming from group to group providing support for technical problems and answering questions about the case. A feedback session followed, where two random syndicate groups for each session presented their findings and further discussions and comments were made by all participants. At the end, a questionnaire was handed out, which the participants were asked to complete and to return to the researchers.

The Questionnaire

A two-page questionnaire was devised in order to explore the MBA students' views about the models. The questionnaire consists of two parts. The first part deals with the participants' job details and the second part with the participants' opinions about the simulation models used as part of the exercise. The main question format used for collecting the users' opinions on the models was the 5-point Likert-type scale ranking from 1 to 5, giving an ordinal, non-metric measurement. The 1–5 response scale is commonly used in social science research (Buckingham and Saunders 2004). Other types of questions included are rank order/multiple-choice, single select (yes/no) questions and open-ended questions.

This is an innovative study in the simulation area looking into managers' perceptions of DES and SD simulation models. We could not find any pre-conceived measures on simulation model use in the simulation literature. Therefore, we created the measures used here from our experience as modellers and the statements made in the literature concerning DES and SD. A pre-test was run with five PhD students from Warwick Business School to check the clarity of the questions and the layout of the survey. As a result, some changes were made in the wording of some of the questions. The questionnaire dealt with participants' opinions in terms of: model understanding, model complexity, model validity, model usefulness and model results. We briefly explain below the questions included in the main body of the questionnaire. (The questionnaire is available on request to the authors).

Questions regarding model understanding dealt with the extent that users feel they understand the model and parts of it. On model complexity, a set of questions focused on the users' opinions about the perceived level of detail in the models and also about the sources of complexity they discern. Questions on model validity dealt with the subjects' opinions about the extent to which they think the models are representative of the case study situation, that model outputs are realistic and about their confidence in using the model in decision-making. The next question asked the participants to rate model usefulness in terms of learning, strategic thinking involved and communication of ideas. Concerning the simulation results, we were interested to find out what type of data (numerical *versus* graphical) users referred to when looking at the results. The aim was to find out what attitude the models induced when handling the results (instrumental *versus* conceptual learning). The next question asked the users about the level of difficulty in interpreting the results, the use of graphs (and randomness associated with them) and about the way of thinking when looking at the results ('Do users look for the factors that cause changes in the results?', which is a characteristic attitude in the SD world).

Survey Results

In this section, the results of the statistical analysis on the data collected from the questionnaire survey are presented. Overall the empirical work does not identify significant differences for most of the comparison criteria for DES and SD model use.

In order to test for differences in users' opinions, nonparametric statistical tests are carried out due to the nature of the data obtained from the questionnaire (ordinal and nominal data). According to Siegel (1957), meaningful statistics for nominal data are frequency counts and the mode, and for ordinal data, the median. Diagnostic (probability–probability) P–P plots are used to explore graphically differences in the distributions of ordinal data comparing answers received from the two groups of users. Fisher (1983) and Law (2007) suggest the use of P–P plots in order to compare two distributions. When the plot is linear or close to linear, the two distributions of answers fit one another, meaning that the variables have identical distributions (Wilk and Gnanadesikan 1968; Fisher 1983; Law 2007). The chi-square test for the nominal data and the Mann–Whitney–Wilcoxon test (Fisher 1983) for the ordinal data are used to check that the differences are statistically significant.

Respondents Profiles

From the questionnaire survey with two different groups of executive MBA students, 34 usable questionnaires were derived from the DES group (implemented in June 2006) and 30 from the SD group (implemented in February 2007). This gave response rates of 65 and 79%, respectively. The participant groups were two mixed groups of executive MBA students in terms of background and management level.

Considering the industry sector participation in the survey sample (Table 3), the majority of the DES group came from the public services

Table 3 Sample representation by industry sector

Industry	DES group (%)	SD group (%)
Public services	32	–
Manufacturing	21	40
Business services	18	13
Financial services	9	3
Transport and communication	9	13
Energy and mining	6	13
Trade	3	3
Construction	3	3
Other	–	10

sector (32%—11 respondents) and from manufacturing (21%—seven respondents), whereas the SD group had no representation from the public services. We can argue that participants from the public services sector are more familiar with problems in the prison population case study and so the DES group could be considered more predisposed to the exercise and the models. The majority of the respondents in the SD group came from the manufacturing sector (40%—12 respondents). There was a smaller representation of the other sectors in both groups.

Respondents were also asked to indicate their functional areas and their position in the management hierarchy. Participants in the DES group consisted of 34% working in the production/operations area, 20% in sales and marketing and 9% in computing/IT services, with a lower representation from finance, procurement, R&D and customer services. A somewhat similar picture was observed in the SD group, with 23% of respondents working in the production/operations area, 27% in sales and marketing, and 13% in computing/IT services and a lower representation of the other areas.

Regarding the participants managerial level (Table 4), the majority of the DES group (61%) came from the lower (line) manager level with higher and middle management having a lower representation. Meanwhile, the SD group had a somewhat different representation, with the proportions being 40 and 47% for middle and lower level management, respectively, while higher management had a lower representation. This suggests that both groups had a somewhat different mix regarding managerial level, which might affect the answers and thus the results. However, middle and line managers counted together represented 88% of the DES group and 87% of the SD group. Having a high representation of line management positions in both samples is considered to be beneficial for the survey. The authors believe that managers of a lower

Table 4 Managerial level for each group

Management level	DES group (%)	SD group (%)
Executive	12	10
Middle management	27	40
Manager	61	47
Other	–	3

Table 5 Prior experience by management level (includes both DES and SD samples)

Management level by prior experience	Count	
	No	Yes
Executive	5	1
Middle management	17	4
Manager	21	12
Other	2	1
Total	45	18

level tend to use simulation to a higher extent as a problem-solving tool. In fact, considering both groups together (Table 5), line managers made up the majority of respondents with prior experience—12 (out of 18)—and only four (out of 18) middle managers had prior experience (Table 5). There was only one instance of a higher-level manager with prior experience of simulation.

Comparing the Level of Understanding Using the DES and SD Model

The respondents were given a series of statements regarding their understanding of the models when using either the DES or the SD model. Understanding deals with overall model understanding, understanding of the relationship between variables, understanding of the model structure, understanding of how to use the model and understanding of the model outputs. The level of understanding for each of these items is measured on a scale of 1–5, where 1 means 'understand very little' and 5 means 'understand very well'. The aim here was to measure the users' opinions about their understanding of the simulation model and parts of it and then to compare the answers from both groups.

The P–P plots reveal differences in DES and SD model users' opinions only for the variables: understanding of the relationship between variables and understanding of how to use the model (Figs. 5 and 6). The P–P plots consist of five data points, where each dot represents the cumulative probability for each Likert- scale measure (1 = understand very little, up to level 5 = understand very well), with 1 on the left and 5 on the right. The DES probabilities are plotted on the *x*-axis and for SD on the *y*-axis.

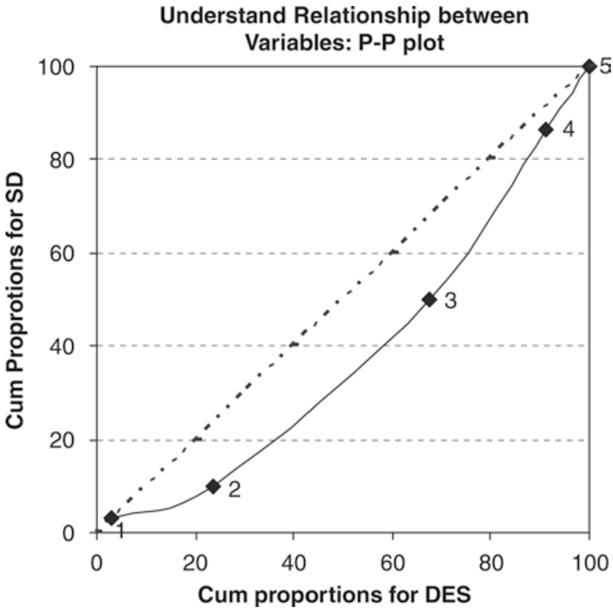


Fig. 5 P–P plot on understanding of the relationship between variables, SD *versus* DES answers, where 1 means understand very little and 5 understand very well

Looking more closely at both graphs, we can observe that the lines are skewed toward the DES model. This means that the DES model users gave a higher proportion of responses in the mid-range (understand little, moderate and understand well, levels 2–4, respectively), while the SD model was mostly rated at the higher levels of the scale (levels 3–5). This implies that SD model users perceived that they had a better level of understanding regarding the relationship between variables and how to use the model. The Mann–Whitney–Wilcoxon test, however, does not identify these differences as significant.

In the P–P plots for the other items on understanding (overall understanding of the model, understanding its structure and understanding of the model outputs), there is little difference between the two groups. This is confirmed by a lack of statistical significance in the differences as well.

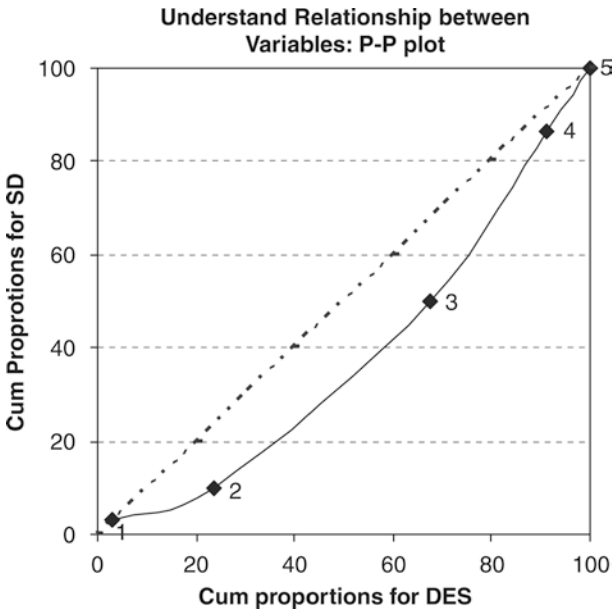


Fig. 6 P–P plot on understanding of how to use the model, SD *versus* DES answers, where 1 means understand very little and 5 understand very well. Points 1 and 2 coincide with the origin of the coordinates (0, 0) because none of the respondents answered with understand very little, and little, for either model

Factors that Help in Model Understanding

The question regarding the factors that help model understanding asked the user to rank in order of importance the factors paper-based material, visual display of the model and animation as the model runs. Looking at the answers received for each factor in Table 6, there is a clear difference in the rankings of the DES group and the SD group for the factors paper-based description and animation as the model runs. This shows that the DES group identified animation as the most important factor that aided model understanding (58.8%), followed by the paper-based description as very important (55.9%). Meanwhile, the SD group identified the paper-based material as the most important factor (62.1%). However, there is no clear difference in the two groups' rankings regarding the visual display of the model. DES and SD users equally rated it as the least important factor.

Table 6 Ranking of factors that helped user understanding of the models (DES and SD)

Factor by model type	Important (%)	Very important (%)	Most important (%)
Paper-based material			
DES	17.6	55.9	26.5
SD	10.3	27.6	62.1
Visual display			
DES	73.5	20.6	5.9
SD	75.9	17.2	6.9
Animation			
DES	11.8	29.4	58.8
SD	24.1	48.3	27.6

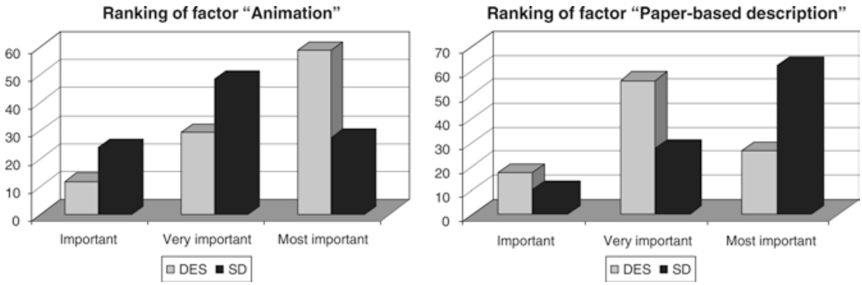


Fig. 7 Frequency diagram showing importance of animation and paper-based description as factors that helped user understanding of the model (DES and SD)

We draw attention here to Fig. 7 that shows the differences between the two groups regarding the factors animation as the model runs and paper-based material. The Mann–Whitney–Wilcoxon test shows that there is indeed a significant difference in users' opinions regarding these factors at a 1.4 and 2.9% significance, respectively. This suggests that for the DES model animation has the greatest impact in helping model understanding, while for the SD model the paper-based material has most effect.

Model Complexity

Concerning the level of detail, a Likert-type question asked the user to rate the simulation models, where 1 represents very detailed and 5 a very

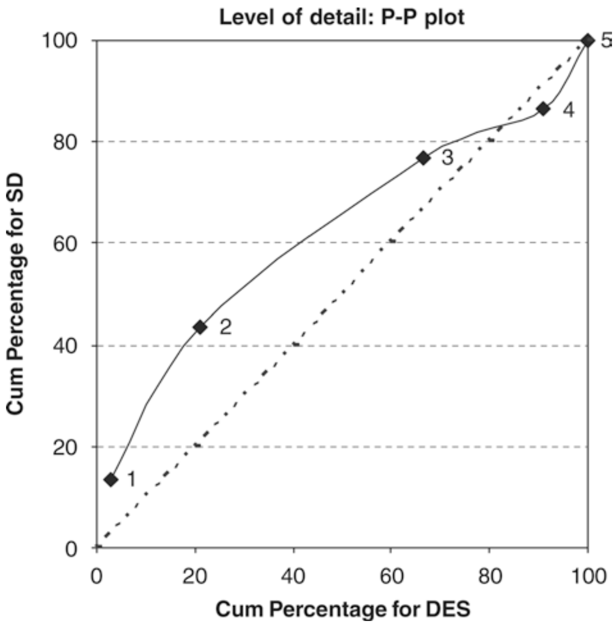


Fig. 8 P–P plot on level of detail of the model, SD *versus* DES answers, where 1 means very detailed and 5 meant very high level

high-level perspective. The P–P plot in Fig. 8 reveals a skew towards the SD model, with the SD model having a higher proportion of answers at the lower level of the scale, corresponding to a greater level of model detail. This is the opposite of what we expected since it is generally thought that DES models are more detailed (Section “Complexity”).

The users could have perceived the SD model as more detailed due to the fact that all the components of the SD model are explicitly presented on screen (as per Fig. 3), whereas for the DES model the structure may not be so explicit (Fig. 2). The actual relationships between variables in DES models are not so apparent to the users when compared to SD models where the stocks, flows and auxiliary variables are displayed on the screen. Despite some skew towards the SD model, results in the P–P plot (Fig. 8), the chi-square and Mann–Whitney–Wilcoxon tests do not reveal any significant differences between the two samples.

Next the questionnaire consisted of an open-ended question asking the users to identify the sources of complexity in the model. The aim was to find out how obvious the feedback in each model (DES and SD) is to both groups of survey participants, without specifically mentioning 'feedback' in the question. It should be noted that the students had received no instruction on feedback as part of the MBA module. We hoped the users would identify the feedback in the model by considering the complexity that arises due to prisoners re-entering prison. Only 20% of the DES group answers and only 3% of the SD group answers are found as correct. Correct answers are considered as those that refer to the relationship and the interdependency between variables or to re-offending. A chi- square test reveals a significant difference in the proportions between the two groups, with a chi-square value χ = 4.33, significant at 3.7% level. Contrary to what we were expecting, this suggests that the feedback effects are more explicit to the DES model users as compared to the SD model users.

One possible reason for this counterintuitive result might be because the users did not actually explore the models enough in order to pick up on their underlying features. In the case of the SD model, the users would not be able to pick up the feedback effects between variables unless they navigated to the model representation page. We are cautious, however, about this finding due to the low response rate to this question (the response rate was 35.3% for the DES group and 13.3% for the SD group).

Model Validity

In terms of model validity, a section of the questionnaire dealt with whether the user found the models to be representative and the outputs realistic. It also asked about the user's confidence in the models. The P–P plots do not show a difference between the two groups, apart from the plot on model representativeness. Observing the P–P plot in Fig. 9, the data are skewed toward the DES group, revealing that the users of the DES model rated it as being representative at lower levels, mostly levels 2 and 3 (little and moderate, respectively), while SD model users rated the model higher, mostly levels 3 and 4 (moderate and much). This implies that the SD model was perceived to be more representative of the case study compared to the DES model. Furthermore, a Mann–Whitney–Wilcoxon test identifies a somewhat significant difference at a 6.5% level.

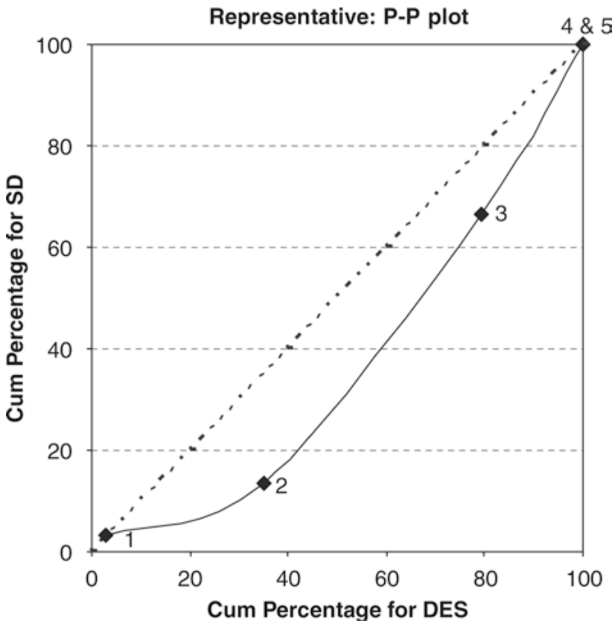


Fig. 9 P–P plot on model representativeness, SD *versus* DES answers, where 1 means very little and 5 very much. Points 4 and 5 coincide because none of the respondents considered the models representative at level 5

When performing a Mann–Whitney–Wilcoxon test on the answers of just users with no prior experience of simulation models, there is a more significant difference between the DES and SD groups, significant at the 1.7% level. This finding suggests that, for users with no prior simulation experience, the SD model was perceived to be more representative of the case study as compared to the DES model. An obvious reason for this result could be that, as discussed above, the SD model structure is more explicit than the DES model structure. One DES model user commented that they would be interested to see the underlying maths.

Model Usefulness

In a separate section, three Likert-type questions and one open-ended question were used to reveal users' opinions regarding the usefulness of

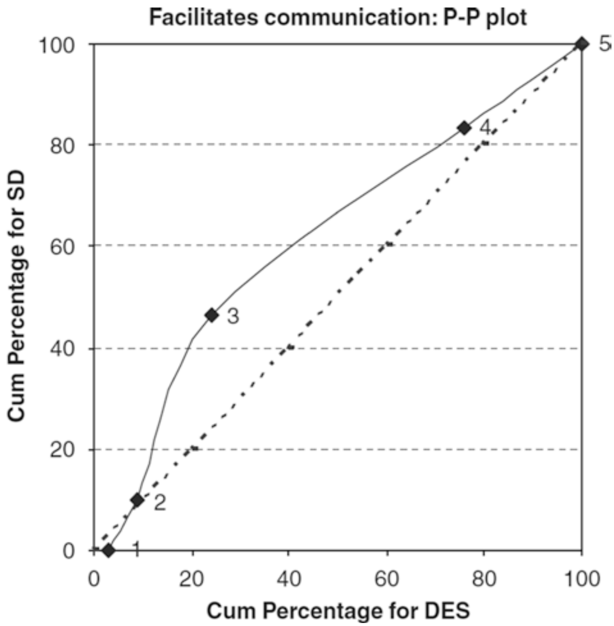


Fig. 10 P–P plot on the capacity of the model to facilitate the communication of ideas, SD *versus* DES answers, where 1 means very little and 5 very well

the two simulation models. The Likert-type questions asked users to express their opinions about whether the use of the models enhanced their learning, whether it helped them think about the problem and whether it facilitated the communication of ideas. Next, with an open- ended question we asked the participants to identify systems that are similar to the context of the prison population model. This question aimed to identify whether after using the prison population model the participants could transfer the knowledge gained to other similar systems. 'Knowledge transfer' can be used as an indicator of the learning achieved (Morecroft and Sterman 1994).

The P–P plots do not identify any differences in the responses to the Likert-type questions apart from the question as to whether the use of the models facilitates the communication of ideas. In the P–P plot in Fig. 10, the line is skewed towards the SD axis in the lower levels of the scale. Here the users rate mostly high and very high the DES model in facilitating the communication of ideas. The Mann–Whitney–Wilcoxon test, however, does not identify any significant differences for the three Likert-type questions.

As for the open-ended question, only 23% of responses from each group are considered as appropriate answers to the question. Examples of correct answers are hospital/bed occupancy and social (unemployment) services. This indicates that the same level of learning was achieved by both groups. However, we are cautious about our findings here because there was a high level of no response to this question (the response rate was 44 and 36.6% for DES and SD group, respectively). On the other hand, it is not clear why some participants did not answer this question. It might be that no answer reflected a lack of learning and so a lack of ability to transfer the knowledge gained.

Model Results

In terms of model results, an issue of importance is the type of results users look at when running a simulation model. DES model users were expected to focus on 'instrumental learning' and so were expected to look more at numerical data. Meanwhile SD model users were expected to use graphs to a higher extent with more of an interest in 'conceptual learning'.

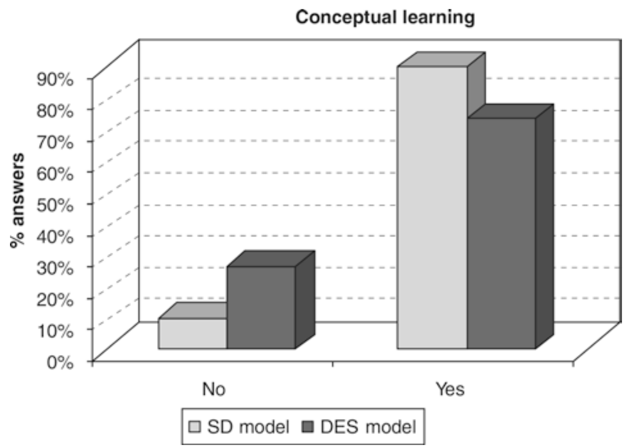


Fig. 11 Bar chart with frequencies of DES and SD model users who used graphical outputs (conceptual learning)—a higher proportion of SD model users

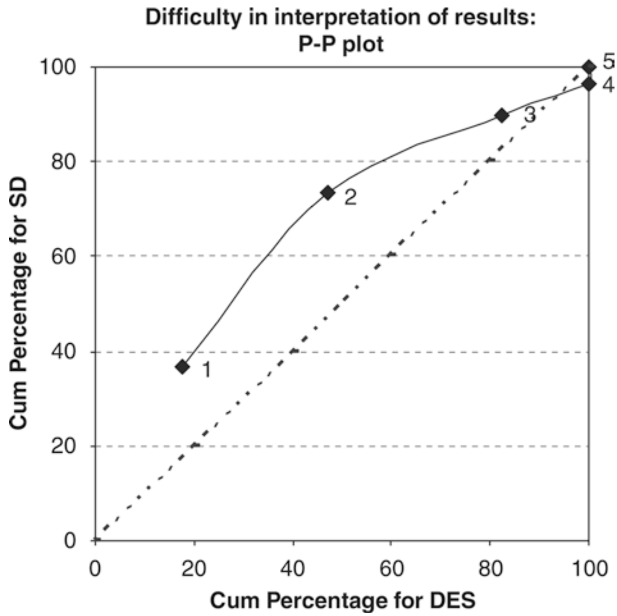


Fig. 12 P–P plot on perceived difficulty in the interpretation of model results, SD *versus* DES answers, where 1 means very straightforward and 5 very difficult

The questionnaire results show that, almost the same proportion of participants from both groups used the numerical results (instrumental learning). On the other hand, a higher proportion of respondents in the SD group claimed to have used the graphs (conceptual learning) as compared to the DES group. The bar chart in Fig. 11 reveals the differences in the level of use of graphs between the two groups. Indeed, a relaxed chi-square test reveals a significant difference between the two groups at a 9.2% level of significance.

The rest of the questions dealt with the user's perceived difficulty in the interpretation of results, the usefulness of graphs and the examination of factors that cause differences in the results. The data reveal a difference in the two groups' opinions regarding the difficulty in the interpretation of results. The P–P plot in Fig. 12 is significantly skewed towards the SD model, meaning that SD model users found the results interpretation less difficult. Also a Mann–Whitney–Wilcoxon test reveals a significant difference at the 3.6% level.

Regarding the users' attitude when interpreting the model results, an open-ended question asked the user to identify what is the main learning from the graphs and whether the user tried to identify the factors that cause changes in the outputs. These questions are intended at finding whether the two models trigger model users to employ different attitudes towards model results. It was expected that DES model users would take notice of the randomness present in the outputs, and therefore in response to the question they would mention randomness as their main learning point from the graphs. In the case of the SD model, the users were expected to be looking for the endogenous factors that cause the changes in the variables' behaviour. However, from the responses received there is no evident difference between the two groups of users regarding their attitude to interpreting the results.

Summary of Results

A summary of the main findings derived from the questionnaire is presented in Table 7.

Understanding, defined as the users' opinion on the level of understanding gained from using the two simulation models, is not found to be significantly different for the two (SD and DES) groups. Some differences are observed regarding the factors that help users understand the model and parts of it. Animation is found as the factor that mostly aids model understanding for the DES group, while for the SD group it is the paper-based description of the model. This complies with the views of Sweetser (1999) and Morecroft and Robinson (2005) that animation and on-screen displays can help model understanding. However, our results suggest that the understanding gained from using a DES model (because of animation and on-screen displays) is not necessarily more than the understanding achieved when using an SD model. Even though, these findings suggest that the level of user understanding is the same, it can be argued that users gain different insights from the two models. However, observation of the DES and SD groups using the models suggests that this was not the case in the current study, because similar issues and policies were considered by both groups during their discussions. The case

Table 7 Summary of results comparing DES and SD model use

Model use	DES	SD
Model understanding		
Level of understanding the model (and parts of it)	No differences identified in users' opinions	
Factors that helped understanding	Animation the most important factor	Accompanying model descriptions is most important
Complexity		
Level of detail	Similar level of perceived detail	
Feedback	Feedback effects more explicit to DES model users	–
Model validity		
Representative of real problem	–	SD model just more representative as compared to DES model
Realistic outputs	Outputs are perceived similarly realistic	
Confidence in outputs	Similar level of confidence in model outputs	
Model usefulness		
Learning	Similar level of learning achieved from using DES and SD models	
Strategic thinking	Same level of perceived strategic thinking involved	
Communication of ideas	Same level of communication perceived to have taken place	
Model results		
Instrumental/conceptual learning	Both SD and DES aid instrumental learning	
	–	SD model aids conceptual learning to a higher extent
Interpretation of results	DES model results were more difficult to interpret	–
Attitude when interpreting results	No differences in the users' attitude	

and accompanying materials were, of course, the same for both sets of users, and so this is not unexpected.

For *complexity*, it is found that users from both groups rated the two simulation models as having a similar level of detail. A counterintuitive finding of this study is that the feedback effects are found to be more explicit to the users of the DES model. Contrary to the general belief that SD is more appropriate in representing feedback structures (Coyle 1985; Sweetser 1999; Morecroft and Robinson 2005), DES can represent feedback effects, which in this case appear to be more explicit to the user. However, we are cautious about this finding because of the small number of answers received to the open-ended question on model complexity. Another issue to be considered is the subjectivity in the choice of the two model representations. The DES and SD models could have been represented in many different ways. We were, of course, only able to choose one mode of display for each model.

Regarding *model validity*, this study suggests that the extent to which the users perceive the models to be representative of the case study is different between the two groups. The SD model is found to be just more representative. For both models, the outputs are perceived to be equally realistic and both groups of users had the same level of confidence in them. The higher level of perceived representativeness related to the SD model can probably be attributed to the overall picture of the system provided with the SD model representation. On the other hand, the finding that model outputs and the confidence in the model are equally rated by both groups implies that overall the level of users' acceptance of both models is not different.

Model usefulness is not identified as different between the two models. Against generally accepted opinions (Sweetser 1999), the findings suggest that both simulation approaches can be used as learning tools and can both trigger the communication of ideas. Even though in the SD literature a range of examples exists that illustrate the use of models for learning and for the communication of ideas (Vennix 1996; Sterman 2000), there are also cases where DES models have been used in facilitating group discussions and problem understanding (Robinson 2001, 2002).

For *model results*, the findings indicate that the users of both the DES and SD models use the numbers (numerical displays) to the same extent. Meanwhile, the SD users focus on graphical displays more than the DES users, suggesting that SD models can aid conceptual learning and thus help users look at the bigger picture. Regarding the level of difficulty in the interpretation of results, our findings support the literature (Brailsford and Hilton 2001) that the DES model results are more difficult to interpret, even though this specific model and the results were fairly simple. No differences are identified in the users' opinions about the use of graphs and in the attitude employed by the users when interpreting the model results. However, a difference in attitudes was observed by the researchers during the group discussions. The SD model users tended to take a 'goal seek' approach, where they blindly changed the inputs in order to get the right output, and then reflected on what policies might be employed to achieve these inputs. The DES group did not employ the same approach and focused on the effect a policy might have on the inputs to the model and then set the input values accordingly.

Discussion and Concluding Remarks

The current study adds to the discussion on the comparison between DES and SD. To the best of our knowledge, this is the only empirical study that tests the differences in using DES and SD simulation models. The survey presented provides empirical evidence about how users' perceive the differences between DES and SD. The comparison criteria used in the survey are based on the generally accepted opinions/statements regarding the differences in using DES and SD found in the literature. Overall, it was not possible to identify many significant differences in the users' opinions regarding the specific DES and SD models used. This may imply that from the user's point of view the type of simulation approach used makes little difference if any. Akkermans (1995) reaches a similar conclusion, identifying that clients are usually indifferent to the simulation language being used. This may not be too surprising, as users are likely to be more interested in what they can learn from a model than about how the model works, that is, as long as the modelling approach is able to address the problem situation. However, we do need to consider whether this is a general conclusion or whether it is a result of some limitations in the validity of the study.

The participant groups involved in the exercise were two mixed groups of executive MBA students in terms of background and level of management and thus comparable to each other. There was a high representation of first line managers, who tend to be more involved in simulation projects, as compared to higher level managers. It should be noted that from the sample used in this study, the proportion of participants with no prior experience was higher than those with prior experience. Both groups commented on the simplicity of the models, but at the same time they appreciated their usefulness for the purpose at hand. Participants tended to be looking for more sophisticated models, considering a wider range of factors such as costs, deaths and other types of sentences.

In the current study, we used the best possible samples to which we had access at the time. Of course, the study could be improved with larger samples. In the DES group, the highest proportion of participants had a background from public services and manufacturing, while in the SD group, manufacturing had the highest proportion. In the latter group there was no representation from the public sector. As a general comment, the DES group expressed a greater interest in the exercise, especially because it was a problem related to their jobs for a reasonable proportion (32%) of the group. This in itself could have biased their answers. It would be considered a more fair experiment if the participants were randomly allocated into each group. However, random assignment of participants in the two groups was not possible because each MBA group took the same course at different times (May 2006 and February 2007) and it would have been difficult to present both simulation models to people with little or no prior experience of using simulation in a session of 1.5 h. It was observed that because the users were exposed to only one of the two simulation models, they tended to take for granted the features of each simulation model, and did not pick up the specific features of each approach, which differ from one another. A solution to this would be to get the participants to work with both simulation models. However, this was not possible due to the limited amount of time available.

There is some level of subjectivity in the choice of the case study and the simulation models. The case study was chosen because it was amenable to both DES and SD modelling. Use of an alternative case study may have provided different findings in terms of the comparison of DES and SD. Meanwhile, a specific DES and a specific SD model were built of the prison population problem. These were only one representation in each approach out of many (if not an infinite number of) possible representations. Would different DES and SD models of the problem have led to different findings? To mitigate this effect, the DES and SD models were developed with the help of experts in their respective fields. It is believed that these models are typical DES and SD models, but it cannot be claimed that they are the only possible models.

Future work could compare DES and SD using different case studies, and a range of different models and simulation packages could be investigated for each case study. The authors of this paper are also studying the differences in terms of model development, involving experts from both simulation modelling approaches. The comparison in this case deals with the concepts and stages that DES and SD modellers go through when building simulation models. The authors intend to implement this study by 'observing' participants while building simulation models (using DES or SD). It is expected that more significant differences between DES and SD will be found in the comparison of the model development process.

References

Akkermans HA (1995). *Modelling with managers. Participative business modelling for effective strategic decision-making*. PhD thesis, Technische Universiteit Eindhoven.

Bard JF (1978). The use of simulation in criminal justice policy evaluation. *J Crim Justice* **6**(2): 99–116.

Brailsford S and Hilton N (2001). A comparison of discrete event simulation and system dynamics for modelling healthcare systems. In: J Riley (ed). *Proceedings of ORAHS 2000*, Glasgow, Scotland, pp. 18–39.

Buckingham A and Saunders P (2004). *The Survey Methods Workbook: From Design to Analysis.* Cambridge; Malden, MA: Polity.

Cox GB, Harrison P and Dightman CR (1978). Computer simulation of adult sentencing proposals. *Eval Program Plann* **1**(4): 297–308.

Coyle RG (1985). Representing discrete events in system dynamics models: A theoretical application to modelling coal production. *J Opl Res Soc* **36**(4): 307–318.

Fisher NI (1983). Graphical methods in nonparametric statistics: A review and annotated bibliography *Int Stat Rev* **51**: 25–58.

Forrester JW (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Grove P, MacLeod J and Godfrey D (1998). Forecasting the prison population. *OR Insight* **11**(1): 3–9.

Korporaal R, Ridder A, Kloprogge P and Dekker R (2000). An analytic model for capacity planning of prisons in the Netherlands. *J Opl Res Soc* **51**(11): 1228–1237.

Kwak NK, Kuzdrall PJ and Schniederjans MJ (1984). Felony case scheduling policies and continuances—a simulation study. *Socioecon Plann Sci* **18**(1): 37–43.

Lane DC (2000). *You just don't understand me: Models of failure and success in the discourse between system dynamics and discrete event simulation*. Working paper, 00.34:26.

Law AM (2007). *Simulation Modeling and Analysis*. 4th ed. Boston and London: McGraw-Hill.

Mak H-Y (1993). *System dynamics and discrete event simulation modelling*. PhD thesis, London School of Economics and Political Science.

Morecroft JDW and Robinson S (2005). Explaining puzzling dynamics: comparing the use of system dynamics and discrete- event simulation. In: JD Sterman, MP Repenning, RS Langer, JI Rowe, JM Yarni (eds). *Proceedings of the 23rd International Conference of the System Dynamics Society*. Boston, MA: System Dynamics Society.

Morecroft JDW and Sterman J (eds) (1994). *Modeling for Learning Organizations*. System Dynamics Series. Portland, OR: Productivity Press.

Pidd M (2004). *Computer Simulation in Management Science*. Chichester: Wiley.

Randers J (1980). *Elements of the System Dynamics Method*. Cambridge, MA; London: MIT Press.

Robinson S (2001). Soft with a hard centre: Discrete-event simulation in facilitation. *J Opl Res Soc* **52**(8): 905.

Robinson S (2002). Modes of simulation practice: approaches to business and military simulation. *Sim Mod Practice and Theory* **10**(8): 513–523.

Robinson S (2004). *Simulation: The Practice of Model Development and Use*. Chichester: Wiley.

Robinson S (2008). Conceptual modeling for simulation Part 1: Definition and requirements. *J Opl Res Soc*, online publication 24 January 2007, doi:https://doi.org/10.1057/palgrave.jors.2602368.

Robinson S, Meadows M, Mingers J, O'Brien FA, Shale EA and Stray S (2003). Teaching OR/MS to MBAs at Warwick Business School: A turnaround story. *Interfaces* **33**(2): 67–76.

Siegel S (1957). Nonparametric statistics. *Am Stat* **13**(3): 13–19.

Sterman J (*2000*). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston; London: Irwin/McGraw-Hill.

Sweetser A (1999). A Comparison of System Dynamics and Discrete Event Simulation. In: RY Cavana, JAM Vennix, EAJA Rouwette, M Stevenson-Wright and J Candlish (eds). *Proceedings of 17th International Conference of the System Dynamics Society and 5th Australian & New Zealand Systems Conference*. Wellington, New Zealand: System Dynamics Society.

Vennix JAM (1996). *Group Model Building: Facilitating Team Learning Using System Dynamics*. Chichester: John Wiley.

Wilk MB and Gnanadesikan R (1968). Probability plotting methods for the analysis of data. *Biometrika* **55**(1): 1–17.

Part III

Applications of System Dynamics at Industry Level

Modelling the Sustainability of Mass Tourism in Island Tourist Economies

Y. Xing and B. Dangerfield

Introduction

As one of the world's largest industries, the tourism industry accounts for approximately 12% of world Gross National Product (GNP) with corresponding receipts of US$747 billion (OECD 2001; World Tourism Organisation (WTO) 2003). It follows that the tourism industry has a responsibility to show leadership in sustainability. Tourism is an extremely complex phenomenon, which cuts across many sectors such as transportation, hotels, fresh water supplies, waste management and energy. These aspects are not always considered as being part of the same sector and their roles in sustainable tourism development may be difficult to

Journal of the Operational Research Society (2011) **62**(9), 1742–1752.
https://doi.org/10.1057/jors.2010.77
Published online 15 September 2010.

Y. Xing
University of Ulster, Newtownabbey, UK

B. Dangerfield (✉)
Department of Management, University of Bristol, Bristol, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_10

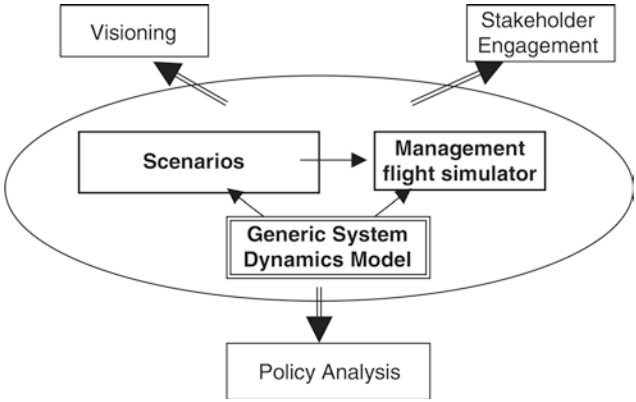


Fig. 1 Research objectives in this study

separate from their other functions. Also, there is a growing awareness of the negative impacts that tourism can have. Examples of such impacts on the environment, especially in coastal and mountain areas and in small islands, are described by Bramwell and Lane (1993). This growing concern, along with the principle of sustainable development (World Commission on Environment & Development 1987), has brought the tourism industry and international organisations to re-assess tourism policymaking in the light of its long-term economic, social and environmental sustainability.

The purpose of this research is to highlight the contribution that system dynamics can make in demonstrating the possibility of boom and bust in island tourist economies as well as for analysing policies that promote sustainable tourism development. As shown in Fig. 1, in this paper a generic sustainable tourism model is described and a set of scenarios for policy analysis are presented. The scenarios are expected to be able to provide insightful information about the possible impacts of policies. In order to help the various stakeholders achieve a holistic view of tourism development and collaborative policymaking (Jamal and Getz 1995; Roberts and Simpson 1999; Hall 2000; Wang and Fesenmaier 2007; Yang 2007), a microworld (or management flight simulator) has also been created. The details of the microworld are omitted here but are set out in Xing (2006).

Conventional Regression Tourism Demand Models

The equation shown below is a typical equation seen in a regression model for tourism demand.

$$Q = \mathrm{F}(Y, TP, ER),$$

where Q = tourist arrivals; Y = income; TP = tour price; and ER = exchange rate.

This relationship effectively analyses the changes in tourist arrivals derived from explanatory variables such as incomes, tour prices and exchange rates. It can be seen that a conventional statistical model such as this oversimplifies tourism demand to offer unreasonable correlations from a few numerical variables. Qualitative variables are overlooked, yet these soft variables can be crucial and important for policymaking. To leave out such variables and concepts is to say explicitly that they have no importance. Further, such models are usually static models that arbitrarily assume equilibrium exists.

The equation may take an explicit form such as $Q_{it} = AP_{it}^{\beta_1} Y_{it}^{\beta_2} P_{st}^{\beta_3} e_{it}$, where Q_{it} is the tourism demand variable measured by tourism arrivals from country region i to the tour destination at time t; P_t is the price of tourism in the tour destination at time t; P_{st} is the price of tourism in the substitute destination at time t and Y_{it} is the income level of the origin country or region i at time t; and e_{it} is the residual term that is used to capture the influence of all other factors that are not included in the demand model. This last term is important as tourism demand is influenced by many economic and non-economic factors that might be excluded because of the non-availability of data. The $Q_{it} = AP_{it}^{\beta_1} Y_{it}^{\beta_2} P_{st}^{\beta_3} e_{it}$ can be transformed to a linear equation in natural logarithm format, such as:

$$\ln Q_{it} = \beta_0 + \beta_1 \ln P_{it} + \beta_2 \ln Y_{it} + \beta_3 \ln P_{st} + \mu_{it},$$

where $\beta_0 = \ln A$; $u_{it} = \ln e_{it}$; and β_1, β_2, β_3 are price, income and substitute price elasticities, respectively. Assuming equilibrium exists by letting $Q_{it} = Q_{it-1}$ (such as in an Autoregressive Distributed Lag Model), it is argued that this could produce the most accurate result. But the question to be raised here is: can an equilibrium be assumed? It is obviously not true in a turbulent tourism development environment. The future might not necessarily repeat history. Tourism regression models concentrate only on the tourist flow generation aspect (only one of six sectors in our model) and ignore the consequences of the volume of tourist arrivals. Without analysing the impacts of the possible actions holistically, sustainability can never be fully understood.

A Generic System Dynamics Model of an Island Tourist Destination

Preliminary Conceptualisation of the Model

According to the WTO the volume of international tourism arrivals from 1950 to 2000 grew at an average 6.8% annually worldwide, 13.2% in Asia and 6.5% in Europe. Tourism 2020 Vision (WTO 2003) forecasts show that international tourist arrivals are expected to reach over 1.56 billion by the year 2020. This demonstrates an annual growth rate of 4.1 per cent per annum over the period 1995–2020 (Fig. 2).

The driving forces for tourist flows can be classified into ‘push factors’ and ‘pull factors’ (Crompton 1979; Pearce and Butler 1993). However, the push and pull factors are interrelated and need to be analysed holistically. The pull and push factors taken together can be described as a ‘destination consideration’ (Fig. 3). Clearly there is a multi-criterion issue with respect to the tourist flow to a particular destination. SWOT has been a common method for assessing a destination’s Strengths, Weaknesses, Opportunities and Threats and it has been widely used among managers to assess business strategy. But it has been argued that SWOT concepts are ambiguous, qualitative and fact-free (Warren 2002).

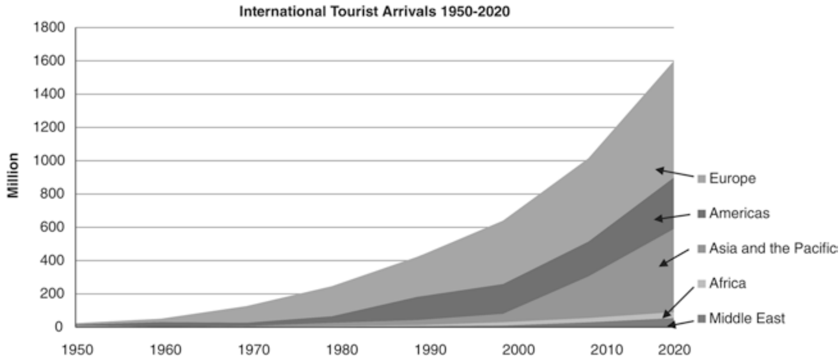


Fig. 2 World tourism development (arrivals). (Source: WTO 2003)

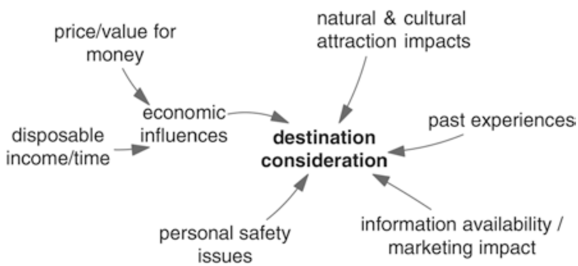


Fig. 3 Destination considerations in respect of potential tourism demand

Accordingly, a SWOT analysis offers little help in answering the quantitative questions related to sustainable tourism development issues.

To analyse tourism development, we have to analyse what type of socio-economic, environmental and personal conditions generate tourist flows. Moreover, once tourist flows are generated, a range of tourism-related activities will follow. Those activities have direct or/and indirect influences on future tourist flows together with socioeconomic and environmental sustainability conditions. These conditions will in turn react on tourist flow generation and tourism-related activities. This system structure is summarised and represented in Fig. 4.

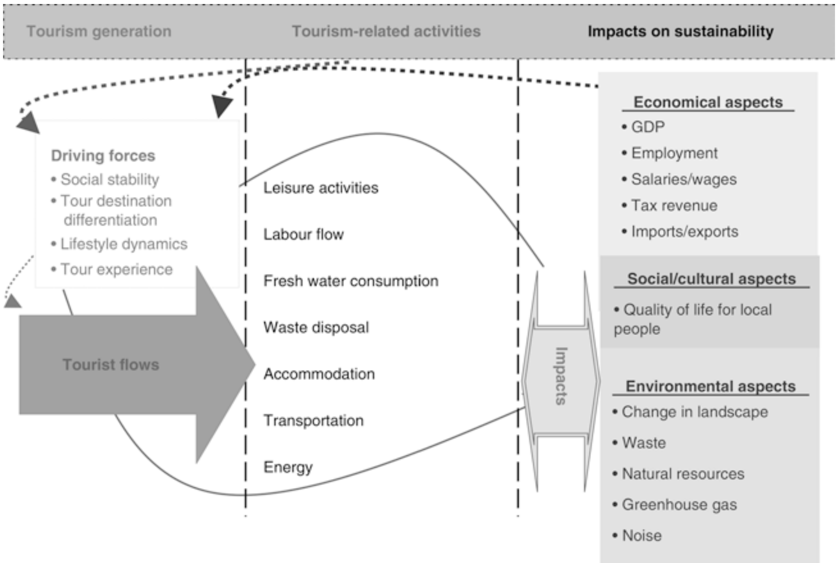


Fig. 4 High-level view of tourism system structure

Basic Structure of the Generic Tourism Model

The tourism system dynamics model (for a simplified overview see Fig. 5, for details of the model see Xing 2006) includes the following sectors: tourist flow generation, labour market, hotels, energy, water and waste, and finally transportation. The sectors contain interacting elements. In the tourist flow generation sector, 'population at tourist generation areas' (i.e. the source population for European tourism) and 'tourists in the tour destination' are modelled as stocks. There are three major factors affecting the number of outgoing tourists: population, holiday making rate (fraction of the population making at least one trip away from the usual residence within the year) and the number of visits per person per time period. The calculation of 'tourist flow generation' (to the tour destination) reflects the effects of changes in the destination's attractiveness and capacity for accepting new tourists.

The effect of the attractiveness index (AI) on potential tourist flows to the tour destination is modelled as a nonlinear relation against the

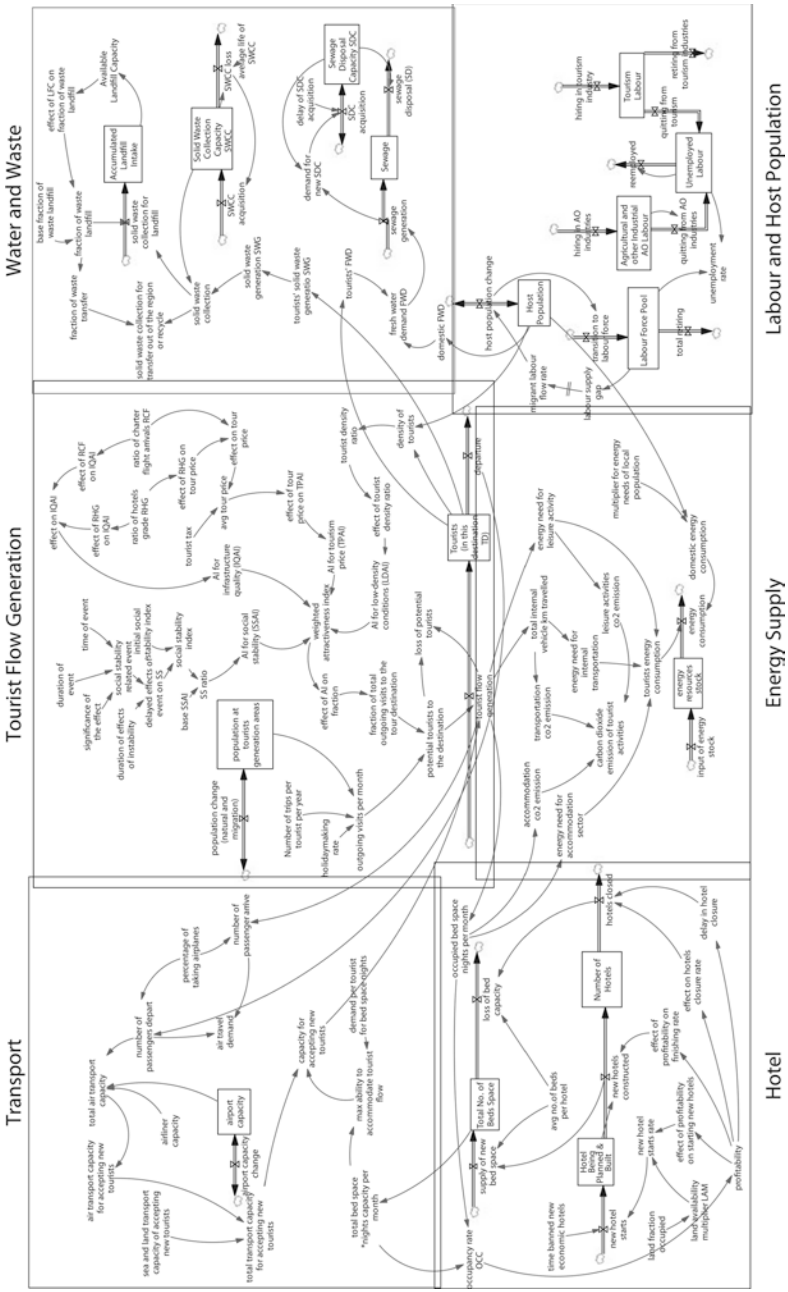


Fig. 5 A simplified overview of the tourism model

weighted AI. The model simulates generic tourism behaviour and mimics the growth of European 'sun and sand' tourism from approximately the early 1960s through to the year 2020. Therefore, using months as the basic time unit, the final time is 720. A month was selected for the basic time unit because the normal short holiday breaks taken would be small fractions of a year and this would have implied an even smaller value for the model's TIME STEP. A detailed description of the model and its equations can be found in Xing (2006).

Model Testing to Improve Users' Confidence

One of the key elements in model validation is to test whether the model fits the purpose of the modelling exercise (Forrester 1961, p. 137; Sterman 2000, p. 89). As yet, there are no islands that have experienced a boom and bust behaviour pattern, but that outcome is certainly feasible if things remain unchecked. Our paper suggests a feasibility—a possible but not assured eventuality. Two of the fundamental questions raised here in modelling the sustainability of mass tourism in island tourist economies are how to avoid a 'Tragedy of the Commons' scenario (Hardin 1968) and a fire-fighting syndrome—the unplanned allocation of resources to fix problems discovered late in a product's development cycle (Repenning 2001). Thus the model's utility has to be judged in respect of drawing attention to a possible eventuality given a formulation that exhibits face validity.

Testing the model is an essential step embedded within the system dynamics model construction process. The ultimate goal of model testing is to improve users' confidence in the model. Richardson and Pugh (1981) point out that 'a system dynamics model addresses a problem, not a system, and is designed to answer a reasonably well-defined set of questions'. The importance of model purpose cannot be over-emphasised: 'Fundamental to the choice of methodology is the need to define the purpose of the model, termed problem definition, and for this purpose to be agreed by all parties concerned' (Dangerfield 2008).

On the basis of theory developed by Forrester and Senge (1980) and Sterman (2000), an iterative model testing process is developed and presented as illustrated in Fig. 6.

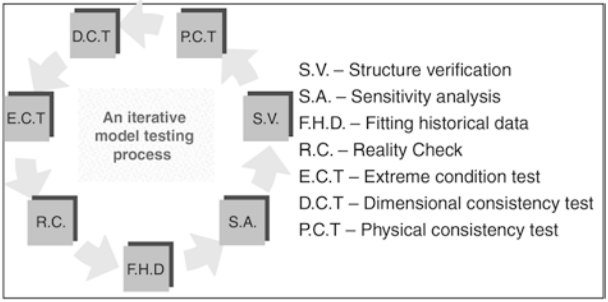


Fig. 6 An iterative model testing process (adapted from Forrester and Senge 1980)

Structure verification tests ask whether the model is consistent with knowledge of the real system and relevant to the purpose. For our generic model this included a half-day presentation of the model structure and assumptions to academic colleagues belonging to the (then) School of Leisure, Hospitality and Tourism at the University of Salford. Other tests conducted in this research included sensitivity analysis. This involved changing assumptions about the value of parameters in the model and examining the resulting output for consequent changes. Monte Carlo simulation (or multivariate sensitivity simulation) was utilised for this test and realised through the Vensim software that renders this procedure automatic. A 'Reality Check®', physical consistency test and extreme condition tests were also successfully carried out in order to improve confidence in this tourism model.

To be an effective policy analysis tool, a system dynamics model should also be able to reproduce relevant aspects of past history (Homer and Keane 1999). Our model allows an assessment of the impact on social stability of a damaging external event occurring at a tourist destination. Although this aspect of the formulation is not restricted solely to terrorist activity, it is illustrated through consideration of the terrorist bombing that occurred on October 12, 2002 in the town of Kuta on the Indonesian island of Bali, killing 202 people and injuring a further 209. Hotel occupancy rates fell to single figures within days and even in 2003 tourists were only just starting to venture back, in part as a result of massive price discounts on the island. The graph in Fig. 7 shows a comparison of

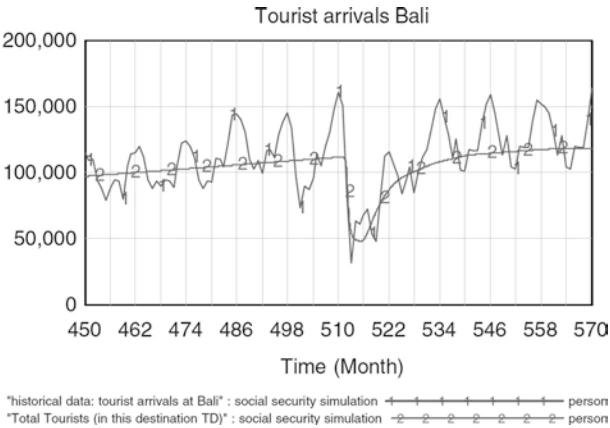


Fig. 7 Comparison of simulation result with actual data. (N.B. The time axis is cropped and covers the period from 1997 to 2007)

monthly arrival data with the simulated result. The impact of the bombing was modelled as a pulse function of delayed effects of the event (time, duration and significance). The facility to model such an eventuality is included in the tourist flow generation sector of Fig. 5.

Turning to the reproduction of past history in the absence of any unanticipated external events, it has to be stressed that, since no island has yet experienced an overshoot and collapse situation in tourist numbers, any historical data (and equivalent simulation of tourist arrivals) will most likely show a continuous growth trend. The validation of the *World Dynamics* and *Limits to Growth* models by reference to past data covered only the growth phase. The projected overshoot has yet to occur and the purpose of those models was to issue a warning call: it is exactly the same here. We have created a simplified structural mechanism supporting tourist flow generation and its consequences. Then the model is articulated by tuning parameters to show plausible scenarios. Validation of a complex socio-economic system model is an on-going process. The model will evolve while more data and facts are established. Presently, however, the model helps us to think harder. As such, a model is very useful for analysts and policymakers to deal with what is a complex set of interacting phenomena in island tourist economies.

Policy Analysis Based Futures for Mass Tourism

There is a significant amount of uncertainty, nonlinear changes and attitudinal data involved in fully understanding the forces behind tourism development. Sustainable tourism planning must be capable of addressing widely different situations (Hunter 1997). It has long been recognised that accurate prediction is not a feasible goal. However, it is possible to formulate scenarios that can shed light on, and offer insights about, possible future developments and thereby improve organisational learning (Van der Heijden 1996; Parry and Carter 1998; Ringland 1998). Scenario planning can help with a higher level of strategic thinking that integrates uncertainty-based futures thinking, a process that is necessary for sustainable tourism policy analysis.

A few attempts have been made to apply scenario planning concepts in a tourism context. For example, the Singapore Tourism Board examined a methodological process and the marketing implications of a series of events using a Delphi approach (Yong and Keng 1989). Weaver classified and analysed four tourism destination development scenarios (Weaver 1998). Eden and Ackermann used scenario planning techniques in strategy building for Scottish Natural Heritage (Eden and Ackermann 1998). The WTO has used scenario planning techniques when dealing with contingency planning (WTO 2004). However, the more widespread use of scenario planning in tourism has not been evident (Yeoman and McMahon-Beattie 2005) and consequently there have been no significant advances in tourism development research and practice. It is argued that the fundamental problem is that written scenarios without support of formal modelling may not be adequate enough to portray the dynamic nature of the change, nor provide managers with a vivid enough picture of the future environment (Georgantzas and Acar 1995; Winch 1999; Forrester 2003; Randers 2005). Compared to conventional scenario analysis approaches system dynamics modelling offers the ability to visualise a dynamic portrayal of possible future developments (Georgantzas 2003), and it employs the twin tools of diagramming techniques in a qualitative manner and quantitative modelling techniques to challenge the current knowledge base (Dangerfield 1999; Dangerfield and Roberts 2000).

Sometimes it is argued that the tourism industry needs constant growth and that maximal amounts of promotion are required to sustain profits and hence jobs. In the current policy context for tourism this may be more aligned with environmental sensitivities than it was 20 years ago and it is also tempered by an increasing mantle of environmental legislation. However, the underlying theme is still that of growth (Buhalis 2000; Bramwell 2003; Sharpley 2004). In this section, a range of price-adjusting policies are examined. It includes changing charter flights, a potential tourist tax and policies for promoting luxury tourism by restricting new budget hotel building.

Changing Charter Flight Arrivals

One of the most aggressive promotion strategies adapted by island tourism authorities in Southern Europe in the past several decades is to support charter flights by subsidising the tour operators for each tourist they send to the islands. An increase in the fraction of charter flight arrivals will certainly encourage further growth of mass tourism. Analyses of the possible impacts are vital for devising appropriate policies for controlling tourism growth and preventing the tourism carrying capacity of an island to be exceeded. Three scenarios are created based on a different fraction of charter flight arrivals, which have been driven by hypothetical policies for these arrivals and assumed to be imposed in month 480 (Fig. 8). Scenario 'RCF Sc1' exhibits a lowered fraction of charter flight arrivals, scenario 'RCF Sc2' is a business as usual scenario and scenario 'RCF Sc3' has the highest fraction of charter flight arrivals.

A higher fraction of charter flight arrivals is usually associated with a higher fraction of package holidays and cheaper accommodation. The simulated result on the total tour expenditure of the three scenarios can be seen in Fig. 9. The scenario 'RCF Sc1' has a lower fraction of charter flight arrivals and thus a higher fraction of scheduled flight arrivals, which indicates a higher expenditure in terms of transportation and associated accommodation expenditure. This scenario has a higher average tour price than the base scenario and, consequently, reduced tourist arrivals. However, it generates the largest tourist expenditure by

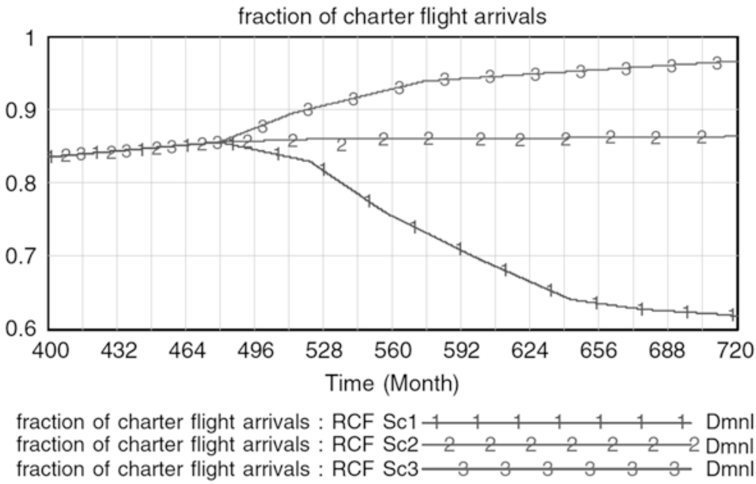


Fig. 8 Three scenarios for the fraction of charter flight arrivals (RCF = Ratio of Charter Flights; Dmnl = dimensionless)

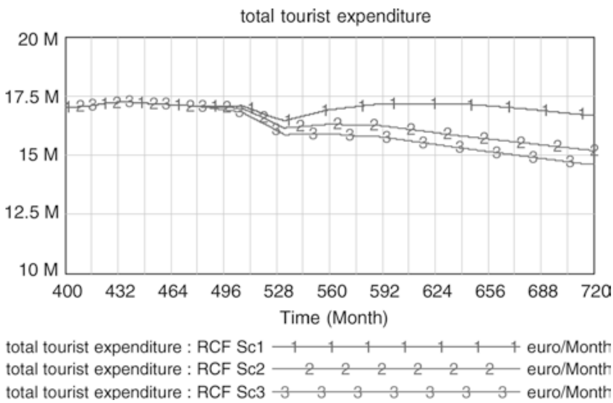


Fig. 9 Impact of charter flight arrival scenarios on total tourist expenditure

compensating for reduced tourist arrivals with a greater margin gained from the higher tour price.

Large tour operators in Europe usually have had a strong influence on the way tourism has evolved, particularly because there are a relatively small number of tour operators at the lower and lower-middle end of the market. Large operators are committed to filling charter flights. This encourages a short-term perspective and allows that market segment to be dominated by customers who holiday abroad because it is cheap, rather than from a desire to experience and appreciate foreign cultures and environments.

A Tourist Tax?

A tourist tax was levied recently in Spain with the intention of mitigating the negative effects from the rapid growth of mass tourism in some of their islands. This had considerable effects on the hotel industry. In April 2001, the Balearic Islands regional government approved Europe's first tourist tax, in spite of opposition from the national government in Madrid and tour operators in the UK and Germany. From early 2002, visitors to Majorca, Ibiza, Minorca and Formentera had to pay an average of 1 euro a day each on checkout if they had been staying in a hotel, hostel, villa or apartment. The intended use of the tax was to fund environmental projects on the Balearic islands (Tremlett 2002).

The tax was unpopular with holidaymakers (particularly those on a budget). It was unpopular with hoteliers who had to collect the tax. It was unpopular with tour operators because they feared for a decline in tourist numbers. Many hotels and villa management companies did not collect the tax and absorbed it into expenses, while some hotels disbursed vouchers that clients could spend on the premises—described by some tourists as the 'Lemonade Tax' (*The Independent* 2003). The tax was collected for the intended environmental purposes, though the impact was not large.

The growing trend of mass tourism in those islands was not deterred by the tourist tax. But hoteliers and tourist firms claimed that the tax on visitors was harming tourism and refused to collect it (*The Independent* 2003). From October 2003, authorities in the Balearic Islands scrapped the tourist tax after local companies rebelled against it.

It is obvious that an integrated approach is required for tourist tax policy analysis. Opposed to the tourist tax policy failure in the Balearics, there is a successful story in Asia. The government of Bhutan in the Himalayas has imposed a tourist tax of 200 US dollars per day on tourists

Table 1 Four tourist taxation scenarios

Tax scenarios	Volume of tourist tax (euro/month/person)	Time of tax levy (month)
taxBase (zero tourist tax scenario)	0	None
taxSc 1 (Low tax scenario, e.g. Balearics tourist tax)	40	120th
taxSc 2 (medium tax scenario)	400	120th
taxSc 3 (high tax scenario, e.g. Bhutan tourist tax)	4000	120th

going into Bhutan (Tourist Authority of Bhutan 2005). This tariff usually covers guides, food and accommodation. This is a radical effort not only to try and reduce tourist numbers but also to increase the revenue coming from tourism—a very successful strategy (Bhattarai et al. 2005). The Bhutan Tourism Authority is emphasising the development of products that are unique to Bhutan. The living culture of Bhutan and eco-tourism are said to be the two main attractions at the moment.

The scope of our model (Fig. 5), while encompassing environmental issues such as water, waste and energy, does not extend to environmental protection, so the impact of any tourist tax is restricted to its effect on visitor numbers. We have analysed the impact on tourist arrivals resulting from the imposition of various rates of tourist tax. The figures in Table 1 present four different scenarios (from zero tourist tax to a high tax). Their effects on the tour price AI, which has an arbitrary scale from 0 to 100, and tourist arrivals follow in Figs. 10 and 11, respectively.

An arbitrary date of the 10th year (month 120) is the assumed date for the introduction of the tax. From the above figures it can be seen that tourist arrivals are very sensitive to the rate of the tourist tax when the tax is above the medium level (400 euro per person per month). However, although a case might be made for a modest tax imposition, in complex social and economic environments, such as island tourist economies, multiple factors need to be considered and dealt with, such as stakeholders' engagement or decision-making based upon demographics. This will involve, *inter alia*, hoteliers, tour operators and the local workforce. In order to avoid conflicts between hoteliers and policymakers over a tourism tax policy, as occurred in the Balearic Islands, mutual consensus between different stakeholders has to be achieved.

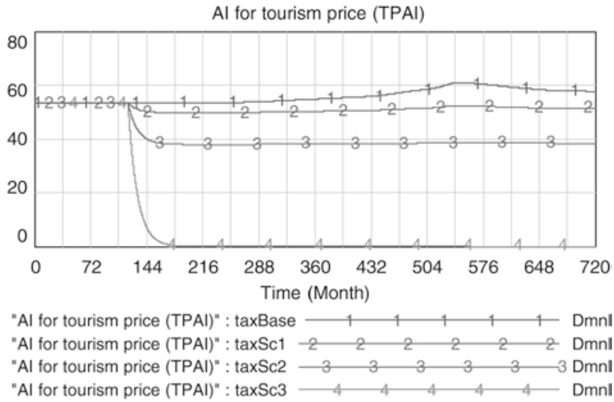


Fig. 10 Impact of tourist taxation on Tour Price Attractiveness Index (TPAI)

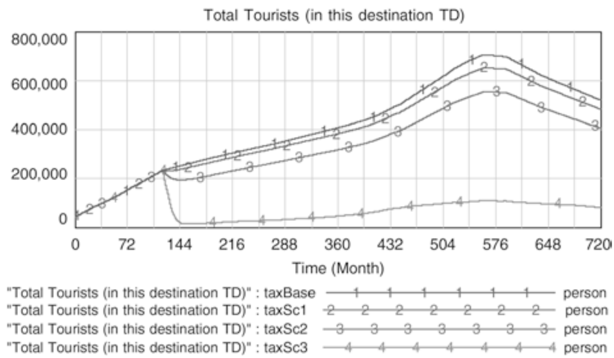


Fig. 11 Impact of tourist taxation on arrivals at a tour destination

For modest tax scenarios the effect on arrivals is temporary and does not ultimately prevent the inexorable rise in the numbers of tourists. Growth hits a peak only when other limits manifest themselves around 35 years later.

Restricting the Construction of New Budget Hotels

Upon realising the possible negative impacts of mass tourism, some South European islands are aiming more at the upper market segments by

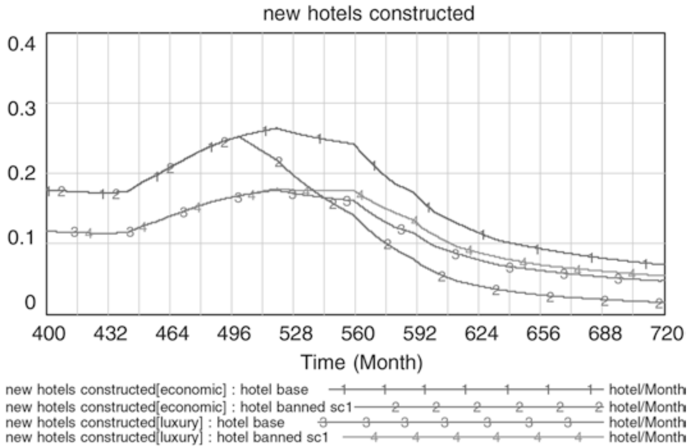


Fig. 12 Completed rates of construction of different hotel types after the policy change in month 498 (half way through 2001)

subsidising new luxury hotel building and restricting the construction of budget hotels. But the effectiveness and possible impact of such a policy on tourism development has barely been studied. In an attempt to evaluate this policy a scenario: 'hotel banned scl'—in which we assume new budget (economic) hotel building is completely banned from the 498th month (equivalent to end-June 2001)—is compared with the base case 'hotel base'.

In the 'hotel banned sc1' scenario, the numbers of new hotels constructed (Fig. 12) takes time to change because of work already in the pipeline, but clearly luxury hotel building increases due to the increased demand for accommodation. Of particular interest is Fig. 13, which shows that it takes a considerable time to change the hotel mix, and hence the economic status of the clientele. For this evaluation, while the absolute numbers can be questioned (around three new hotels per annum), the emphasis is on a comparison between the two policies. In this respect the policy precept is that it takes too long for the effects (of the draconian policy of termination of the construction of budget hotels) to manifest themselves. Growth limits are likely to be reached anyway and this will affect all new hotel construction.

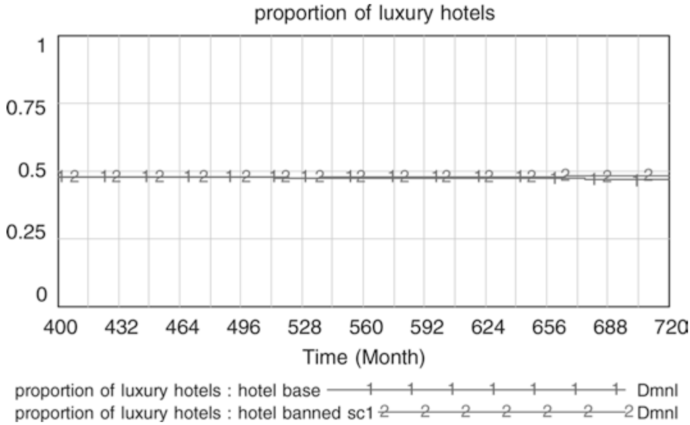


Fig. 13 Proportion of luxury hotels in the accommodation mix

Fig. 14 Effects of pressure on survival in a cheap holiday market

Discussion

It has been observed that the insistence that ‘cheap’ is beautiful has been an illness in the package tourism industry. For too long this industry has suffered from a self-perpetuating cycle of sending more tourists greater distances for less profit. Under these circumstances, companies are more concerned about staying in business than protecting their hosts’ livelihood from unsustainable damage.

This vicious circle is illustrated in Fig. 14, which depicts a positive feedback loop in which more budget-price holidays are impacting on the tour companies’ profits and which in turn exacerbates the pressure. Companies could become insolvent by operating on a too-low price base and then keep going from year to year by simply changing the name of the business. This is in the interests neither of the business nor of the customer, yet it is possible because of free market entry. Some tour operators claim that this problem hinders collective action by operators to increase margins. However, any collective agreements to increase stability or raise margins would be opposed by the Office of Fair Trading in the UK.

Nonetheless, the direction of the vicious cycle must be reversed. Researchers have argued that practices for sustainable tourism offer techniques that can reverse the trend by offering a variable holiday product. An increase in the range and quality of holidays should be associated with greater margins and the chance to compete on more sustainable resources than just price. Evidence suggests that this could be a long-term opportunity for operators to add value to the service they provide. A genuinely sustainable approach to tourism should have benefits for all. For those involved in the industry it means long-term profitability and a need to avoid a potential boom-and-bust. As we have shown, there are always limits to growth. If some island tourist destinations become over-dependent on tourism—even over many decades—they may eventually experience sharply contracting visitor numbers (and profits) thereby destroying their original attractions. An example is the boom and bust tourism development in mainland Southern Spain during the 1980s (Forsyth 1996).

Sustainable tourism development, compared to the practice of price-cutting, has a number of benefits for all the stakeholders, such as:

- Adding value to holiday packages by offering more to tourists than the standard sun, sea and sand.
- Cutting costs by recycling waste products and reducing unnecessary fuel consumption.
- Active involvement with local authorities and communities by liaising with industry suppliers to provide products that support local industries and avoid environmental damage.
- Sustainability is not peripheral to the tourism industry, but is in fact central to breaking the downward spiral of sending more and more tourists greater distances for less profit.

However, enforcing sustainable tourism policies is extremely difficult. Policymakers and destination managers need to formulate an integrated public-private partnership and develop opportunities for understanding by all stakeholders in order that they might realise the importance of maintaining a fairly standardised pricing structure and policy. Furthermore, sustainable tourism does not come from imposing polices alone, or from simple checklists or isolated initiatives. Rather, it depends on an insightful understanding of exactly how the tourism system functions and interacts through time with the other industries in which it operates. Put simply: there is a need for a systemic approach to tourism policy with the purpose of surfacing the often conflicting actions of the various system stakeholders who are driven by their own missions and goals. If policy remains as un-coordinated as at present, then the likelihood is of a 'Tragedy of the Commons' scenario (Hardin 1968) and the possible demise of one or more island tourist economies.

Conclusions

Sustainable tourism development problems are replete with nonlinearities, feedback and considerable complexity. However, it is impossible to prove that any simulator that aims to tackle this complexity is a correct or 'true' model of the real system at the time of modelling. As yet, no European island has experienced boom and bust, but that behaviour is certainly feasible if the tourist destinations are not managed wisely. Tourism dynamics, as shown in this paper, provide a warning sign that such behaviour is feasible. Sustainable development models cannot be validated by any one test such as their ability to fit historical data. A good fit to data during the growth phase says nothing about the timing and magnitude of any incipient peak and eventual decline in tourist arrivals. Model testing should be regarded as the process of bringing the user's confidence to an acceptable level such that any policy inference about the system, derived from running the model, is one upon which a high degree of reliance can be placed.

Sustainable development problems are system problems: the solutions must involve looking at the impact of changes on as much of the system as possible. Partial solutions are likely to be ineffective and may even make things worse. Changes throughout the system must be co-ordinated. It is not sufficient for individual units to change without understanding the impact such changes will make on other parts of the system. In order to formulate sustainable tourism development policies, the following three steps have to be taken:

1. To conceptualise a whole-system picture that captures the most important variables and interrelationships.
2. To carry out a detailed analysis based on integration of hard and soft data and methodologies.
3. To improve stakeholder engagement and participation in policy analysis and policymaking.

The artefacts developed in this research (the generic tourism system dynamics model, policy scenarios and the microworld—see Xing 2006) provide a powerful means to enhance the accomplishment of the three tasks identified. The tourism system dynamics model described in this paper identifies a number of essential feedback structures that have profound effects on sustainable tourism development. Rather than providing a forecast of a predetermined future, the model develops a means of testing alternative scenarios for policy analysis and stakeholder collaboration. This research focuses on the evolution of a holistic framework together with a generic model for achieving the aspiration of the sustainable development of (particularly island) tourism.

From a methodological perspective this research shows that system dynamics provides a way of visualising tourism as a network of integrated systems, including demographic, cultural, economic and energy, while rigorously inferring their performance through quantification and the use of computer simulation. System dynamics modelling can serve as a vehicle for integrating multiple data resources and multiple methods from marketing, finance, operations and other functional spheres of tourism. Explicit mapping and analysis of feedback in a system dynamics model reveals an intuitive grasp of dynamics and enhances the quality of debate while eliciting new knowledge. It keeps us thinking hard about how to design the future.

Future work can be drawn out from the analysis presented in this paper in several ways. First, the generic model can be parameterized to represent the dynamics of tourism development for a particular island more precisely. Second, the generic model and modelling process presented in this paper can be applied to sustainable development analysis for other industries. Third, the analysis of sustainable tourism development can be integrated with other long-term sustainable development modes, such as sustainable urban development or sustainable regional development.

Acknowledgements Thanks are due to the University of Salford, UK for funding this research.

References

Bhattarai K, Conway D and Shrestha N (2005). Tourism, terrorism and turmoil in Nepal. *Ann Tourism Res* **32**: 669–688.

Bramwell B (ed). (2003). *Coastal Mass Tourism: Diversification and Sustainable Development in Southern Europe*. Clevedon: Channel View.

Bramwell B and Lane B (1993). Sustainable tourism: An evolving global approach. *J Sustainable Tourism* **1**: 1–5.

Buhalis D (2000). Marketing the competitive destination of the future. *Tourism Mngt* **21**: 97–116.

Crompton JL (1979). Motivations for pleasure vacation. *Ann Tourism Res* **6**: 408–428.

Dangerfield B (1999). System dynamics applications to European health care issues. *J Opl Res Soc* **50**: 343–353.

Dangerfield B (2008). System dynamics advances strategic economic transition planning in a developing nation. In: Qudrat-Ullah H, Spector M and Davidsen P (eds). *Complex Decision-Making: Theory & Practice*. New York: Springer, pp 185–209.

Dangerfield B and Roberts C (2000). A strategic evaluation of capacity retirements in the steel industry. *J Opl Res Soc* **51**: 53–60.

Eden C and Ackermann F (1998). *Making Strategy: The Journey of Strategic Management*. London: Sage.

Forrester JW (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press, now available from Pegasus Communications, Waltham, MA.

Forrester JW (2003). Economic theory for the new millennium. In: Eberlein R (ed). *Proceedings of the International System Dynamics Conference*. New York: System Dynamics Society (CD-ROM).

Forrester JW and Senge P (1980). Tests for building confidence in system dynamics models. In: Augusto AL, Forrester JW and Lyneis JM (eds). *TIMS Studies in the Management Sciences*. Vol. 14, Oxford: North-Holland, pp 209–228.

Forsyth T (1996). *Sustainable Tourism: Moving from Theory to Practice Prepared by Tourism Concern. World Wide Fund for Nature (UK)*, England: Godalming.

Georgantzas N (2003). Tourism dynamics: Cyprus' hotel value chain and profitability. *Syst Dynam Rev* **19**: 175–212.

Georgantzas N and Acar W (1995). *Scenario-Driven Planning: Learning to Manage Strategic Uncertainty*. Westport, CT: Quorum Books.

Hall M (2000). *Tourism Planning: Policies, Processes and Relationships*. Essex, UK: Pearson Education Limited.

Hardin G (1968). The tragedy of the commons. *Science* **162**: 1243–1248.

Homer JB and Keane TE (1999). Evaluating strategies to improve railroad performance—A system dynamics approach. In: Farrington PA, Nembhard HB, Sturrock DT and Evans GW (eds.), *Winter Simulation Conference*. Arizona: Proceedings IEEE, pp 1186–1193.

Hunter C (1997). Sustainable tourism as an adaptive paradigm. *Ann Tourism Res* **24**: 850–867.

The Independent (2003). Balearic isles scrap eco-tax on tourists. 30 October, London, p. 15.

Jamal TB and Getz D (1995). Collaboration theory and community tourism planning. *Ann Tourism Res* **22**: 186–204.

OECD (2001). *Tourism*. Vol. 2003. OECD, http://www.oecd.org/department/0,2688,en_2649_34389_1_1_1_1_1,00.html, accessed January 2004.

Parry M and Carter T (1998). *Climate Impact and Adaptation Assessment*. London, UK: Earthscan Publications.

Pearce D and Butler R (1993). *Tourism Research—Critiques and Challenges*. London: Routledge.

Randers J (2005). 30 years beyond the limits: An interview with Jørgen Randers: Co-author of 'The Limits to Growth'. In: 'News and Analysis for Futures Studies & Scenario Planning'. *Plausible Futures Newsletter*. http://www.plausiblefutures.com/2007/04/30-years-beyond-the-limits-an-interview-with-jørgen-randers/, accessed September 2010.

Repenning N (2001). Understanding fire fighting in new product development. *J Prod Innovat Mngt* **18**(5): 285–300.

Richardson GP and Pugh AL (1981). *Introduction to System Dynamics Modelling with DYNAMO*. Cambridge, MA: MIT Press, now available from Pegasus Communications, Waltham, MA.

Ringland G (1998). *Scenario Planning—Managing for the Future*. Chichester, UK: Wiley.

Roberts L and Simpson F (1999). Developing partnership approaches to tourism in Central and Eastern Europe. *J Sustainable Tourism* **7**: 314–330.

Sharpley R (2004). Tourism, modernisation and development on the Island of Cyprus: Challenges and policy responses. In: Bramwell B (ed). *Coastal Mass Tourism—Diversification and Sustainable Development in Southern Europe*. Clevedon: Channel View Publications, pp 321–340.

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

Tourist Authority of Bhutan (2005). Official tourist information. Tourism Authority of Bhutan. http://www.tourism.gov.bt/, accessed December 2005.

Tremlett G (2002). Green tax helps islands clean up. *The Guardian*, May, London, p. 11.

Van der Heijden K (1996). *Scenarios: The Art of Strategic Conversation*. Chichester: John Wiley & Sons.

Wang Y and Fesenmaier DR (2007). Collaborative destination marketing: A case study of Elkhart county, Indiana. *Tourism Mngt* **28**: 863–875.

Warren K (2002). *Competitive Strategy Dynamics*. Chichester: Wiley.

Weaver DB (1998). Introduction to ecotourism. In: Weaver DB (ed). *Ecotourism in the Less Developed World*. Oxon: CAB, pp 1–33.

Winch G (1999). Dynamic visioning for dynamic environments. *J Opl Res Soc* **50**: 354–361.

World Commission on Environment and Development (1987). *Our Common Future*. Oxford, UK: Oxford University Press.

WTO (2003). Long-term prospects: Tourism 2020 vision. World Tourism Organisation. http://www.world-tourism.org/facts/eng/vision.htm, accessed January 2006.

WTO (2004). *Indicators of Sustainable Development for Tourism Destinations: a Guidebook*. Madrid: World Tourism Organization.

Xing Y (2006). *Exploring the sustainability of mass tourism in island tourist economies: a system dynamics approach*. PhD thesis, University of Salford, Salford, UK.

Yang J-T (2007). Knowledge sharing: Investigating appropriate leadership roles and collaborative culture. *Tourism Mngt* **28**: 530–543.

Yeoman I and McMahon-Beattie U (2005). Developing a scenario planning process using a blank piece of paper. *Tourism Hospitality Res* **5**: 273–285.
Yong Y and Keng K (1989). A Delphi forecast for the Singapore tourism industry: Future scenario and marketing implications. *Eur J Marketing* **23**: 3546.

Modelling for Policy Assessment in the Natural Gas Industry

Y. Olaya and I. Dyner

Introduction

The Colombian energy policy aims to increase the security and sustainability of energy supply. To attain these goals without increasing government participation, the energy policy promotes competition in the market place. Competition and efficiency of markets depend to a great extent on the possibility of substitution.

Journal of the Operational Research Society (2005) **56**(10) 1122–1131.
https://doi.org/10.1057/palgrave.jors.2601895
Published online 8 December 2004.

Y. Olaya (✉)
Departamento de Ciencias de la Computación and Decisión, Faculted de Minas, Universidad Nacional de Colombia, Medellín, Colombia

Carrera, Bogota

I. Dyner
Energy Institute, Universidad Nacional de Colombia, Medellín, Colombia

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_11

Natural gas can play a decisive role in increasing the efficiency of energy markets and energy use. This relatively abundant and environmentally friendly fuel is a good substitute for some residential uses and for transportation and power-generation, and can therefore help in many instances to increase market efficiency and reduce emissions. How consumption of natural gas would be affected by its penetration in new markets is a question that links the behaviour of consumers and industry with the operation of the gas system and the evolution of gas reserves and other energy sub-sectors.

Despite the considerable potential of proven and probable reserves, and despite the government's plan to expand natural gas consumption that started in late 1980s, natural gas markets in Colombia are just emerging. Among the barriers that prevent further expansion of gas markets are the limitations on transportation capacity, and the imperfect knowledge of demand behaviour, historical production, reserves and investments. In addition, the development of the natural gas market is affected by prices and regulations in other sectors, such as the fuel market. Policy makers, investors and regulators face thus the problem of evaluating alternative strategic decisions under conditions of complexity and uncertainty. For them it is important to have a perspective on the integrated systems in order to examine the impact of their decisions over some of its components. Complexity of policy analysis calls for the understanding of behaviour and decisions via system simulation.

Natural gas is a non-renewable resource; the initial growth of production and discoveries is therefore followed by depletion. Exhaustibility of natural gas is an issue for policy makers in a long-term analysis, but in the short term it is the market operation what influences the decisions of producers and consumers, and it is the market dynamics what determines the shape of the natural gas extraction path. The system operation, which is largely influenced by factors such as transportation capacity, inter-fuel competition and regulatory conditions, has a strong effect on the extraction rate and on the depletion of the source. In efficient markets, prices reflect scarcity and provide signals for the development of new reserves and technology innovation.

Many simulation models for the natural gas industry consider depletion of resources and the sector's responses to technology changes in order to shape the long-term industry development and behaviour. Fossil II [1] is a system dynamics model with life cycles of natural gas and petroleum. It captures the long-term dynamics of discovery, production and depletion as well as the transition of both industries from conventional to non-conventional resources and technologies.

Unlike petroleum, natural gas transportation is expensive and highly dependent on economies of scale. In addition, prices for natural gas are often determined in local rather than in national and international markets. Exploration activities and discoveries of gas and petroleum, however, are often related. System dynamics models for petroleum industries [2, 3] simulate the evolution of reserves and hydrocarbon production, considering the fact that technological advances reduce the cost of finding and producing petroleum.

Other models, such as the Gas System Analysis Model of the US Department of Energy [4, 5], integrate different methodologies, thereby allowing the evaluation of decisions for both the short and medium term. All these models were developed for mature markets, where historic data on reserves, discoveries, prices and industrial activities are abundant. None of these conditions apply to the Colombian case.

System simulation, in the form of both micro-worlds and stand-alone models for problem solving, has been successfully applied in a variety of areas [6, 7]. In the energy sector, there are a number of applications for policy support [8–13], as well as for improving the understanding of the market and its participants [14, 15].

In Colombia, there are instances of simulation modelling, based on system dynamics, for policy support in the energy industry. Issues such as rational energy use and in particular, gasoline substitution in urban transportation [16], and efficient use of energy in the residential sector have been studied and modelled [17, 18]. As a consequence of the complexity of energy systems, those models integrate different components and methodologies for strategy and policy support, using varied information quality, perspectives and scope.

This paper shows the need of modelling support for policy assessment in the Colombian natural gas sector. It methodologically synthesizes optimization and system dynamics in order to address issues related to industry sustainability. The next section is dedicated to the discussion of the system dynamics model for the natural gas market, emphasizing the main causal relations that drive the evolution of reserves and price. This is followed by two sections where we first present the optimization model used to simulate the gas network operation, and later the integration of the general market model with the gas transportation model, creating a tool for policy analysis. Model evaluation and validation is presented next. Then there is a discussion of results of simulations and their implications for policy analysis. Finally, we present the conclusions.

Modelling the Natural Gas Industry

Existence of proven reserves alone does not guarantee the creation of a market for natural gas. In the Colombian case, only when government invested in the transportation system and involved private participation in the construction of distribution networks, did the market for natural gas eventually emerge.

The expansion in demand and production of natural gas brings in new customers such as thermal generation plants and the domestic demand of the urban population. The increased number of actors and transactions, along with the deregulation of the market and the limitation of transportation capacity, generate additional uncertainty in the security of the natural gas supply.

Although the intensity of exploration activity in Colombia is low, there have been significant gas discoveries in four of the five most-explored sedimentary basins. Exploration is mainly undertaken by private companies in association with Ecopetrol, the Colombian state-owned company, which, having little participation as an exploration company itself, has oriented its policies towards the creation of incentives for foreign investment in this activity.

In this section, we will explain the system dynamics model that has been developed to understand the behaviour of the natural gas industry in Colombia. The long-term relationships are based on Naill's models [1, 19, 20]. Natural gas depends on the interaction between the exploration industry, transportation and customers as well as on proven reserves.

In a deregulated market, this interaction results in the formation of the market price that will drive the future supply and demand.

In the following subsections, we will present the basic components of the system dynamics model. We explain the dynamics of discoveries and the influence of technology over industry costs. Both components are then coupled in a general market model.

Dynamics of Discoveries

There are various methods for estimating and classifying the volume of petroleum or gas in a sedimentary basin; differences in assumptions, geological knowledge and evaluation criteria cause estimates of reserves to vary across methods and time [2, 21, 22].

In Colombia, the inventory of historical natural gas production, consumption and reserves is not complete; consequently, a detailed description and quantification of natural gas reserves is not possible. We divided reserves into two broad categories: proven reserves, which are the estimated volumes that can be economically produced under the existing technical and economic conditions; and probable reserves, which are an estimate of the remaining undiscovered resources.

The initial volume of gas identified as proven reserves is obtained from engineering and geological data acquired during exploratory drilling. Usually, this estimate of proven reserves is conservative and, during the development stage of a field, the geological and fluids data from new wells are used to recalculate the reserves. In a later stage, information from production is used for the validation and revaluation of proven reserves.

As shown in Fig. 1, new discoveries increase the amount of proven reserves but production reduces them. Proven reserves are reduced at a rate that depends on production, while new reserves are added as a result of price, investment and exploration. Note that investment increases transport capacity, which contributes to increases in production rates.

Figure 1 shows that discoveries increase the amount of proven reserves, and those reserves decrease as they are produced at a rate depending on

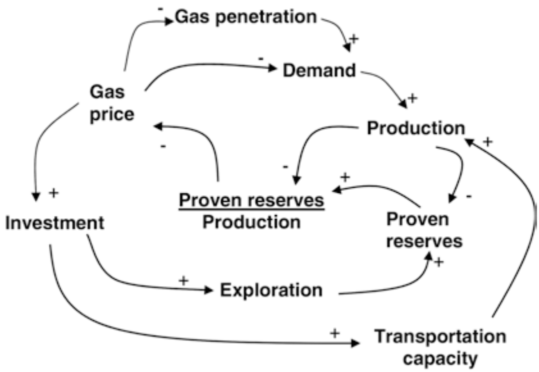


Fig. 1 Natural gas reserves development-cycle

the demand. Physical exhaustion does not always occur, for extraction always leaves a remaining fraction of natural gas trapped underground, and because fields are abandoned when their marginal production cost is higher than the market price. Scarcity reflects higher gas prices, which in turn affect demand. While non-captive customers can switch to other fuels, captive markets may reduce their consumption. At the same time, higher prices provide incentive for investment in new technologies and in new exploration projects, which in turn increase reserves.

The dynamics illustrated in Fig. 1 help to explain how energy policy would work. As mentioned above, one of the strategies to increase market and energy efficiency is to promote competition and substitution. Increments in gas prices reduce the penetration of gas in other markets, through substitution, and in turn affect demand. Depending on the marginal cost of gas production, and on the elasticity of demand, the policy of gas price deregulation can either increase or decrease the share of gas consumption.

Operation and capacity of the transportation network can also be the target of policies aimed to increase competition, such as open access to the transportation facilities, or transportation pricing and expansion. The impact of some transportation policies will be analysed in the Transportation section; but as Fig. 1 shows, transportation policies affect the access of suppliers to the gas market, and the production capacity, promoting competition and price reduction.

The following sections explain how the relationships and the dynamics depicted in Fig. 1 are modelled. Penetration of natural gas in the residential sector is modelled in terms of the number of new households connected to the distribution networks. Demand growth in power generation depends on the new thermoelectricity plants being built. The vehicular demand for natural gas, NGV, is estimated using a separate substitution model for the transportation sector and is added to the total demand.

In Fig. 1, price is the only factor affecting exploration investment. In actual decisions, however, investors evaluate also factors such as costs, technology, geological potential, infrastructure, and economic and political risk, some of which are incorporated in the model as indicated in the following sections.

Costs and Technology

The cost of gas depends on the value of the exploration, production and transportation—infrastructure costs. Exploration cost can be defined as the investment needed to find a unit of the resource. This cost is usually lower at the beginning of the exploration activities, when the most accessible and economical fields are found, and tends to increase as the cumulative discoveries approximate to the estimated value of probable reserves [24, 25]. Maturity of exploration is then associated with higher exploration and discovery costs.

The cost of finding and producing a unit of natural gas varies according to the physical properties of the reservoir as well as the nature of the technology, infrastructure and taxes in the region. Although depletion of reserves is associated with higher costs of exploration and production, it has been noted that technology progress can change the shape of the depletion curve.

Technology improvements have allowed the development of non-conventional resources at competitive prices, and at the same time, have increased the efficiency and productivity, as illustrated in Fig. 2. Technology has not only increased general energy intensity, but has also helped to reduce the total costs of the industry [25–27].

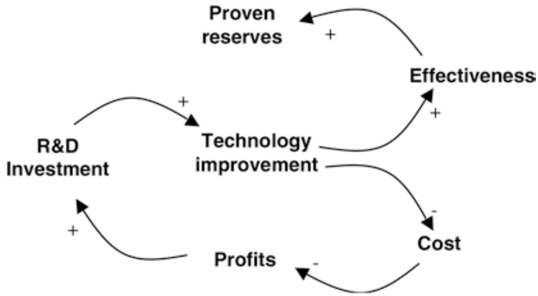


Fig. 2 Influence of technology in the natural gas industry

Berg proposes function (1) for estimating the cost of exploration, G, depending on: a depletion parameter, γ; the amount of remaining reserves, $(D-D_0)$; time, t; and, a parameter representing technological progress in exploration, δ.

$$G = \beta e^{\gamma(D-D_0)-\delta t} \tag{1}$$

Similarly production cost depends on the cumulative production and the technological advance, resulting in investment in research and development. Berg et al. [25] proposes function (2) for estimating the costs of producing oil, which also describes the behaviour of gas production costs:

$$C = C_0 e^{(\eta A-\tau t)} \tag{2}$$

where, C_0 is the initial cost of producing a unit of resource, A is the cumulative production, τ is the rate of technological advance and η is a depletion parameter.

Figure 2 shows that investment in research and development will eventually reduce exploration, production and operation costs, increasing profits and contributing to further investments. Costs and investments, in turn, are linked with the dynamics of fuel substitution and further exploration that were illustrated in Fig. 1.

Investments in the hydrocarbon industry are usually large and fields are often located in areas of poor infrastructure and difficult drilling conditions. In addition, exploration for natural gas is a high-risk activity, because of the large costs involved and the geologic uncertainty of the reserves. Profits drive the industry; therefore, if expected prices and demand are likely to compensate spending on operations, taxes and royalties, and there is a good geologic potential, companies would consider investment.

Energy requirements depend on social, cultural and legal factors but, in general, they change with population growth, economic development and technological progress. For many years, consumption of natural gas in Colombia was restricted to industrial and power generation facilities located near producing fields. Building the transportation network was the first step towards developing other markets for natural gas, such as residential or vehicular.

Investment in both exploration and expansion of production and transportation capacity is of great importance to support the system's evolution. The discovery and development of potential reserves are not possible without continued investment.

We disaggregated both supply and demand of natural gas into well-defined regions, corresponding to the main production fields and consumer centres, in order to understand issues related to the transportation network. The following section investigates the specification of a gas transportation model and this is followed by a section where we addresses the issue of coupling it with the market model previously discussed, for policy assessment purposes.

Transportation

As discussed before, the inappropriate transportation infrastructure in Colombia confined the influence of natural gas to areas near production sites for nearly 20 years. During the 1990s, construction of gas pipelines began with the objective of connecting large urban centres with the production fields. Private companies, under government concessions, are undertaking the construction of both pipeline networks and urban distribution systems.

Power generation and other thermoenergy-based industries are the main consumers in the growing natural gas markets. In recent years, the residential sector has increased its share of total demand, while the participation of the compressed natural gas sector is still insignificant. Growth of those potential markets is highly dependent on the gas transportation system.

The balance between production and demand takes place in the pipeline. Producers inject their gas at the connection nodes. Gas flows through the distribution sub-system and it is delivered to customers, according to agreements. Production capacity depends upon the available reserves and the transportation capacity. If the pipeline capacity is insufficient, or its operation is inefficient, exploration, reserves and demand are affected. In addition, increasing demand may be a signal for investment in more gas transportation capacity.

If distribution of production is optimally operated, as may be intended, a mathematical program can calculate the transported volumes with minimal production and transportation costs. An optimal solution will satisfy regional demands and the restrictions in the volumes that can be carried through the different pipelines of the network. The proposed model should seek to minimize the total costs of production and transportation with the restrictions imposed by demand and pipeline capacity, as follows:

Minimize:

$$\sum_{i=1}^{np}\sum_{j=1}^{nd} PC_i x_{ij} + \sum_{i=1}^{np}\sum_{j=1}^{nd} TC_{ij} x_{ij} \tag{3}$$

subject to:

$$\sum_j x_{ij} - \sum_j x_{ji} = b_i \tag{4}$$

$$0 \le x_{ij} \le u_{ij}, \forall i, j \tag{5}$$

where PC_i is the production cost per unit at node i, TC_{ij} the transportation cost per unit from i to j, x_{ij} the gas flow from i to j, u_{ij} the capacity, or upper limit to, flow from i to j, b_i the net gas flow, np the number of demand nodes, j, and nd is the number of supply nodes, i.

The optimal production–distribution dispatch plan, found with this optimization model (3, 4, 5), provides a clear signal for production and transportation expansion. In the following section, we discuss why this model and a vehicular demand model for natural gas may be coupled with the system dynamics model previously discussed.

Integration of Models and Policy Analysis

As energy costs and availability play an important role in economic development, policy makers seek for the appropriate utilization of energy resources. All economic activities use different sources of energy that are often technically interchangeable at a reasonable cost. This fact implies that the policies applied in one energy sub-sector can influence the behaviour of other consuming or producing sectors, and that there is a need for tools to evaluate decisions and assess their impacts.

During recent years, the importance of the Colombian natural gas sector has grown. The big discoveries of the last decade and the construction of the gas network drove the expansion of demand and production. Future development of the sector will depend on the ability of natural gas to substitute the use of other energy sources, such as gasoline in transportation, or electricity in cooking. Models for policy analysis must then account for the complex relations within the gas industry and for its interactions with other sectors.

We have previously described the behaviour of the reserves and market in the natural gas sector and the system dynamics model built to simulate it. Although this model captures the main feedback relations of a mature market, it is not well suited to analyse specific policies for new markets. In particular, it cannot answer questions such as how to operate the transportation network in order to secure supply for the demand regions, or what the impact of a vehicular gas program would be.

Integration of models, like those described in previous sections, provides a solution to the problem of evaluating the cross-impacts of policies. The term integration can be understood from the point of view of control, information exchange or interaction with analysts. With respect to control, the integration accounts for the level of knowledge that models capture from other models. The aspect of information exchange has to do with access to models' data, and refers to the interaction between analysts and models.

Figure 3 shows some of the aspects of model integration for the natural gas sector. Each one of the boxes in Fig. 3 corresponds to one of the integrated-model components, and the arrows indicate that there are information flows between them. For example, the 'Environment' component calculates the impacts from natural gas consumption using as inputs the demand calculated in the 'Natural gas demand' and 'Urban transportation' modules. Investment in exploration is calculated in the 'Exploration and production' module, using the industrial profits calculated in the 'Market' module. The components for gas exploration, market, demand and pipeline system are the ones discussed in this paper. The urban transportation model [16] is an aggregated model for the analysis of demand and substitution in the vehicular sector. The environment model accounts for the emission of pollutants in the power generation and urban transportation sectors.

Macroeconomic indicators, such as projections for GNP and gasoline prices, are some of the input data for the models. Other inputs are the initial values of reserves, industry costs and pipeline capacity. Exploration, demand and transportation of natural gas interact in the market module, and the resulting price determines the yield of investment in exploration, the substitution of natural gas in industrial sectors and the substitution of gasoline for gas in urban transportation.

In the pipeline model, an optimal production–dispatch plan is sought, based on the demand and capacity restrictions. The production plan feeds the exploration/production model, and the market model receives information about gas volumes dispatched to demand subsystems; unsatisfied demands provide a signal for pipeline capacity expansion. Total consumption of natural gas and gasoline are the input data for the environment component, which calculates pollutant-emissions.

Models for the natural gas market and demand are embedded in the model for exploration and production. The urban transportation model provides the information for gas demand and the calculation of emissions,

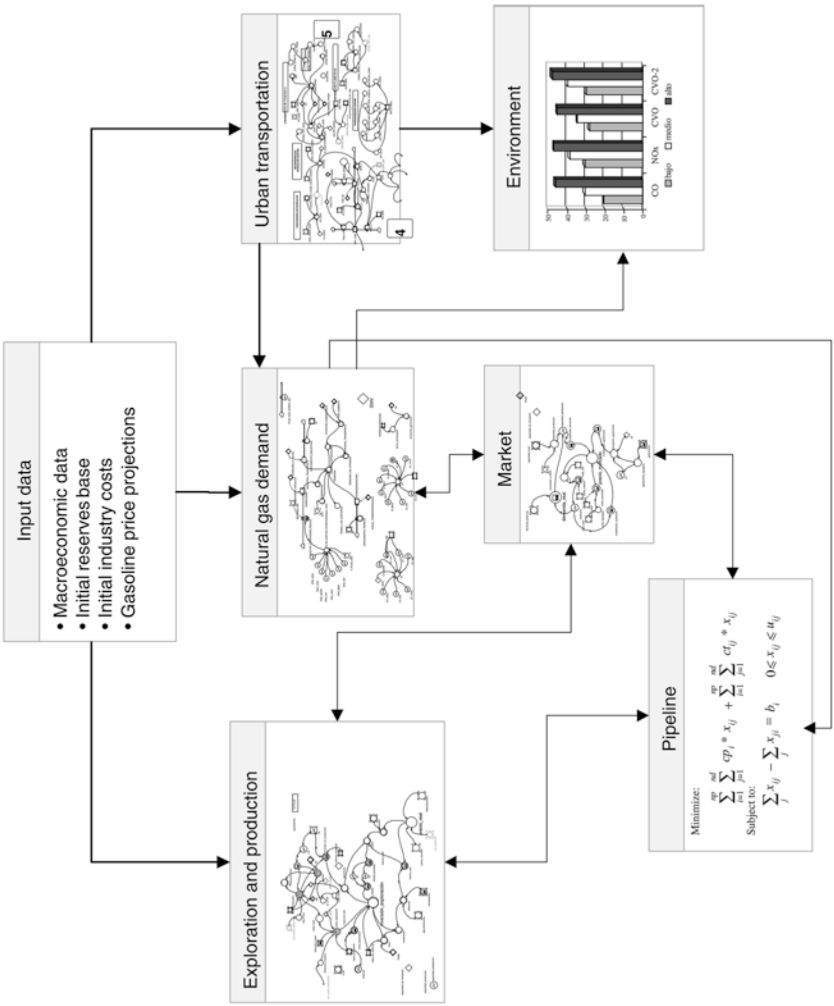


Fig. 3 Integrated model with components

which are based on volumetric emission factors. The gas network optimization model is an independent component that interacts with the market component every time-step (but analysts may choose how often they want it) during each simulation run. An interface controls the integration and execution of the models.

Having explained the policy-assessment environment for the investigation of certain crucial issues, related to market operation within the natural gas industry, we now turn our attention to evaluating the model for policy analysis.

Model Validation

We evaluated model consistency by comparing simulation results of the evolution of reserves with actual historic data for the period 1980–1997. Historic data of proven reserves are aggregated, inaccurate and do not account for flared production but, overall, they follow the actual trend of natural gas reserves in Colombia.

Figure 4 compares the actual proven reserves with those obtained via model simulation. Although model reserves grow as historic reserves did, there are significant differences between both trajectories. The system dynamics model assumes that once investments are made, continuous

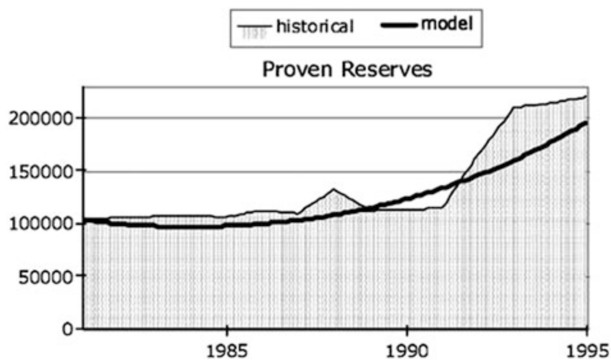


Fig. 4 Comparison of historic proven-reserves with those obtained via model simulation

addition of reserves will take place. But the fact is that discoveries follow a random process and finding a giant field, like those discovered in Colombia in the early 1990s, cannot be predicted with certainty.

During the period 1980–1997, gas production reflected the surge in proven reserves. In the simulated period, demand of natural gas grew slowly and was limited to regions in the vicinity of the production centres. The discovery of Cusiana and Cupiagua added near 100000Mm^3 of gas to proven reserves and the reserves–production ratio almost doubled, as shown in Fig. 5.

The randomness associated with exploration reduces the predictive capacity of the model, and can only partly explain the difference between the system dynamics model and actual behaviour. Another factor that can affect results is the fact that the estimated amount of reserves in place is dynamic, and also subject to uncertainty. It has been found [22] that most of the new proven-reserves additions in the US result from revaluation of reserves rather than from new discoveries. Initial geological knowledge, field complexity, technology and economic environment are some of the factors that Morehouse [22] indicates as causes of the reserves-growth phenomenon.

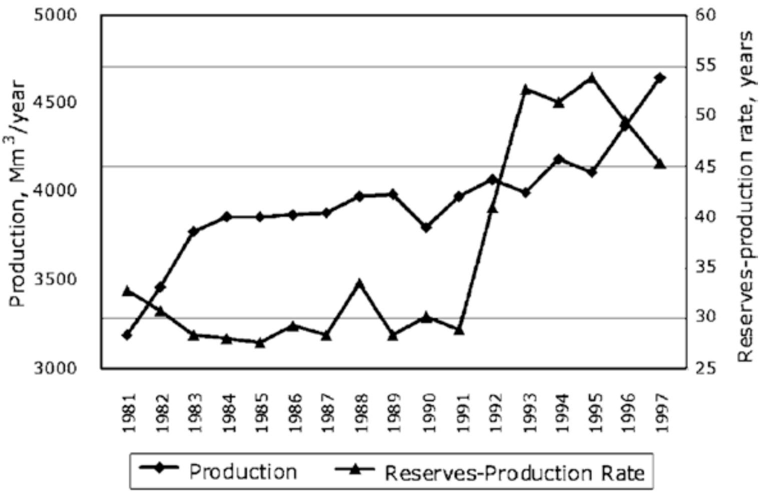


Fig. 5 Historic production of gas

To account for the uncertainty regarding discoveries, appropriate scenarios have been built into the model. In this way, we will be able to analyse market behaviour under some extreme conditions. Under a scenario approach, which has been chosen to deal with uncertainty at different levels, we can proceed by evaluating some important alternative policy issues. The policies that have been investigated, and which we report in the following section, include: substitution of gasoline by gas in the ground-transport sector, its positive impact on the environment, and issues related to the expansion of the pipeline system.

Policy Assessment

The promotion of the use of natural gas in ground transportation and the operation of the transportation networks have been proposed as policies that would increase competition and market efficiency in the energy sector. The UPME and other government agencies have estimated the viability of these policies. A detailed cost–benefit analysis of such policies is beyond the scope of this study; our objective is to provide a basis for cost–benefit and other policy analysis by quantifying a variety of policy impacts, such as natural gas penetration and emissions reduction. Although the results of the model can be used to estimate costs and benefits of policies directly, we emphasize the ability of the model integration to evaluate the secondary and lagged effects of policies, and to illustrate the long-term effects of decision making.

Gas Demand and Substitution of Gasoline in Transportation

Substitution of gasoline by natural gas in the transportation sector has been proposed as a policy to increase energy efficiency, and at the same time, to control the emissions of gases in cities. It is expected that the natural gas market will then grow as a result of a substitution program for transportation.

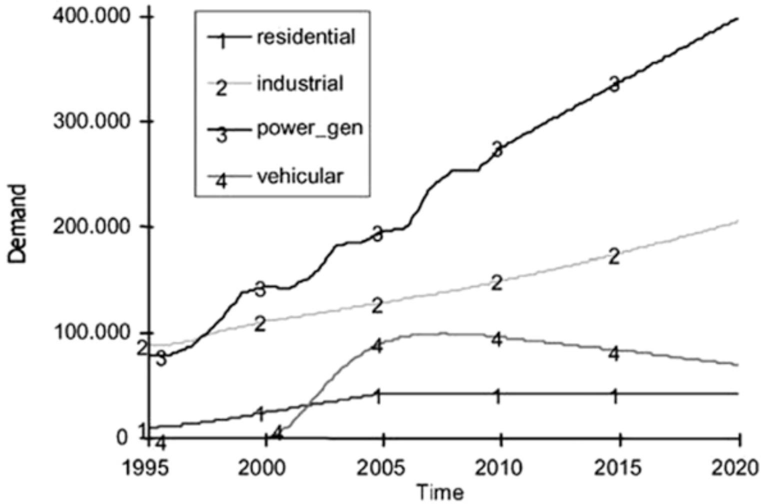


Fig. 6 Simulated natural gas demand in all sectors

Figure 6 shows the simulation results for the aggregated natural gas demand in all sectors. The demand of the urban transportation sector may have an important impact on the natural gas market during the initial years. In addition, Fig. 6 indicates that electricity generation is expected to be the most active sector for expansion of natural gas demand, followed by other industry demands and urban transportation, leaving the residential sector as the smallest. Demand for automobiles and transport, is very sensitive to income and technology changes. The decline of demand for gas in the transportation sector is the result of assumptions that have been made with respect to technology changes and the evolution of GDP.

The increases of gasoline price provide an indisputable signal for the substitution of this fuel by others, among which gas is one alternative; results indicate that 20% of conventional vehicles would participate in a natural gas conversion programme.

Natural gas can substitute many different fuels. In the transportation sector, natural gas can compete with gasoline, whose price is determined

Table 1 Price scenarios for substitution

	Scenario
NG priceis30% <average substitute price	1
NG price is 40% >average substitute price	2
NG price is similar to average substitute price	3

NG stands for natural gas

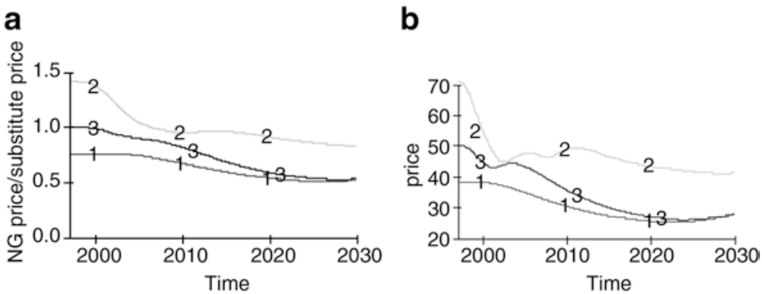


Fig. 7 Evolution of relative and absolute natural gas prices under alternative scenarios

in global markets. In the power generation market, natural gas competes with oil products, coal and hydropower. In the proposed model, the natural gas industry perceives an initial average price both for gas and its substitutes; the scenarios for the relative prices of substitutes are shown in Table 1.

Figure 7a illustrates simulation results of relative prices under the three scenarios of Table 1; Fig. 7b shows the evolution of gas prices under the above scenarios. Line 1 in Fig. 9 corresponds to the scenario in which the initial price of gas is lower than that of its substitute. In this case, the only factor that affects price is the reduction of costs caused by technological progress (R&D). When there are no big differences between natural gas and substitutes' prices, gas prices will behave as in line 3, and will tend to reflect the cost evolution. If price of natural gas is substantially higher than those of its substitute, as in line 2, a large portion of demand will switch to the substitute and price of gas will fall, until it stabilizes at a fairly high level.

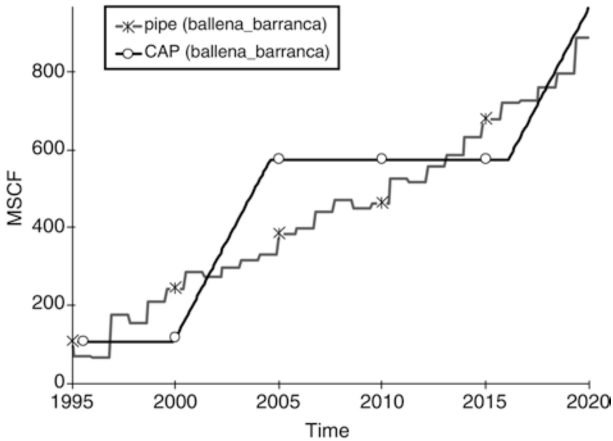


Fig. 8 Total production of natural gas under different scenarios of substitution

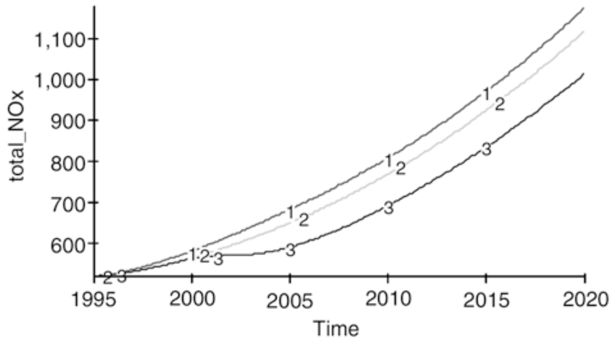


Fig. 9 Effects of policies in NO_x emissions

Figure 8 shows the natural gas production for the substitution scenarios in Table 1. When prices of gas are relatively low, demand grows and production will follow an ascending path (line 1). If prices of gas are much higher than prices of its substitutes, demand for gas falls and also does production, as line 2 illustrates. The reduction in consumption and production is lower in Scenario 3, when initial prices for gas and substitutes are similar. The discontinuities in production are due to the optimization routine and changes in substitution patterns.

Emissions Control

Besides its economic advantages, substitution of gasoline by natural gas is an efficient policy for the control of air emissions in the transportation sector. Figure 9 compares emissions of NO_x under several environmental policies. The emissions, when no control policies are applied, appear in line 1 of Fig. 9. Line 2 shows the emissions when trip lengths are reduced, and line 3 corresponds to the reduced emissions from substitution of gasoline by natural gas. It is clear that the reductions from substitution are greater than those from trip optimization.

As shown in Fig. 8, demand for natural gas in all sectors is expected to grow significantly. The response of the natural gas industry is what will determine whether or not this demand will be satisfied. Depletion of proven reserves is not yet a restriction for production, and the critical factor to analyse is the production capacity, which depends on the operation and capacity of the gas transportation network.

Effect of the Gas Transport Network on the Market

With growing demand and production, some of the segments of the gas transport network will soon be operating at full capacity, as Fig. 10 illustrates for the segment from *Ballenas to Barranca*. In Fig. 10, the line with stars, 'pipe series', corresponds to the gas volumes that should be transported, according to the gas-network model; whereas, the line with dots, 'the cap series', corresponds to the simulated capacity evolution of the pipeline. This figure shows that unsatisfied demand due to transportation constraints provides a signal for capacity expansion. It can be inferred that in this case the optimization problem presented in the Transportation section finds no solution, as restrictions cannot be satisfied.

Let us turn now to analyse the transportation system as a whole, leaving aside particular pipeline segment, as above. The restrictions imposed by the limited capacity of natural gas transmission network will have an effect on the gas price at the market place. Line 1 of Fig. 11 illustrates how average gas-price rises when no expansion of the transportation network takes place. When prices climb, non-captive clients can use

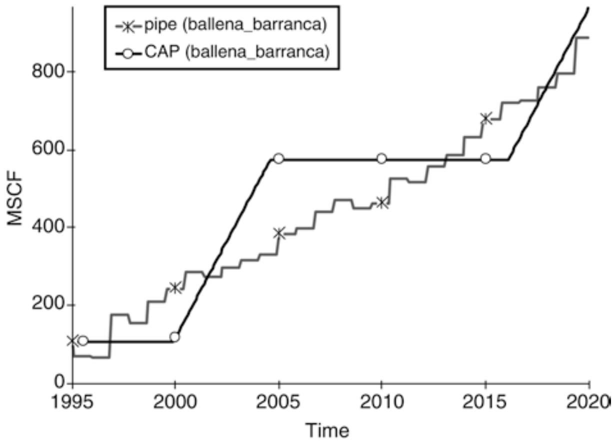


Fig. 10 Pipeline capacity and volumes transported in MSCF per day

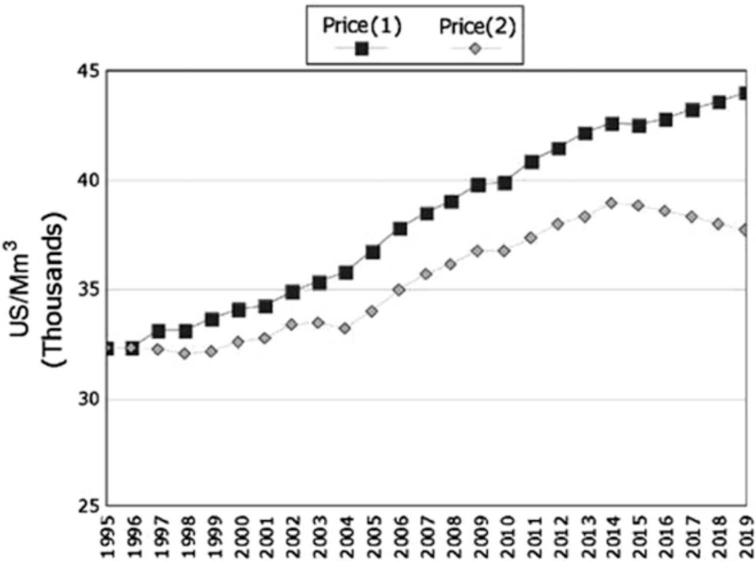


Fig. 11 Market price, US$/Mm^3

alternative substitutes, and demand decreases, making prices fluctuate. By contrast, when pipeline expands to avoid demand dissatisfaction, prices remain low, as indicated by line 2 of Fig. 11. Note that in this case, we assume that both generators and transporters can switch to alternative fuels, which is not unrealistic under normal circumstances.

Results show that the outlook for the natural gas industry seems promising in Colombia, as natural gas is a suitable alternative for gasoline, coal and electricity. During the time period analysed, the level of natural gas reserves can support the industry expansion, even in the most pessimistic scenario. With increasing demand and the creation of a new industry environment, there will be favourable conditions for more exploration and discoveries.

Capacity of transportation networks may turn out to be a bottleneck for the development of natural gas markets. The regulatory entities must establish conditions for coordination within the sector as well as with other sectors. In that way, the obstacles emerging from the restrictions in the gas transportation system can be removed.

Conclusions

The presented model facilitates policy analysis by integrating different methodologies and modelling approaches. The integration of SD with optimization is required as cost-minimization is a reasonable approximation for satisfying supply and demand, under network constraints.

By following a path consisting of integrating methodologies for validation and policy analysis purposes, the analysis platform handles the system complexity being modelled, improves its understanding and widens the perspectives of analysts and policy-makers.

From the energy perspective, the discussed tool facilitates understanding the dynamics of the natural gas industry and the possible impacts of policies and regulations on the system, considering all major sub-sectors and their relationships. In this paper, a gasoline substitution programme proved to be promising from the energy-efficiency point of view as well as being environmentally friendly. Other policies, such as pipeline expansion and exploration, seem to operate properly under price liberalization, but there are no definite conclusions to this end. We do not investigate this issue further as this goes beyond the focus of the paper. However, the approach appears to be robust enough for the study of such policy and regulation issues.

Acknowledgements We acknowledge the financial support of The COLCIENCIAS (The Colombian Research Council), UPME (Colombian Ministry of Mines and Energy) and the Universidad Nacional de Colombia. This paper was completed while the co-author spent his sabbatical at Imperial College London. We thank the anonymous referees for their useful comments and constructive criticism.

References

1. Naill RG, Klinger BA and Petersen E (1992). An analysis of the cost effectiveness of US energy policies to mitigate global warming. *System Dyn Rev* **8**(2): 111–128.
2. Sterman J and Richardson G (1985). An experiment to evaluate methods for estimating fossil fuel resources. *J Forecasting* **4**: 197–229.
3. Davidsen P, Sterman J and Richardson G (1990). A petroleum life cycle model for the United States with endogenous technology, exploration, recovery, and demand. *System Dyn Rev* **6**(1): 66–93.
4. Becker A, Godec M, Pepper W and Zammerilli A (1995). *Gas system analysis model. Technology and policy assessment of North American natural gas potential*. Society of Petroleum Engineers, SPE PAPER 30187, Dallas.
5. Baron R (1997). *Recent developments in the gas system analysis model (GSAM).* US Department of Energy, Washington.
6. Laurillard D (1993). *Rethinking University Teaching: A Framework for the Effective Use of Educational Technology*. London: Routledge.
7. Curry B and Moutinho L (1992). Using computer simulation in management education. *Mngt Educ Dev* **23**(2): 155–167.
8. Dyner I (2000). System dynamics platforms for integrated energy analysis. *J Opl Res Soc* **51**: 136–144.
9. Bunn D and Larsen E (eds) (1997). *Systems Modelling for Energy Policy*. New York: Wiley.

10. Ford A (1997). System dynamics and the electric power industry. *System Dyn Rev* **13**: 53–86.
11. Morecroft J and Sterman J (1994). *Modeling for Learning Organizations*. Portland, OR: Productivity Press.
12. Bunn D, Larsen E and Vlahos K (1993). Complementary modelling approaches for analysing several effects of privatization on electricity investment. *J Opl Res Soc* **44**: 957–971.
13. Sage AP (1991). *Decision Support Systems Engineering*. New York: Wiley.
14. Lane D (1995). On a resurgence of management simulations and games. *J Opl Res Soc* **46**: 604–625.
15. Langley PA and Larsen E (1995). Converging technologies: multi-media and gaming simulations. *J Intell Systems* **5**(2–4): 151–177.
16. Smith R, Dyner I, Olaya Y and Arango S (1999). Penetración del gas en el sector transporte terrestre en Colombia. *Energética* **12**(21): 41–56.
17. Dyner I, Smith R and Peña G (1995). System dynamics modelling for residential energy efficiency analysis and management. *J Opl Res Soc* **46**: 1163–1173.
18. Dyner I and Bunn D (1997). A systems simulation platform to support energy policy in Colombia. In: Bunn D and Larsen E (eds). *Systems Modelling for Energy Policy* New York: Wiley.
19. Naill RG (1973). The discovery cycle of a finite resource: a case study of US Natural gas. In: Meadows DH and Meadows DL (eds). *Toward Global Equilibrium*. Cambridge, MA: Wright-Allen Press.
20. Naill RG and Behrens III WW (1974). Non-renewable source sector. In: Meadows DH, Meadows DL, Randers J and Behren III WM (eds). *Dynamics of Growth in a Finite World*. Cambridge, MA: Wright-Allen Press.
21. Gordon L, Gautier D, Mast R and Root D (1993). US Geological survey estimates of natural gas energy resources. In: *The future of energy gases*. USGS Geological Paper No. 1570, Washington, pp. 495–506.
22. Morehouse DF (1997). The intricate puzzle of oil and gas reserves growth. *Natural Gas Monthly*, Energy Information Administration. US DOE, Washington.
23. Ghouri S (1996). Pakistan's new petroleum policy 1994. An immense promise. *Energy Policy* **24**(5): 477–488.
24. Livernois J and Russell U (1987). Extraction costs and the economics of nonrenewable resources. *J Polit Econ* **95**(1): 195–203.

25. Berg E, Kverndokk S and Rosendahl K (1999). Optimal oil exploration under climate treaties. Statistics Norway. Discussion paper No 245. Available at http://www.ssb.no/publikasjoner/ etter_serie/dp/
26. Alazard N (1996). Le progres scientifique et technique en exploration–production: impact sur les reserves et les coûts. *Rev Energie* **47**(481): 79–82.
27. Appert O and Boy De La Tour X (1997). Exploration et production des hydrocarbures: les enjeux technologiques pour l'Europe. *Rev Energie* **48**(485): 105–118.

Understanding the Drivers of Broadband Adoption: The Case of Rural and Remote Scotland

S. Howick and J. Whalley

Introduction

National broadband coverage has been a policy objective adopted in many developed countries. Although some countries have advanced towards this goal more than others, an increasing number of countries can now claim to have achieved widespread broadband availability. However, it does not follow that widespread broadband availability automatically leads onto its widespread adoption. Broadband may be seen by

Journal of the Operational Research Society (2008) **59**(10), 1299–1311.
https://doi.org/10.1057/palgrave.jors.2602486
Published online 22 August 2007.

S. Howick (✉)
University of Strathclyde, Glasgow, UK

Department of Management Science, University of Strathcly, Glasgow, UK

J. Whalley
University of Strathclyde, Glasgow, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_12

some potential adopters as being too costly, while others may feel that there is no compelling service that warrants its uptake. Consequently, it is pertinent to ask what factors drive broadband adoption?

If the socio-economic benefits of broadband are to be realized, such as access to educational services or markets located elsewhere, then adoption needs to be both understood and encouraged. This is particularly important in rural and remote areas. It has been argued that rural and remote areas will enjoy significant socio-economic benefits from the introduction of broadband (Scottish Executive 2001a, 2002), with some going as far as suggesting that broadband will have the same transformational impact as electricity. The introduction of broadband will contribute to the death of distance that telecommunications has so long promised to bring about.

This paper focuses on a specific rural and remote area of the UK, rural and remote Scotland. The paper aims to explore the key drivers of broadband adoption in this area. After presenting an overview of broadband and rural and remote Scotland, broadband availability and adoption are discussed. A model is then presented which identifies the key drivers of broadband adoption in rural and remote Scotland and how they interact with one another. The paper closes with recommendations for future policy initiatives in this area.

Broadband

There is no agreed definition of broadband. Although the ITU states that broadband equates to transmission speeds faster than 1.5 or 2 mbps (ITU 2003, p. 9), a range of alternative definitions have been suggested. The OECD, for instance, adopts a downstream access threshold of 256 kbps, a rate that is slightly faster than the 200 kbps rate adopted by the FCC (ITU 2006, p. 21). Ofcom, in contrast, has opted for a broader definition encompassing always on and data rates of at least 128 kbps (Ofcom 2004). The Ofcom definition is adopted here.

Although there may be some disagreement as to the definition of broadband, it is widely accepted that there are many advantages associated with it (Varian et al. 2002; Vigden et al. 2004; Cava-Ferreruela and

Alabau-Mun˜oz 2006). These advantages are both economic as well as social. Broadband contributes to national economic competitiveness (ITU 2003), allowing industry to access distant markets as well as develop and deliver new services. There are also social benefits associated with broadband. The greater download speeds associated with broadband broaden and improve the services that can be offered as part of telemedicine and tele-education, as well as allow dispersed families to remain in contact through, for instance, the online sharing of pictures or the use of Skype. While Fransman (2006) is not alone in noting that broadband lacks a 'killer application,' it has made the downloading of music and videos more attractive and common than was the case with dial-up Internet access.

Drawing on the literature it is possible to identify a range of factors that encourage ICT and broadband adoption. The first identifiable factor is that of cost. Although Bauer et al. (2005) argue that the relationship between broadband penetration and cost is not yet fully understood, Geroski (2000) states that cost is an important issue for those dial-up Internet users considering switching to broadband. As broadband Internet access is typically more expensive than its dial-up counterpart, potential switchers who are unaware of the benefits of broadband may be reluctant to pay more for what they perceive to be the same service.

It is worth noting, however, that while the difference in price between dial-up and broadband may not be that great, the cost of the service is often not the only cost that has to be considered by potential switchers as new computer equipment and training may also be required. In the late 1990s research found that households in rural America were less likely to own a computer than their urban counterparts, thereby adding to the cost of using the Internet (Strover 2001). Moreover, Biggs and Kelly (2006) draw attention to broadband pricing and speed trends—drawing on ITU data they show, on average, that between 2003 and 2005 broadband prices have fallen while speeds have increased. In other words, broadband prices are falling in absolute terms and are thus less of a barrier to adoption than was previously the case.

Secondly, the attributes or characteristics of broadband can also influence its adoption. Savage and Waldman (2005) identified three such attributes—speed, service reliability and 'always-on'—whose influence on adoption varies depending on the social status of the potential adopter. Potential adopters with higher incomes value these more than those with lower incomes, and those with a degree value speed more, always-on less and reliability about the same as those without a degree. Warren (2004) found that the majority of computer literate farmers struggle with slow connection speeds, with the consequence that the faster speeds of broadband are likely to be attractive to them.

A third factor that can encourage broadband adoption is the social context within which the potential adopted is located. Savage and Waldman (2005) found that broadband adoption is most likely in households with a higher income, college education and multiple computers. Ofcom (2006c, p. 64) also draws attention to the supporting role played by friends and family when it comes to learning about digital services and products. Across the UK as a whole, the most popular way to learn about digital products and services was through reading the manual, with the second most popular being through asking friends and family for assistance. Scotland was the only part of the UK where the reverse was found.

Fourthly, would-be adopters should feel a need to use the Internet and that access is best serviced through a broadband connection. In a 2006 survey by Ofcom, a lack of need or interest was found to be the overwhelming reason why the Internet was not used at home (Ofcom 2006a, p. 71). Galloway and Mochie (2005) reported a perceived lack of need among rural small and medium sized enterprises, while more broadly in rural England it has been observed that businesses adopt ICT at a significantly slower rate than their urban counterparts (Warren 2004; DEFRA 2005).

A major concern in encouraging broadband adoption is to ensure that some areas are not left behind, as they will not be able to enjoy the advantages that it brings. Within the UK, concern has been expressed about geographically remote areas such as Devon, Cornwall and rural Wales. One area that has received considerable interest is rural and remote Scotland, not least because the Scottish Executive has actively sought to ensure that those in rural and remote areas have the same access to broadband as those within urban areas. Through broadband the economy will be revitalized and population levels stabilized, thereby allowing social services to be provided (Scottish Executive 2001a). There are many rural and remote areas for which exploration of the drivers of broadband adoption would be beneficial, however, the focus of this paper is on how broadband adoption can be encouraged in rural and remote Scotland.

Scotland

Located in the north of the British Isles, Scotland is the second largest of the four countries that form the United Kingdom. The population of Scotland is just over five million people, the majority of whom live within the central belt that connects Glasgow in the west with Edinburgh in the east. Around one-third of the population live in one of Scotland's six cities, which are shown in Fig. 1, with the consequence that population densities vary considerably across Scotland. The highest population densities can be found in Glasgow—3290 people per square kilometre—and the lowest in the Highlands—8 people per square (National Statistics 2003, p. 19). Figure 1 also highlights the areas served by Scotland's two regional development agencies, Highlands & Islands Enterprise (HIE) and Scottish Enterprise (SE).

The Scottish Executive is the devolved government of Scotland. The purview of the devolved government is determined by the Scotland Act, 1998. Some policy areas such as education and health are devolved to Edinburgh while others like the UK single market and defence were retained in London. Those policy areas that were retained in London are referred to as 'reserved matters' and as telecommunications relates to the UK single market responsibility remained in London. As a consequence, Ofcom regulates all of the industry across the UK. This should not be taken, however, as suggesting that telecommunications policy is decided in London without consultation with others parties. Ofcom has established committees representing each of the four countries within the UK as well as users, and regularly engages in consultation on specific issues. The two Scottish regional development agencies have developed their own broadband initiatives within the wider UK framework, and have sought to influence Ofcom through contributing to consultations.

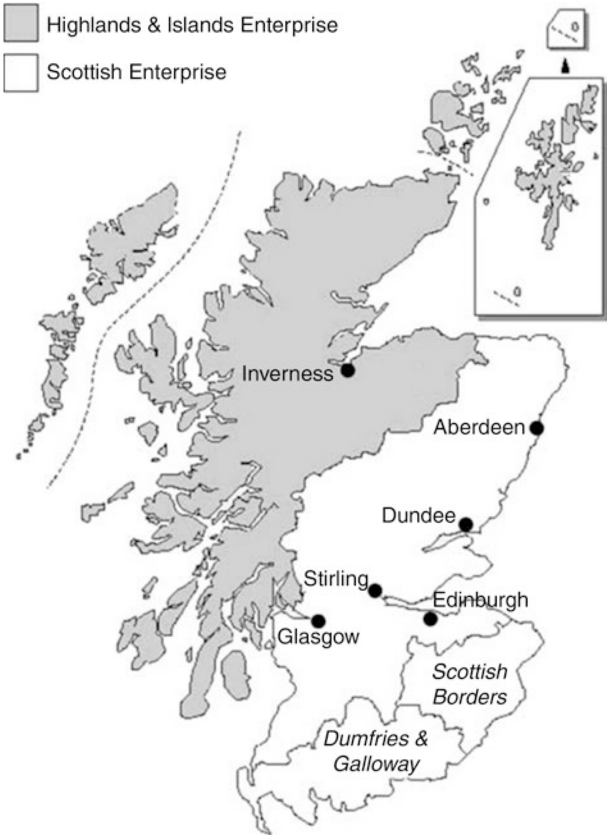


Fig. 1 Scotland

Broadband Availability

According to a recent report, four technologies are used in the UK to deliver broadband services: digital subscriber line (DSL), cable, fixed wireless access and satellite (Ovum 2006). Only two of the technologies are widely available throughout the UK, that is, DSL and cable (Ovum 2006, p. 13). However, the dominant broadband technology in rural and remote areas is DSL with the consequence that the role of BT in broadband availability is particularly important.

As befits an incumbent operator, BT owns and maintains telephone exchanges across rural and remote Scotland that have been enabled to provide broadband services in a piecemeal fashion. A substantial number of exchanges were enabled when sufficient numbers of potential broadband users had registered to breach the ‘trigger levels’ (a threshold of potential users above which BT would enable the exchange for broadband) set by BT. As a consequence of this policy, around 85% of the UK had access to broadband services by April 2004 when BT announced the end of its ‘trigger level’ scheme. Subsequently, those exchanges that had reached 90% of their trigger level were upgraded more or less immediately while all the remaining exchanges with a trigger level were to be upgraded by the summer of 2005. On completion, broadband would be available to around 99.6% of businesses and households across the UK.

In April 2005 BT announced that it would upgrade most, but not all, of its remaining exchanges within Scotland. A total of 378 out of 399 exchanges would be upgraded as a consequence of financial support offered from primarily the Scottish Executive as well as the European Regional Development Fund. Although this increased broadband availability within Scotland, some areas were still without access as not all exchanges were upgraded. In these areas, public–private partnerships, such as the Connected Communities project in the Western Isles, have been used to provide broadband services in areas where BT does not do so. This project, which has total capital expenditure of just over £5 million, delivers broadband to public sector bodies such as schools and hospitals as well as businesses and households through its affiliated ISP (Lattemann 2006; Ofcom 2006e). While this project has provided broadband infrastructure where none existed, other public–private partnerships elsewhere in Scotland have improved the quality of the infrastructure that is available.

Finally, it is worth noting that as a consequence of the strategic review of telecommunications on the one hand and the focus of Ofcom on local loop unbundling on the other hand, service competition is possible throughout Scotland (Whalley 2005). This means that while there may be only a single infrastructure owned by BT in many parts of Scotland, competition between different companies is possible at the broadband package level. When the strategy adopted by BT is combined with public–private partnerships and the emergence of service competition, broadband is available throughout Scotland with the consequence that the pertinent question to ask is not how broadband availability can be encouraged but rather what drives broadband adoption. Therefore, this paper focuses on 'What are the key drivers of broadband adoption in rural and remote Scotland and which of these drivers should policies focus on to have the greatest impact on encouraging broadband adoption?'

Broadband Adoption

Although broadband is available to 99.9% of households and businesses across the UK (Ofcom 2006a, p. 37), broadband adoption rates are far lower. As can be seen from Table 1, Internet adoption varies between 48 and 59% depending on which part of the UK is looked at. These figures are, however, for dial-up and broadband Internet households with the consequence that broadband adoption is actually lower.

Around 36% of households across the UK are broadband households, though there are considerable differences between England (37.76%), Scotland (30.6%), Wales (26.46%) and Northern Ireland (24.96%). Significantly Ofcom (2006a, p. 54) found that the rural and remote areas were more likely to use the Internet, albeit via dial-up, than their urban counterpart with one exception: Scotland. In Scotland, urban adoption of the Internet is greater than that in rural and remote areas.

To gain an understanding of the main factors that contribute to businesses or households deciding to adopt broadband or, more importantly,

Table 1 Broadband and PC adoption rates 2005

	UK	England	Scotland	wales	N Ireland
Internet household fine broadband (%)	57	59	51	49	48
Broadband as % Internet households	63	64	60	54	52

Source: Ofcom (2006a, p. 54).

what hinders their adoption of broadband, information was gathered from three sources: literature on new product diffusion, broadband literature and interviews with key stakeholders in the broadband market in rural and remote Scotland. This information was used to build a causal diagram in order to appreciate how the various drivers of broadband adoption interact with one another. The process used to build up the causal diagram using the above three sources of information will be discussed next.

New Product Diffusion

New product diffusion is a widely researched issue (e.g. Rogers 2003) and many new product diffusion models exist in the literature (Mahajan et al. 1990, 2000); however, one of the most well known and widely used is the Bass model (Bass 1969). Since its introduction, many researchers have proposed extensions to the original form of this model (Mahajan et al. 1990, 2000; Parker 1994). In addition, a summary of the applications of the extensions of the Bass model to telecoms can be found in Fildes and Kumar (2002). Due to its wide applicability to many areas of new product diffusion and, in particular, its goodness of fit to technological diffusion of information and telecommunications innovations (Teng et al. 2002; Kim and Kim 2004), the Bass model structure was used as a starting point to aid understanding of broadband adoption in rural and remote Scotland.

A causal diagram representing the main elements of the Bass model (Sterman 2000) can be viewed in Fig. 2. The Bass diffusion model assumes that adoption for a product stems from two main sources; innovators who adopt the product due to external sources of awareness, usually interpreted as the effect of advertising and from imitators who adopt the product as a result of contact with previous adopters, that is, from word of mouth.

Figure 2 highlights three feedback loops. The first is a positive feedback loop; as more people adopt the product, there are more people to communicate its benefits through word of mouth and hence further people will adopt. There are also two negative feedback loops; as more people

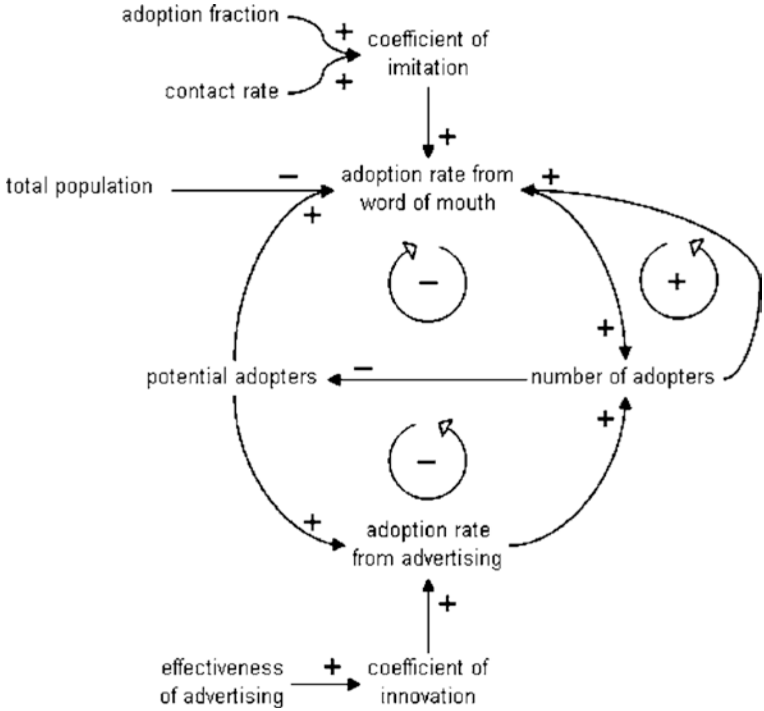


Fig. 2 Causal diagram representing the Bass diffusion model

adopt the product, there are less potential adopters left (due to a finite population of potential adopters) to adopt through the effects of either word of mouth or advertising.

The new product diffusion literature provided a base model with which to begin considering the drivers of broadband adoption in rural and remote Scotland. The next section discusses how this was extended using the broadband literature and interviews with key stakeholders.

Broadband Literature and Interviews

Following a review of the broadband literature the first step in the modelling process was to split the population into four categories: households with a dial-up connection, households without a dial-up connection, businesses with a dial-up connection and businesses without a dial-up connection. This split of the population was deemed necessary as it was felt that these categories may have varying drivers to adoption and may also be influenced by differing coefficients of imitation and innovation. In order to reflect the specific situation of broadband adoption in rural and remote Scotland, two areas of the model were expanded as follows:

1. There are obviously many factors that will influence a potential adopter in their decision to adopt broadband. Many of these can be captured by considering the various decision criteria that are taken account of by a potential adopter. This will consist of the broadband attributes that are considered to be of most importance to potential adopters. These decision criteria will affect the coefficients of innovation and imitation.
2. The factors that affect the pool of potential adopters.

In addition to information gained from reviewing literature on broadband, interviews with key stakeholders from the broadband industry in rural and remote Scotland provided clarification on what are considered to be the key drivers and their impact on broadband adoption. The interviewees included senior managers from the Scottish Executive, HIE, Scottish Enterprise Borders, Ofcom, consultants from both a major telecommunications focused practice as well as an independent consultancy. In addition, HIE represent the interests of businesses within the region they serve with the consequence that their views will have been included, albeit indirectly.

Each of the organizations that were interviewed are all keen to understand 'What are the key drivers of broadband adoption in rural and remote Scotland and which of these drivers should policies focus on to have the greatest impact on encouraging broadband adoption?' HIE and Scottish Enterprise Borders were interested due to the economic benefits that broadband brings, and while this is also true of the Scottish Executive they were also motivated to participate by the welfare benefits that broadband may bring. Ofcom, as the telecommunications regulator, has a natural interest in ensuring the functioning of the market and understanding how market failure may be addressed.

Interviewees were asked to comment on the drivers of adoption that were taken from the literature and also asked to expand on these based on their own experiences. The following sections summarize the main drivers that were agreed upon by the interviewees.

Decision Criteria Function

A decision criteria function for households without dial-up is shown in Fig. 3.

Figure 3 illustrates the key factors which are believed to influence the decision of a householder, who currently has dial-up, on whether or not to adopt broadband. This figure includes factors such as the relative cost of the decision (Scottish Executive 2001b, p. 7) as well as whether households use, or would like to use, products and services like digital cameras and Internet shopping (that is, the 'use of other technologies') whose use is enhanced in some way by broadband Internet access (Fransman 2006; Ofcom 2006c). Also included is whether households appreciate the attributes of broadband, or have any concerns regarding either the content that is available on the Internet or its security (Ofcom 2006a). If there are any children within the household, the level of ICT competence is raised with the consequence that adopting broadband is not as great a hurdle as would otherwise be the case (Robertson et al. 2004). In other words, the complexity of adopting broadband is decreased.

It should be noted that the cost of broadband is represented as an exogenous factor. It could be argued that as the total number of adopters increases, the price of broadband falls. However, the regulation of local loop unbundling, which has effectively created a floor price, along with the extensive size of the market reduces the extent to which prices will fall in future.

A feedback loop is identified in Fig. 3. This loop shows us that the dial-up users for whom the decision criteria/utility function is highest are those with the highest dial-up costs as they are the heaviest users. Since these users are the most likely to switch to broadband, the average dial-up cost will drop as the heavier users of dial-up migrate to broadband, with the consequence that those dial-up users who remain are those with a lower utility function. This is a balancing or 'goal seeking' feedback loop

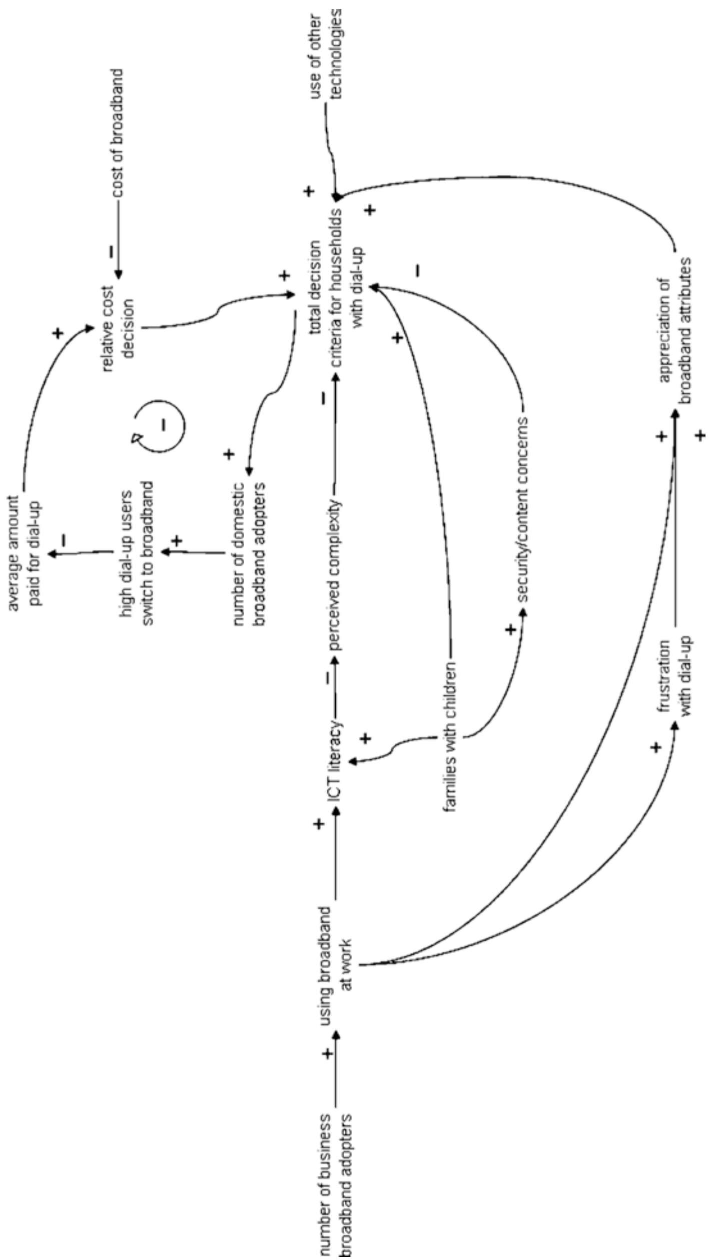


Fig. 3 Key factors affecting decision criteria for households with dial-up

that indicates that dial-up users will continue to migrate to broadband as long as there are discernible financial benefits in such a move.

For households without dial-up the decision criteria function is similar to that shown in Fig. 3, with the following amendments:

1. 'Frustration with dial-up' will not occur.
2. The financial decision criteria are based on an absolute cost which will include the absolute cost of broadband, cost of a computer and the household's level of income.

The decision criteria function for businesses with dial-up is shown in Fig. 4.

Figure 4 shows that for businesses with dial-up the decision on whether or not they should adopt broadband is based on two key factors: the cost and benefit of broadband. The cost involves the relative cost of broadband versus dial-up. An incentive payment may also be available to some businesses (Scottish Enterprise 2004). In addition, some business may

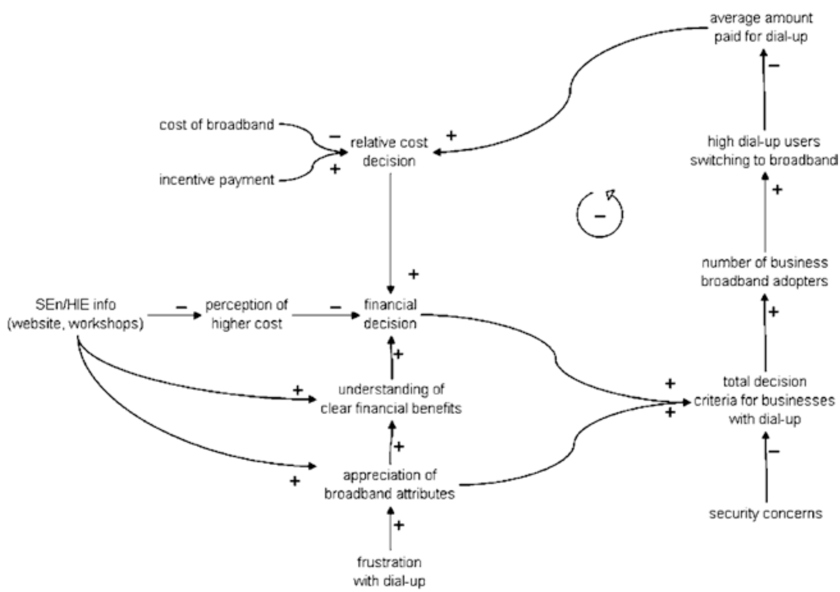


Fig. 4 Key factors affecting decision criteria for businesses with dial-up

have a perception that costs are higher than they actually are (Scottish Executive 2001b, p. 11), however campaigns such as SE/HIE workshops (where technology is demonstrated and the benefits that it brings discussed) may have helped to alleviate this perception. The benefits of broadband are seen to be an appreciation of its attributes and through understanding the potential financial benefits that it could bring to a business. A third factor that is also taken into account is concerns about security. A business will need to ensure that any transactions or information that is processed through broadband will be safe. A feedback loop is identified in Fig. 4. This loop is the same as that for households shown in Fig. 3. It is a balancing feedback loop that indicates that dial-up business users will continue to migrate to broadband as long as there are discernible financial benefits in such a move.

For businesses without dial-up the decision criteria function is similar to that shown in Fig. 4, with the following amendments:

1. Frustration with dial-up will not be an influencing factor.
2. The financial decision criteria are based on an absolute cost which will include the absolute cost of broadband, any incentive payments if they are available, costs associated with security issues, the income level of the business.
3. Campaigns such as the SE/HIE workshops will also promote ICT literacy among businesses and hence reduce any perceived complexity which will have a positive effect on the decision criteria.

Potential Adopters

Figure 5 captures the key factors that influence the pool of households without dial-up that are potential broadband adopters.

A household can only adopt broadband if the infrastructure is available to them and they know that it is available to them. If householders use broadband at work, this results in them having more knowledge about infrastructure availability (Hollifield and Donnermeyer 2003). In addition, some households simply do not want to use the Internet, or indeed a PC at all. For these people, the Internet currently offers no compelling service that would encourage their adoption of broadband

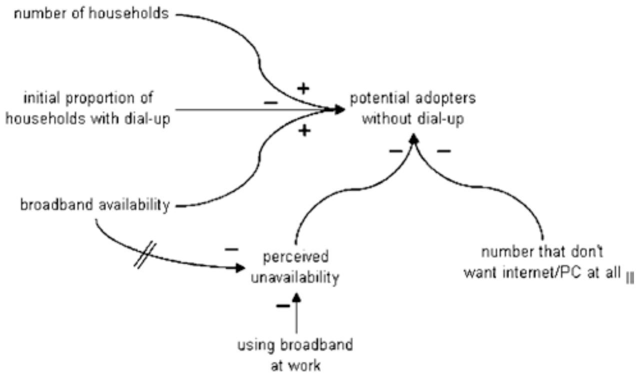


Fig. 5 Key factors that influence the pool of households without dial-up that are potential broadband adopters

(Fransman 2006) nor are they encouraged by the decline in prices (Brignall 2006; Ofcom 2006d, p. 119). Ofcom (2006a, b) terms this as 'voluntary exclusion' and is currently reported as 24% of people over the whole of Scotland (Ofcom 2006b). The two lines across the arrow linking 'broadband availability' and 'perceived unavailability' indicate a delay in people's perception of the availability of broadband.

A similar causal diagram to Fig. 5 exists for households with dial-up. The only difference is that the factor 'number that don't want internet/PC at all' will have no influence in this situation.

The factors that influence the pool of businesses without dial-up that are potential broadband adopters can be seen in Fig. 6.

Figure 6 is very similar to Fig. 5. The main difference is that perceived unavailability of broadband infrastructure is influenced by campaigns such as the SE/HIE workshops which will aid in reducing this perception.

Advertising

In addition to expanding the key factors that affect the decision criteria and pool of potential adopters, there were a number of factors in the literature that impacted the role of advertising in broadband adoption. Figure 7 illustrates the key factors affecting advertising.

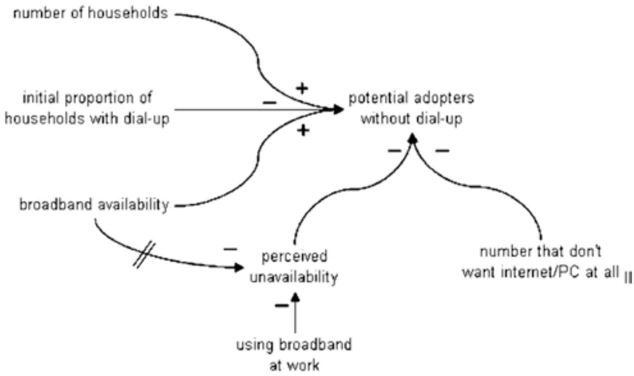


Fig. 6 Key factors that influence the pool of businesses without dial-up that are potential broadband adopters

The concepts shown in bold are those that appear in Fig. 2, that is they indicate how the factors influencing advertising impact on the original Bass model. In this figure, SE/HIE media campaigns include the 'Speak up for Broadband' and 'Broadband for Scotland' promotions. It should be noted that ISP advertising is shown to influence both the coefficient of imitation and innovation. It is believed that although the main influence on the coefficient of imitation is through word of mouth with other adopters, the fact that someone has seen ISP advertising will impact the meeting they have with the existing adopter.

The Impact of Policies on the Causal Diagram

Bringing together the causal diagrams represented in Figs. 2–7 produces an overall casual model for broadband adoption (allowing for all four sub-groups of the population). This model provides a clear explanation of the key drivers of broadband adoption in rural and remote Scotland. This model can be used to consider the impact of past or future policy initiatives to encourage broadband adoption. A number of policy initiatives have been introduced by SE and HIE (Tookey et al. 2006). The following

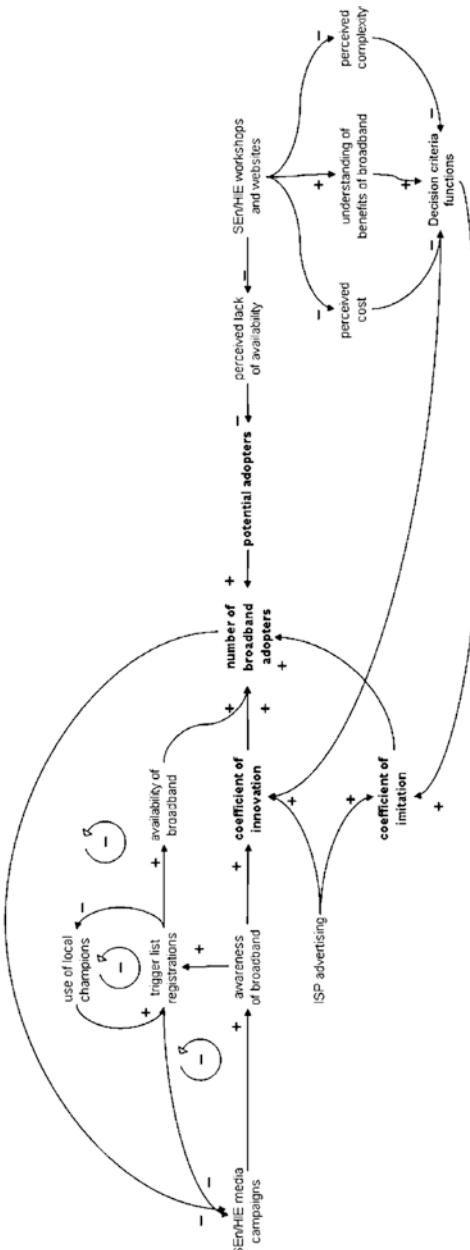


Fig. 7 Key factors that influence advertising

discusses the key initiatives and how they impact on broadband diffusion:

1. *TV and website advertising* (*Speak up for Broadband*, *Broadband for Scotland*): The TV advertising campaigns drew attention to broadband and encouraged individuals to register their interest in broadband on the trigger lists run by BT. This campaign ended when BT decided to upgrade all of its exchanges with trigger lists in place. In contrast, the second campaign is ongoing and is more wide ranging in nature as it seeks to provide impartial and technological neutral advice on broadband for businesses and individuals alike. These impact on the awareness of broadband shown on Fig. 7, which in turns impacts on the coefficient of innovation. However, to be influenced by website advertising, access to the Internet, either at home via dial-up or at work, is required.
2. *Demonstration of broadband benefits* (*www.work-global.com*, *e-business demo centres*, *touring demonstration bus*): These campaigns provide illustrations of broadband in use so that potential adopters can appreciate its attributes. The web-based campaigns provide case studies showing examples of broadband use while the touring demonstration bus allows would-be adopters to appreciate these for themselves through a hands-on session. These impact on the appreciation of broadband attributes shown in Figs. 3 and 4, which in turn influence the decision criteria for households and businesses.
3. *Incentive schemes* (*Business incentive scheme*): These are targeted towards businesses and provide a financial incentive to adopt broadband through subsidizing the cost of connection. These are shown as incentive payments in Fig. 4 and impact on the relative cost decision, which in turn influences the decision criteria for businesses.

Each of the above policies have impacted the coefficients of innovation and/or imitation, mainly through the decision criteria function. These have therefore had an impact on the rate at which potential adopters have adopted broadband. However, none of them have directly impacted the pool of potential adopters, thus restricting the uptake of broadband.

As the causal model indicates that the pool of potential adopters is restricting broadband adoption, it is important to explore whether or not this is a long-term constraint. Therefore, to assess the over time behaviour which is produced from the relationships contained in the causal model a quantitative system dynamics (Forrester 1961; Sterman 2000) simulation model was produced.

Simulation Model

System dynamics has been used to understand the diffusion process at an aggregate level in a number of previous studies (Milling 1996; Maier 1998) and to specifically explore the diffusion of telecommunications (Osborne 1999). The particular value of system dynamics in modelling the diffusion process stems from the large number of factors that influence a potential customer's decision to adopt a product. The influence of these factors on the decision to adopt will vary over time. System dynamics is of particular use to model the behaviour of the impact of these factors over time (Lyons et al. 2003). In addition, the strategic analysis of the diffusion of a new technology will involve variables for whom there is limited data. System dynamics is able to take account of such variables if they are believed to be affect system behaviour (Fildes and Kumar 2002).

Model Structure

Figure 8 represents an extract of the System Dynamics simulation model that was built. This figure captures the main factors that influence the adoption rate for businesses without dial-up and reflects the key factors and relationships that were captured in the causal diagram. The remainder of the system dynamics model includes similar structures for businesses with dial-up, households without dial-up and households with dial-up.

In order to appreciate how the structure of the quantitative simulation structure relates to the structure of the causal diagram, four key variables can be highlighted:

1. 'bus ndu criteria' on the left-hand side of Fig. 8 represents the decision criteria function for non-dial up businesses. This is influenced by security concerns (bus security concerns), perceived complexity (bus ndu complexity) appreciation of broadband attributes (bus ndu attributes) and financial decisions (Bus ndu finance).

   The decision criteria function is taken to be a weighted average of all the factors that influence it. This is in line with the utility function approach used in diffusion modelling, reviewed by Roberts and Lattin (2000). Others have also adopted such an approach, for example, Savage and Waldman (2005) propose a utility function to describe how adopters of broadband select between Internet service providers. Madden and Simpson (1997) developed a utility function when modelling the adoption of cable broadband services. It is also in line with the technology acceptance model used by Oh et al. (2003) which combines factors that contribute to a potential adopter's attitude towards broadband.
2. 'bus_pot_ndu' in the middle of Fig. 8 represents the potential non-dial-up business adopters. This is influenced by the total number of businesses in rural and remote Scotland (total_bus), the initial proportion of businesses that have dial-up (initial_bus_dialup_users), the number of non-dial-up businesses that have already adopted broadband (business_adopters_no_du), the number of businesses that are simply not interested in broadband (bus_not_interested) and the perceived availability of broadband (perc_bus_avail).
3. 'word_of_mouth_bus_ndu' in the upper right of Fig. 8 represents the number of non-dial-up businesses that adopt broadband each time period due to a word-of-mouth experience. As seen in Fig. 2, this is influenced by the total business population (total_bus), the number of business adopters (bus_adopters_no_du), the number of potential adopters (bus pot ndu), the contact rate (bus_ndu_contact_rate), the adoption fraction (bus_ndu_adopt_fraction) and the decision criteria (bus_ndu_criteria).
4. 'advert_adopt_bus_ndu' in the lower right of Fig. 8 represents the number of non-dial-up businesses that adopt broadband each time period due to the effect of advertising.

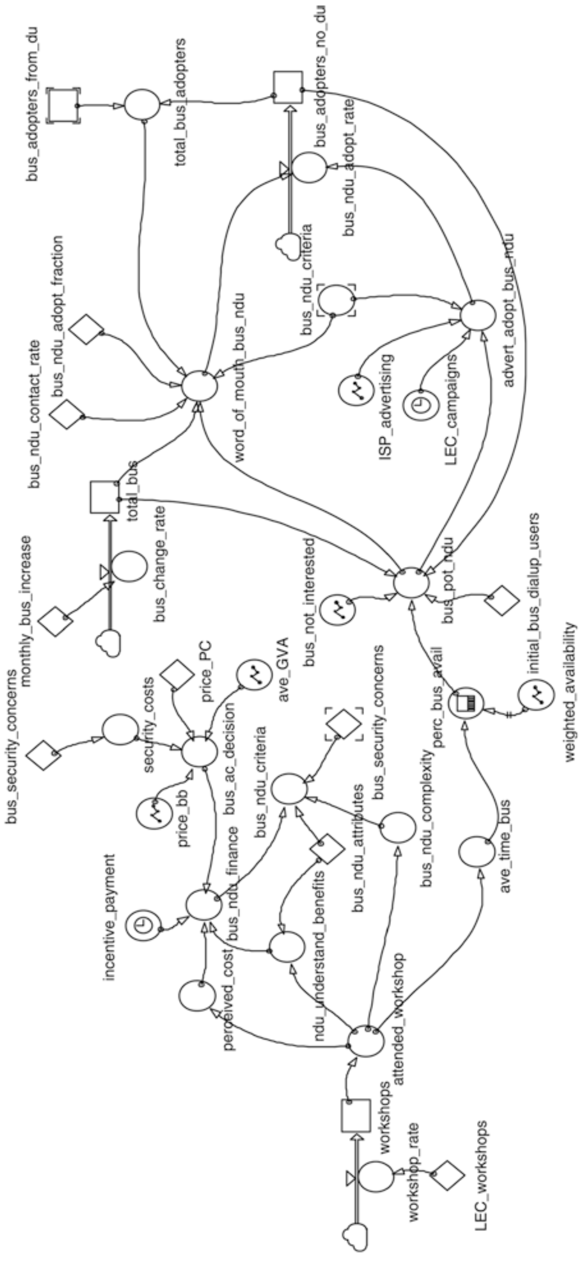


Fig. 8 Portion of System Dynamics model capturing factors that influence broadband adoption for businesses without dial-up

It should be noted that it has been assumed that if non-dial-up users wish to connect to the Internet during the time period being considered, they will move directly to broadband.

Populating the Model with Data

Various sources of data were used to populate the model. Some of the data is publicly available, for example the number of households and businesses in rural and remote Scotland (General Register Office for Scotland 2004; National Statistics 2003). Another example is the fraction of households using dial-up at the beginning of the simulation run (Ofcom 2006d; Scottish Executive 2005). However, for some parts of the model expert judgement was required, for example the future trend for the cost of domestic broadband.

During model construction and after the model had been populated by data, standard system dynamics model validation, or confidence building, tests were carried out (Forrester and Senge 1980; Sterman 2000). There are three commonly used sets of tests: tests of model structure, test of model behaviour and test of policy implications. The tests of model structure involve ensuring the structure and parameters contained in the model are consistent with the knowledge of the system and that all the concepts that are important to the problem are included in the model. The review of the literature and the interviews with interested stakeholders drew out relevant concepts and enabled those involved to gain confidence in the structure of both the qualitative and quantitative models. Gaining confidence in model behaviour involves testing the structure and parameters of the model by observing the resulting behaviour. As a part of this the interviewees were asked to produce and discuss reference mode sketches of their view of broadband adoption over a 20-year period for both businesses and households. Standard methods were used for eliciting the reference mode sketches (Andersen and Richardson 1997; Saeed 1998). Finally, tests for policy implications include observing the real system once recommended policies have been implemented to track their actual impact over time. Such tests were not relevant at this stage of the analysis.

System Dynamics Model Output

Figure 9 shows the number of household and business adopters over a 20-year (i.e. 240 month) period commencing January 2001. This date was chosen as the starting point of the simulation since this corresponded with the introduction of broadband into the market (Ofcom 2006d). In addition, 2001 also corresponds to the point when the Scottish Executive began to express an interest in broadband (Scottish Executive 2001a, b, 2002).

The shape of the graphs shown in Fig. 9 follow the typical Bass model S-shaped growth. The initial exponential growth arises from the rate of adoption from word of mouth (i.e. the positive feedback loop shown in Fig. 2). The growth then slows as the number of potential adopters reduces through saturation of the market (i.e. the negative feedback loops shown in Fig. 2). Of particular interest here is the steepness of the curves, especially for householders. Within a 2.5-year period (month 60 to

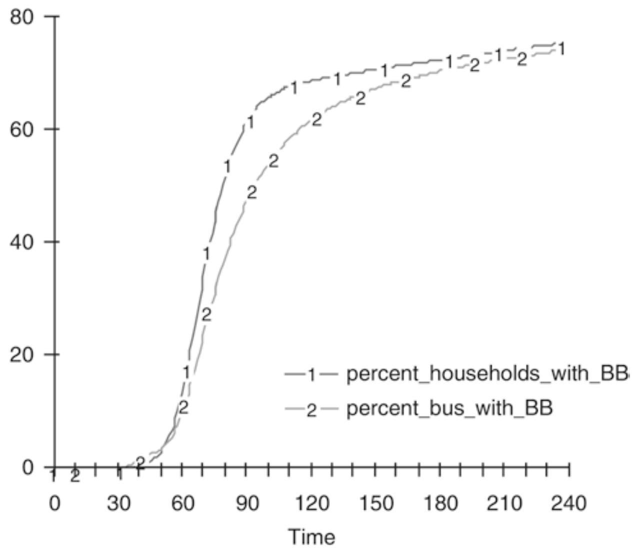


Fig. 9 Simulation model output assuming no future policies implemented to promote broadband adoption

month 90) the percentage of households with broadband increases from approximately 13–60%.

A second feature of Fig. 9 is that business adoption is lagging behind household adoption. This can be largely explained by the types of businesses that exist in rural and remote Scotland. Farmers and rural SMEs often find it difficult to appreciate how broadband can add value to their business and thus may feel that they will not benefit from its use (Warren 2004; DEFRA 2005). On the other hand, many householders will be employed by a few large employers such as local authorities and are therefore more likely to experience broadband in their employment and thus adopt.

A third feature of Fig. 9 is that 100% adoption is not reached and, indeed, the rate of adoption slows down well before this level is reached. However, the rate of adoption does not completely stop during the 20-year period. Although very gradual, there is continual evidence of adoption. The main reason for this outcome is due to the expected changing composition of the overall population in rural and remote Scotland. Although population forecasts do not anticipate the total number of population to dramatically change (Registrar General for Scotland 2005), there is an expectation of in-migration from central Scotland resulting in a large percentage of the population having experience of broadband and thus wishing to adopt it in rural and remote Scotland (Hope et al. 2004).

It should be noted that the nature of some of the data used to populate the model meant that sensitivity analysis was required in order to explore the impact on model behaviour when model inputs were changed. Although the position of the two time series shown in Fig. 9 altered during sensitivity analysis, the three features described above were reasonably robust that is S-shaped growth, business adoption lagging behind household adoption and not reaching 100% adoption. Figure 10 illustrates the impact on the simulation output when the weightings that should be applied to the key factors which were believed to influence the decision of a householder whether or not to adopt broadband (see Fig. 3) were altered. The five time series represent the use of different weightings including extremes such as having all the weighting put on just one criteria. The result is that the main conclusions of S-shaped growth and not

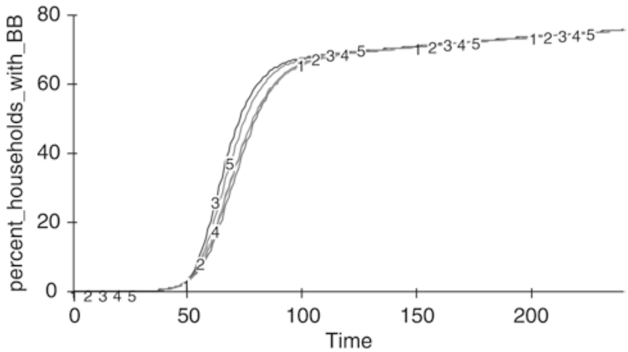


Fig. 10 Sensitivity analysis for the weightings included in the householders decision criteria

reaching 100% adoption still hold with only a small variation in the slope of the time series.

If the socio-economic benefits arising from the use of broadband are to be achieved in rural and remote Scotland, policies initiatives are required in order to speed up the diffusion process. With respect to Fig. 9, this means implementing policies which either increase the slope of the curve or lift the point at which adoption slows down. Returning to the causal model, each of the inputs to this model were explored as potential areas for policies initiatives to focus upon.

Making changes to many of the inputs to the causal model was seen to have an impact on the slope of the graphs. For example, Fig. 11 shows the impact of policies to encourage the appreciation of broadband attributes. This was implemented midway through 2007 in the model. Time series 1 and 2 are the original output from the simulation model. Time series 3 and 4 represent the diffusion curves after the policy has been implemented. There is minimal impact to the diffusion of broadband adoption amongst householders, however the slope of diffusion amongst businesses has increased.

Many of the inputs impacted the overall decision criteria for a business or a household and changing the factors contained in the decision criteria did not have a significant impact on the results as it was assumed that a

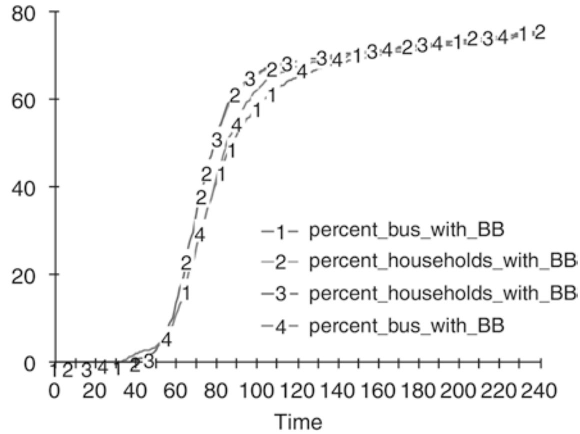


Fig. 11 Implementing policies to encourage the appreciation of broadband attributes

number of factors contribute to a household or business' decision to adopt.

However, the most significant change in output was gained from the simulation when changes were made to those factors that impact the pool of potential adopters that is lifting the point at which adoption slows down. Exploration of the causal model indicated that the limited pool of potential adopters is restricting broadband adoption and the quantitative model has confirmed that this will remain a constraint in the long term. Therefore, policies that could potentially have the largest impact on adoption are those that could target and reduce the number of households and businesses that believe that they do not want the Internet, or in fact a PC, at all. It is appreciated that many of these people may never wish to experience the Internet or PC. However, an understanding of the needs of these people is required to explore whether or not there are particular services they could benefit from and thus enable them to experience the socio-economic benefits of broadband. If, for example, it was assumed that the number who believed they do not want the Internet or a PC halved, the resulting diffusion curve would be as shown in Fig. 12.

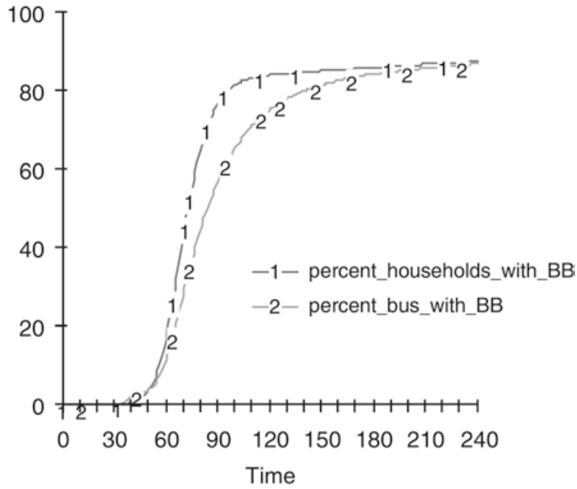


Fig. 12 Halving the number of people who not want the Internet, or PC, at all

Policy Discussion

It has been highlighted that a key group of people to target is the number of households and businesses that believe that they do not want the Internet, or in fact a PC, at all. Policies that could be adopted to target this group include the following:

- *Focused marketing campaigns*: This is not a new policy but rather the restatement of an old policy that stopped once availability was achieved. These campaigns, in contrast to those in the past, will target key groups of businesses and households with low broadband adoption rates.
- *Local champions*: Broadband could be brought closer to those who have not yet adopted through community-based local champions.
- *Incentives*: For many it could be argued that broadband is increasingly affordable, but in rural and remote Scotland incomes are substantially less with the consequence that it is more of an issue. Although those people who are simply not interested in the Internet are unlikely to be swayed by a reduction in subscription cost, introducing incentives

such as providing new computer equipment as well as demonstrating the benefits it could bring may trigger their interest.

- *Online public services*: The move of public services online could be used to force adoption but this would inevitably result in disenfranchisement of some parts of society. Online tax filing has proved to be more successful than anticipated, though whether it has the same sort of attractiveness in rural and remote areas is another matter.
- *Understanding the needs of 'those not interested'*: By definition, those people who are not interested in the Internet at all, are not motivated by any particular service. Thus a study into the attractiveness of services to this particular group may allow for tailored campaigns to be developed that would encourage uptake by this group. In addition, there is a more general issue of identifying whether a 'killer application' for broadband actually exists.

Conclusions and Future Research

Although this paper focuses on broadband within rural and remote Scotland, particular attention has been spent on understanding the drivers of adoption. Through a series of policy initiatives, some of which were UK wide and some of which were Scotland specific, broadband is now available to almost all communities within rural and remote Scotland. Following this there was a reduction in advertising by the regional development agencies that coincided with the ending of BT's trigger lists and the availability of broadband in most Scottish telephone exchanges. However, this does not mean that everyone actually has adopted broadband. Broadband has been adopted more in urban areas than rural and remote areas, with the consequence that the undoubted benefits of adoption are not being enjoyed in these areas. Thus, the first conclusion that we can make is that policy initiatives need to be developed that focus on adoption even when broadband is available.

Understanding the drivers of broadband adoption and their interaction was believed to be key to the development of appropriate policy initiatives. Therefore, a causal model was developed using various data sources. From this model it was concluded that past policies have had the greatest impact on the rate at which potential adopters have adopted broadband, however, they have not directly impacted on the pool of potential adopters thus restricting the uptake of broadband.

This conclusion was reinforced with investigation through a quantitative simulation model. In particular, it has been suggested that future policies that could potentially have the largest impact on adoption are those that target and reduce the number of households and businesses that believe they do not want the Internet, or in fact a PC, at all.

The above conclusions highlight the need for policymakers to address broadband adoption. It is hoped that through the dissemination of the results to all parties involved in the data collection process that key policymakers within Scotland will use the results to support future efforts to develop appropriate policies.

Although this paper has focused on rural and remote Scotland, it has implications for other rural and remote areas around the world. In localities as diverse as Alaska (Hudson 2006), Scandinavia (Lindmark and Björstedt 2006) and Germany (Lattemann 2006) there is a desire to ensure that broadband is available and then adopted by all those who live in rural and remote areas as well as urban areas. Through identifying the drivers of adoption as well as appropriate policy initiatives, the models suggested here could be mapped onto other geographies. Although the model developed in this paper focuses on rural and remote Scotland, much of its structure would apply to many urban areas. Urban areas are not homogeneous, differing in terms of population densities, income and communication use. Integral to the models articulated in this paper is the notion that adoption is driven by a series of uses that differ depending on the circumstances of the business or household. The models have identified in which general areas future policy initiatives should be developed. By understanding the impact of the different services on adoption, future research will result in more tailored policy initiatives. Future research can also be undertaken that targets those who express no interest in adopting broadband. The understanding of how services impact on this group is unclear, and thus crucial to increasing broadband adoption.

References

Andersen D and Richardson G (1997). Scripts for group model building. *Syst Dynam Rev* **13**(2): 107–129.

Bass FM (1969). A new product growth model for consumer durables. *Mngt Sci* **15**(5): 215–227.

Bauer JM, Kim JH and Wildman SS (2005). Effects of national policy on the diffusion of broadband in OECD countries. Paper presented at the *UFL-LBS Workshop the Future of Broadband: Wired and Wireless*, Gainsville, FL, February.

Biggs P and Kelly T (2006). Broadband pricing strategies. *Info* **8**(6): 3–14.

Brignall S (2006). Broadband and phone services going down the tube, say customers. *The Guardian*, 14 November, available at technology.guardian.co.uk

Cava-Ferreruela I and Alabau-Mun˜oz A (2006). Broadband policy assessment: A cross-national empirical analysis. *Telecommun Pol* **30**(8/9): 445–463.

DEFRA (2005). ICT in England's rural economies: A final report to DEFRA, available online at www.defra.gov.uk

Fildes R and Kumar V (2002). Telecommunications demand forecasting—A review. *Int J Fore* **18**: 489–522.

Forrester JW (1961). *Industrial Dynamics*. Portland, OR: Productivity Press.

Forrester JW and Senge PM (1980). Tests for building confidence in system dynamics models. In: Legasto AA Jr, Forrester JW and Lyneis JM (eds). *TIMS Studies in the Management Sciences*, vol. 14. Amsterdam: North-Holland.

Fransman M (2006). *Global Broadband Battle—Why the US and Europe Lag While Asia Leads*. Stanford, CA: Stanford Business Books.

Galloway L and Mochie R (2005). The use of ICT in rural firms: A policy-orientated literature review. *Info* **7**(3): 33–46.

General Register Office for Scotland (2004). Household projections for Scotland: 2002-based, available online at www.gro-scotland.gov.uk

Geroski PA (2000). Models of technology diffusion. *Res Pol* **29**(4/5): 603–625.

Hollifield CA and Donnermeyer JF (2003). Creating demand: Influencing information technology diffusion in rural communities. *Gov Inform Quart* **20**(2): 135–150.

Hope S, Murray L and Martin C (2004). *In-Migration to the Highlands & Islands*. Edinburgh: NFO Soc Research.

Hudson HE (2006). *From Rural Village to Global Village*. Mahwah, NJ: LEA.

ITU (2003). *Birth of Broadband—ITU Internet Reports*. Geneva: ITU.

ITU (2006). *Trends in Telecommunication Reform 2006—Regulating in the Broadband World*. Geneva: ITU.
Kim M and Kim H (2004). Innovation diffusion of telecommunications: General patterns, diffusion clusters and differences by technological attribute. *Int J Innov Mngt* **8**(2): 223–241.
Lattemann C (2006). *Public Private Partnerships im Brietband-sektor—Fallbeispiele aus den La¨ndern Schweden, Großbritannien, Frankreich und Deutschland*. Germany: MMC Multimedia Campus, Kiel.
Lindmark S and Bjo¨rstedt P (2006). The Swedish broadband market. In: Fransman M (ed). *Global Broadband Battle—Why the US and Europe Lag While Asia Leads*. Stanford, CA: Stanford Business Books.
Lyons MH, Adjali I, Collings D and Jensen KO (2003). Complex systems models for strategic decision making. *BT Technol J* **21**(2): 11–27.
Madden G and Simpson M (1997). Residential broadband subscription demand: An econometric analysis of Australian choice experiment data. *Appl Econ* **29**(8): 1073–1078.
Mahajan V, Muller E and Bass FM (1990). New product diffusion models in marketing: A review and directions for research. *J Marketing* **54**(1): 1–26.
Mahajan V, Muller E and Wind Y (2000). *New-Product Diffusion Models*. Dordrecht, Netherlands: Kluwer Academic Publishers.
Maier FH (1998). New product diffusion models in innovation management—A system dynamics perspective. *Syst Dynam Rev* **14**(4): 285–308.
Milling PM (1996). Modeling innovation processes for decision support and management simulation. *Syst Dynam Rev* **12**(3): 211–234.
National Statistics (2003). *Yearbook 2003—The Official Yearbook of the United Kingdom of Great Britain and Northern Ireland*. London: Office of National Statistics.
Ofcom (2004). *The Ofcom Internet and Broadband Update*. London: Ofcom, April.
Ofcom (2006a). *The Communications Market: Nations and Regions*. London: Ofcom, April.
Ofcom (2006b). *The Communications Market: Nations and Regions—Scotland*. London: Ofcom, April.
Ofcom (2006c). *Media Literacy Audit: Report on Media Literacy Amongst People in the Nations and Regions*. London: Ofcom, April.
Ofcom (2006d). *The Communications Market 2006*. London: Ofco, 10 August.
Ofcom (2006e). *Policy Implications Arising from the Communications Market: Nations and Regions Research*. London: Ofcom, 12 October.

Oh S, Ahn J and Kim B (2003). Adoption of broadband internet in Korea: The role of experience in building attitudes. *J Inform Technol* **18**(4): 267–280.

Osborne J (1999). Dynamic modelling to assist in the understanding of consumer take-up and the diffusion of new telecommunications services. Paper presented at the. *17th International Conference of the System Dynamics Society and 5th Australian & New Zealand Systems Conference*, 20–23 July, Wellington, New Zealand.

Ovum (2006). *UK broadband status summary March 2006—covering the period October 2005 to end December 2005. A Report for the Department of Trade and Industry*. London: Ovum.

Parker PM (1994). Aggregate diffusion forecasting models in marketing: A critical review. *Int J Forecast* **10**: 353–380.

Registrar General for Scotland (2005). *Projected Population of Scotland (2004-based)*. Edinburgh: HMSO.

Roberts JH and Lattin JM (2000). Disaggregate-level diffusion model. In: Mahajan V, Muller E and Wind Y (eds). *New-Product Diffusion Models*. New York: Springer.

Robertson A, Soopramanien D and Fildes R (2004). Understanding residential Internet service adoption patterns in the UK. *Teletronikk* **100**(4): 84–94.

Rogers EM (2003). *Diffusion of Innovations*, 5th ed, New York: Free Press.

Saeed K (1998). Constructing reference mode. Paper presented at the *16th International Conference of the System Dynamics Society*, Quebec City.

Savage SJ and Waldman D (2005). Broadband internet access, awareness and use: Analysis of United States household data. *Telecommun Pol* **29**: 615–633.

Scottish Enterprise (2004). Still time to plug into broadband payment, 26 January, available at www.scottish-enterprise.com

Scottish Executive (2001a). *Connecting Scotland*. Edinburgh: Scottish Executive.

Scottish Executive (2001b). *Digital Inclusion*. Edinburgh: Scottish Executive.

Scottish Executive (2002). *Making it Happen*. Edinburgh: Scottish Executive.

Scottish Executive (2005). *Scotland's People: Annual Report: Results from the 2003/2004 Scottish Household Survey*. Edinburgh: Scottish Executive.

Sterman JD (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Chicago: Irwin/McGraw-Hill.

Strover S (2001). Rural Internet Connectivity. *Telecommun Pol* **25**(5): 331–347.

Teng JTC, Grover V and Guttler W (2002). Information technology innovation: General diffusion patterns and its relationships to innovation characteristics. *IEEE Trans Eng Mngt* **49**: 13–27.

Tookey A, Whalley J and Howick S (2006). Broadband diffusion in remote and rural Scotland. *Telecommun Pol* **30**(8/9): 481–495.

Varian H, Litan RE, Elder A and Shutter J (2002). The net impact study: The projected economic benefits of the Internet in the United States, United Kingdom, France and Germany, available online at http://www.netimpact-study.com

Vigden R, Francis D, Powell P and Woerndl W (2004). Web service business transformation: Collaborative commerce opportunities in SMEs. *J Enterprise Inform Mngt* **17**(5): 372–381.

Warren M (2004). Farmers online: Drivers and impediments in adoption of Internet in UK agricultural businesses. *J Small Bus Environ Develop* **11**(3): 371–381.

Whalley JL (2005). Local loop unbundling in the United Kingdom: Strategic reviews, institutional responses and equivalence. Paper presented at *International Telecommunications Society European Regional Conference*, Porto, Portugal, September.

Part IV

System Dynamics in Healthcare

System Dynamics Mapping of Acute Patient Flows

D.C. Lane and E. Husemann

Introduction

In the United Kingdom the majority of healthcare is provided by the state, funded from general taxation. The provider—the National Health Service, or 'NHS'—is organized via a complex web of geographically based coordinating bodies and semi-autonomous hospital trusts. The NHS has around 90,000 physicians and 300,000 nurses. It handles one million scheduled visits to GPs, 33,000 unscheduled visits to A&E and

Journal of the Operational Research Society (2008) **59**(2), 213–224.
https://doi.org/10.1057/palgrave.jors.2602498
Published online 26 September 2007.

D.C. Lane (✉)
Henley Business School, University of Reading, Reading, London, UK

London School of Economics, London, UK

E. Husemann
London School of Economics, London, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_13

25,000 operations daily. Each of these days costs £140 million (Euros 210 million) (Royston 2005, www.mashnet.org.uk). Although public support for a system of universal healthcare, 'free at the point of delivery' and equitable in its handling of individuals, has remained high, over the last two decades there has been an increase in public concern about unacceptable performance on various fronts. The three Labour administrations first elected to office in 1997 have had the improvement of patient experience within the NHS as a continuing campaigning point and many improvement programmes and policy innovations have resulted.

This paper concerns the development and use of a hybrid form of qualitative mapping derived from system dynamics (SD) (Forrester 1961) and used to study flows of acute patients within the NHS. The DoH provided funding and broad direction, with the aim of improving the experience of patients within the NHS. This paper briefly describes all stages of the project but concentrates on the activities associated with the workshops.

Initiating the Project

The DoH's Health Services Division ('HSD') initiated the project, prompted by an SD simulation study of A&E waiting times. That study dealt with patients in A&E but also with those on wards and on waiting lists for scheduled operations. By taking an aggregate view it had been possible to create a broad-brush model of the system and then parameterize it with reference to a collaborating hospital (Lane et al. 2000, 2003). A short account (Lane et al. 1998) had been distributed throughout the NHS and was seen by staff in the HSD who made contact with a view to extending the work by setting the handling of acute patients in an even wider context. Their initial idea was to become more familiar with approaches for building broader models; ones which dealt with various interconnected NHS processes involving acute patients. At this stage, the different experiences with modelling that NHS staff and the DoH had had informed their ideas on the shape that any project should take.

A recent review concluded that one of the 'unique selling points of significant strength within the British OR research agenda [is] … applications in health care' (Bouyssou et al. 2004) and this is evidenced by a previous *JORS* collection on this topic (Davies and Bensley 2005). So DoH staff have experience of a range of OR approaches. This range includes 'problem structuring methods' (PSMs), which help different stakeholders to explore and then agree on a problem definition (Rosenhead 1989) and some 'soft' systems approaches (Checkland 1996). Tools from the SD field were also known (Royston et al. 1999) but the concentration was on the qualitative 'systems thinking' popularized by Senge (1990). Methods were developed for applying the systems thinking tools of causal loop diagrams and archetypes specifically to healthcare issues, these becoming known in the NHS as 'whole systems working' (Pratt et al. 1999). At the more quantitative end of OR approaches, DoH staff in the internal 'Economics and OR' (EOR) groups have a record of analysis of patient flows (Bensley et al. 1995) and they, along with other groups working with the NHS, have a long and continuing history of such analysis (Davies and Davies 1994; Millard and McClean 1996; Bennett and Worthington 1998; Gallivan et al. 2002; Brailsford et al. 2004). Generally this work has used discrete event simulation and concentrated on specific parts of hospitals or particular treatment types (Harper and Shahani 2002; Ashton et al. 2005; Griffiths et al. 2005).

An interest in bringing more breadth into their models, combined with an understanding that healthcare systems contain multiple interconnections, attracted DoH staff to the idea of experimenting further with the usefulness of SD. They wanted a process which ensured that any models did not present themselves as 'black boxes', understood only by their expert builders (Lane 1992; Pidd 1992). However, they also wanted something distinct from the PSM-like mapping approach of systems thinking. They saw a need for a systems modelling approach involving formal representations of system structure which also preserved the spirit of their PSM work by drawing many people into the analysis, not just a small group with specialist training in the approach. The range of interests described here resulted in the DoH agreeing to fund the research project. The exact focus of the project is outlined next.

Project Purpose and Scope

Though initiated by the DoH's HSD this study was aimed at contributing to the work of the Emergency Services Action Team ('ESAT') a high-level policy group reporting to the Secretary of State for Health. The project details were agreed with a Steering Committee of staff members from HSD and from elsewhere in the DoH, including the EOR Division. The main features that were agreed fitted the HSD's needs, while remaining within their very practical funding and timing/access constraints. These features were threefold; the organizing concern of the project, the main elements of the work and the intended outputs. These are now described briefly.

The concern was broad, being: 'the wider patterns of patient management in acute hospitals, and patient blockages in the whole system'. As well as considering existing patient experiences and the quality of service provided, we would look at what might happen to patients if different levels of resources were applied or if different treatment pathways were used. The idea was to identify possible ways of improving things. We agreed the following three main elements of the work. First, creating an interim qualitative system map of a 'general acute hospital', that is hospitals admitting urgent as well as scheduled cases. Second, running workshops for invited NHS staff (who were assumed to have no previous SD experience).

The workshop aims were to allow the correction of the map, and to facilitate discussions about ways of improving patient management. Subsequent analysis and summarization of the workshop responses was the third element.

The two intended outputs flowed directly from these elements. First, the workshops would provide an opportunity to assess the benefit of applying SD ideas within the NHS, particularly involving staff not expert in the approach. Second, ESAT and HSD would receive a report on the re-worked system map, along with a set of systemically informed suggested interventions for improving the processes of patient management.

This was an exciting opportunity, affording excellent access to NHS professionals.

Designing an Approach: Practical Constraints and Methodological Issues

The ambitious aims combined with the practical constraints of the project had raised some methodological issues. The Steering Committee wanted 20–30 NHS staff involved in the daily work of patient treatment to take part in the project and to suggest improvements. The knowledge provided by this group would be the 'database' of the project (Forrester 1992), the workshops acting as a channel for their ideas, with HSD and ESAT receiving those ideas. The modelling was constrained to being qualitative only. The resulting map had to be as general as possible and approved by all participants. Finally, time with the participants would be strictly limited; the workshops were restricted to a half day.

As referred to earlier, the tools of SD modelling have the potential to contribute distinctively to healthcare issues this being recently evidenced by work in a special issue of the *American Journal of Public Health* (Homer and Hirsch 2006; Leischow and Milstein 2006; Sterman 2006). The question here was which selection of tools would best suit the needs of this project. Given the constraints specified by DoH, the participative development of an SD simulation model was not appropriate. Ways exist for developing such models with groups (Vennix 1996; Andersen and Richardson 1997), including groups of healthcare workers (Vennix and Gubbels 1992). However, these require more time commitment from participants than this project afforded and would arguably not achieve the large-scale staff involvement that was required. On the other hand, we had concerns about the extent to which systems thinking alone could make a contribution in this case. PSMs can help groups agree a common definition of the problem and help them explore ways of dealing with it (Rosenhead 1989). This can involve mapping to represent ideas. In contrast, SD focuses on the causal mechanisms of a system and how they might produce different modes of behaviour over time. A rigorous understanding of that structure/behaviour link requires equation formulation, parameterization and model simulation. The version of SD popularized by Senge (1990) introduced feedback thinking to a broad audience in the form of word/arrow 'causal loop diagrams'. However, many readers did not follow Senge's advice about the benefits of simulation models. The resulting systems thinking movement has an ambivalent relationship with the SD field. The field's originator, Jay Forrester (1994), commented: 'Systems thinking is coming to mean little more than thinking about systems, talking about systems, and acknowledging that systems are important. In other words, systems thinking implies a rather general and superficial awareness of systems'.

Simply ignoring feedback certainly can significantly reduce the effectiveness of policies (Sterman 1989; Kleinmuntz 1993) and so systems thinking can be very helpful because it sensitizes people to the possible existence and consequences of feedback. Nevertheless, there are limits to the contribution that such maps can make. System dynamicists are very aware of the need for a rigorous approach to understanding behaviour (Sterman 1994b; Moxnes 1998). Maps alone cannot achieve this (Richmond 1994; Booth-Sweeney and Sterman 2000) and there are particular problems with the maps used in the whole systems working approach—causal loop diagrams (Richardson 1986; Lane 2000, 2008; Warren 2004) and archetypes (Sterman 1994a; Lane 1998).

The project therefore needed a mapping approach located somewhere between systems thinking mapping and SD simulation, an approach which offered more rigour via its maps but which could be implemented within the tight constraints of the project. The approach that we subsequently designed for the project had two parts (see Figs. 1 and 2). The preliminary activities were planned to consist of interviews with DoH and NHS staff and site visits to three hospitals. These would be followed by three workshops around Britain and analysis of the contributions elicited during them.

Methodologically, the most important mapping decision was to use stock/flow diagrams—SFDs (Forrester 1961, 1968). This allowed for the representation of system stocks and flows along with information feedback effects. However, as described later, this required the subsequent creation of both a conceptual framework to introduce these maps to participants and an associated set of questions to help them use them.

A key question was how to balance the content and process aspects of the project (Eden 1987). Clearly we had to provide some modelling content involving sound systems analysis. Also needed, however, was a num-

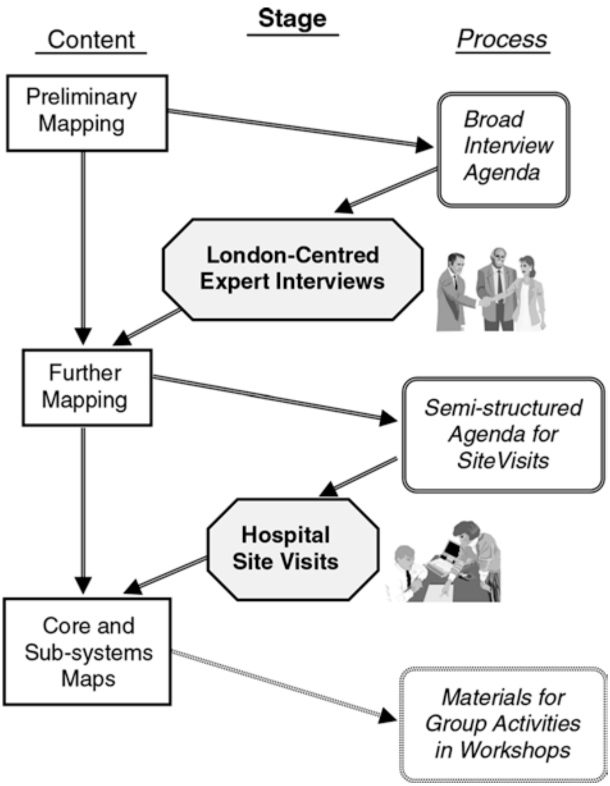


Fig. 1 Preliminary activities in the acute patient flow mapping project

ber of appropriate group processes which would help participants to work with that analysis and make them feel involved and committed (Eden et al. 1983). The following account describes both the content and process aspects. It deals with the preliminary activities in a single section but then devotes more space to the workshop activities.

Preliminary Activities

These activities are illustrated in Fig. 1. We started with some preliminary mapping of patient flows based on previous experience from the A&E project and examination of DoH reports. The broad aspirations of the

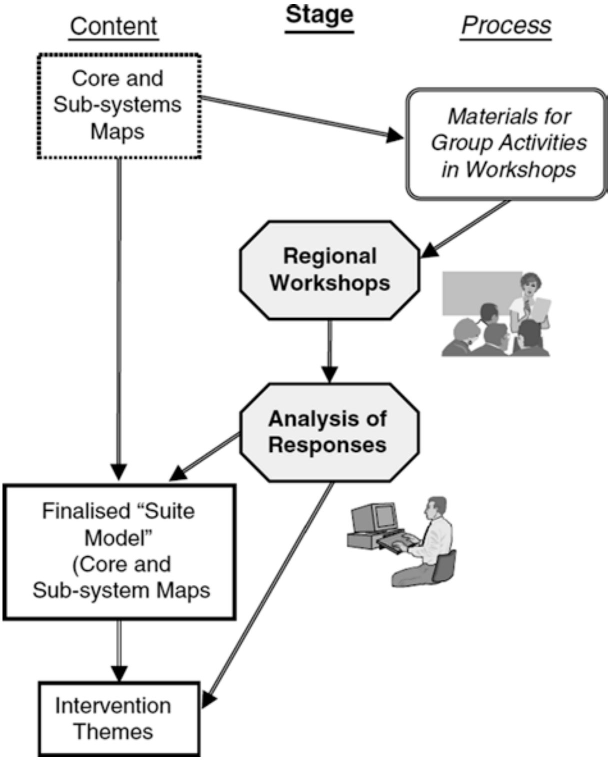


Fig. 2 Workshop activities

project meant that flows of patients between many different parts of the NHS had to be included. This tentative content generated questions for the subsequent interviews.

To further improve the quality of the maps interviews were undertaken with individuals in London: three with DoH experts and two with senior NHS staff in hospitals. Using the preliminary maps and the interviewee responses we did further mapping. At this stage we had a single map, showing stocks of patients and flows of patient movements, as well as influencing factors. We used this to create an approach for the site visits. The single map contained a great deal of detail. It included the processing of patients by GPs, by A&E departments and by out-patient clinics. We included in the maps patients scheduled for elective surgery. Finally, we represented the flow of patients into community care.

We conveyed this complexity during the site visits using one overall map and a series of different versions which also included more detail of one of the different sectors just mentioned. We photocopied the maps onto A1 paper and used them as the basis of the interviews. Essentially, this over-all map created a context for the discussion. Staff were then shown the various other maps and asked to comment on their deficiencies and adequacies. This approach meant that rather than swamping interviewees with all of the information collected in one go, they were instead able to respond to different parts in detail at different times, in turn focussing the discussion on each aspect of acute patient management. However, across the whole series of interviews the contents of all of the maps was exposed to critical comment.

We used this approach during visits to three hospitals: Peterborough, Oxford Radcliffe and King's London. The comments collected led to the idea that the project needed a slightly different approach to the mapping and workshops. We therefore created a Core Map which gave an overview of patient flows (both acute flows and other presentation pathways) and then a series of separate, more detailed maps which concentrated on the detailed features of some of the sub-systems.

Conceptual Framework for NHS Resources and Pathways

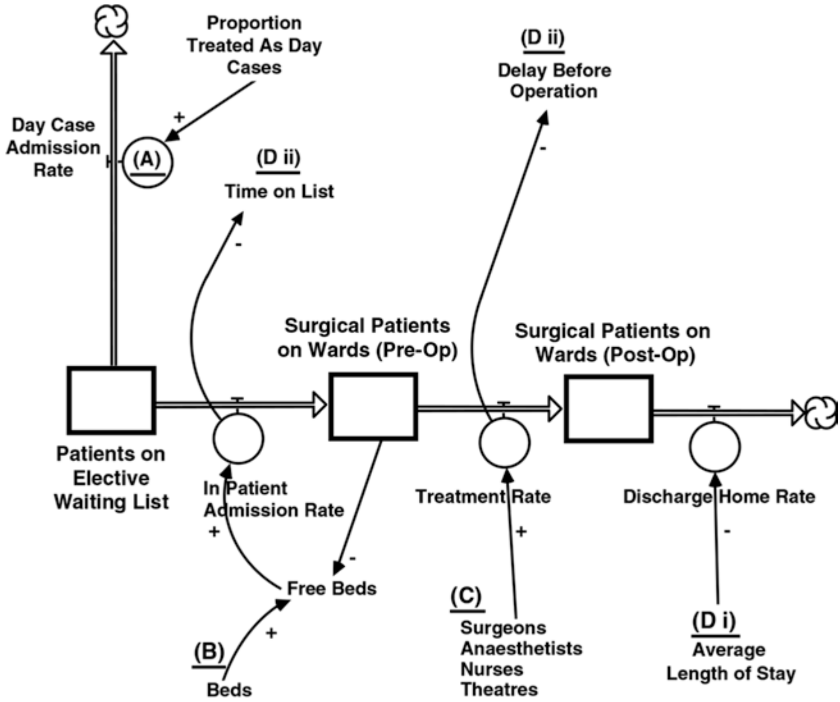
The workshop-related activities are illustrated in Fig. 2. A critical element was the design of a process that would help the attendees to understand the existing maps, to comment on them, and then to work with the mapping symbols to create ideas about how to change the system. It was at this point that the practical constraints and methodological issues described above became most relevant to the approach used. Something was required which could be used with very little prior knowledge but which still offered a level of rigour by drawing on key SD ideas and helping people to represent their understanding of how the system components fitted together and influenced patient services. What we came up with was a hybrid approach with SFDs, involving a 'conceptual framework' and an associated, healthcare-related set of five generic questions.

To ensure that the SFDs could be used by NHS staff with no previous SD background, SFD mapping was introduced using a 'conceptual framework' for mapping treatment stages, patient flows and influencing factors (Fig. 3). Those stock/flow ideas were then used to ask healthcare workers a series of five questions which were both very general in systems terms but also tailored to the requirements of the NHS. The framework and questions were designed to interest NHS staff and to draw them into a discussion concerning the larger maps.

The framework is couched in commonsense language yet within it lurk some powerful ideas from SD. The rectangles are state variables; in SD language, stocks or levels. The double-lined arrows indicate flows. These influence stocks via accumulation or draining processes. The round valves and taps indicate the amount of patients moving between treatment stages. The single lines show instantaneous causal influences. We introduced the framework using a simplified healthcare example; patients waiting for some type of treatment. After diagnosis the patients go onto a waiting list (box lower left). Perhaps it is possible to perform the treatment on an outpatient basis. In that case the patient moves fairly quickly up the day care flow on the left of the figure and out of this little system. Otherwise the patient must wait for admission to a ward (doubled-lined arrow into central box), which naturally requires that a bed be free. The treatment happens when the necessary resources are applied and the patient recovers on the ward (box lower right) before being discharged.

This example visualizes some of the influences on the service that patients receive. It also serves to help people learn the language of SFDs. This then helps them to understand the more complex SFD maps and to think in stock/flow terms. It also relates the five generic questions to stock/flow thinking. Those questions come down to the mnemonic 'ABCDE' (Fig. 3).

'A' stands for 'Alternative Pathways'. In the example a patient might be treated elsewhere than in hospital and so takes a different (upward) path. Generally, the framework encourages users' creativity by asking the question: *Are there possible alternative pathways along which patients could be processed*?



Alternative Pathways	Are there possible alternative pathways along which patients could be processed?
Blocking Resources	What resources can passively limit patients flowing to the next stage?
Conversion Resources	What resources actively create throughput to the next stage?
Dwell Times	How long do patients wait in different stages?
Extra Comments	Is there any other observation that anyone wants to make?

Fig. 3 Conceptual framework for NHS resources and pathways

'B' stands for 'Blocking Resources'. In the example, patients cannot be admitted without free beds. This contains the idea that most stocks have finite capacity, so when they are filled their inflows are blocked, or shut down. The general question is: *What resources can passively limit patients flowing on to the next stage*?

'C' is 'Conversion Resources'. If B's are passive then C's are active. In many cases patients convert from one healthcare stage to the next because resources such as surgeons and operating theatres are brought to bear. The framework poses the question: *What resources actively create throughput to the next stage*?

'D' is 'Dwell Times in Stages'. These durations come in two forms. Sometimes D's are simple inputs (Di): in the example the average length of stay is the recuperation time; adding more nurses will not reduce this. Alternatively, D's can be outputs (Dii): as in the two cases here (waiting for admission and waiting for an operation) D's can depend on flow rates which themselves depend on B and C type resources. Here the framework makes people ask: *How long do patients wait in different stages*?

'E' stands for 'Extra Comments'. This is a catch-all which allows users to express any other thought or idea which did not seem appropriate or was in any way excluded by the previous categories. So: *Is there any other observation that anyone wants to make*?

The mnemonic was chosen to assist the discussions by helping participants remember and understand the set of questions being posed. In systems thinking terms this framework embodied some simple but powerful mapping symbols and some fundamental ideas from SD. The A's explore range of possible conserved flows (=treatment paths) and therefore help users to think about the different stocks (=treatment stages) that might exist. The B's are a non-technical way of looking for negative feedback loops and flow shutdown non-linearities. The C's adopt a resource-based view (Wernerfelt 1984) of why flows actually take place, what their enabling influences are. Finally, the D's are important performance measures. However, they also distinguish between two important concepts in SD: those dwell times resulting from 'uncapacitated delays' (Di, resource-independent inputs) and dwell times arising from 'capacitated delays' (Dii, resource-dependent outputs). The ABCDE mnemonic therefore aimed to deliver both process and content benefits.

This framework also visualizes some of the influences on patients' experience and on the quality of service that they receive. The A's elicit comment on whether patients are directed to the part of the NHS most appropriate for their needs. The B's evoke effects which stop patients receiving the service they would wish for, while the C's deal with the resources which advance their treatments. Finally, the D's bring to the surface a key aspect of patient experience and service quality: the waiting times encountered in moving through the NHS.

We sent this framework to the NHS staff who had agreed to attend the workshops. We also sent a copy of the Core Map (see Fig. 4) and asked them to spend a few minutes looking at these materials in preparation for the day. To support the use of this framework we created other materials. These are described next, in the context of the workshop.

Running the Workshops

Three workshops were run over a 2-week period in London (at the headquarters of the British Medical Association), Sheffield and Manchester. In total, 43 people attended the workshops, drawn from middle to senior levels in the NHS, all involved in the daily work of patient treatment. These included hospital bed managers, A&E physicians, healthcare strategists, a Director of Nursing, a Social Services manager and a manager from the Ambulance Service. This group therefore represented a broad cross-section of NHS activities.

The format of the day is shown in Table 1. During the informal start each participant received a personalized pack of materials. By way of introduction, both the core map and the conceptual framework were reviewed, and the format and intended benefits of the day described. Participants were then split into four groups of three or four people (each participant's pack designated a group, arranged to get a good mix of job functions). To obtain a degree of triangulation, two groups looked at the work of local doctors and two at how emergency treatment works. In prepared break-out rooms, each had four items to work with (A0 charts taped to the wall). These were: the core map, the conceptual framework,

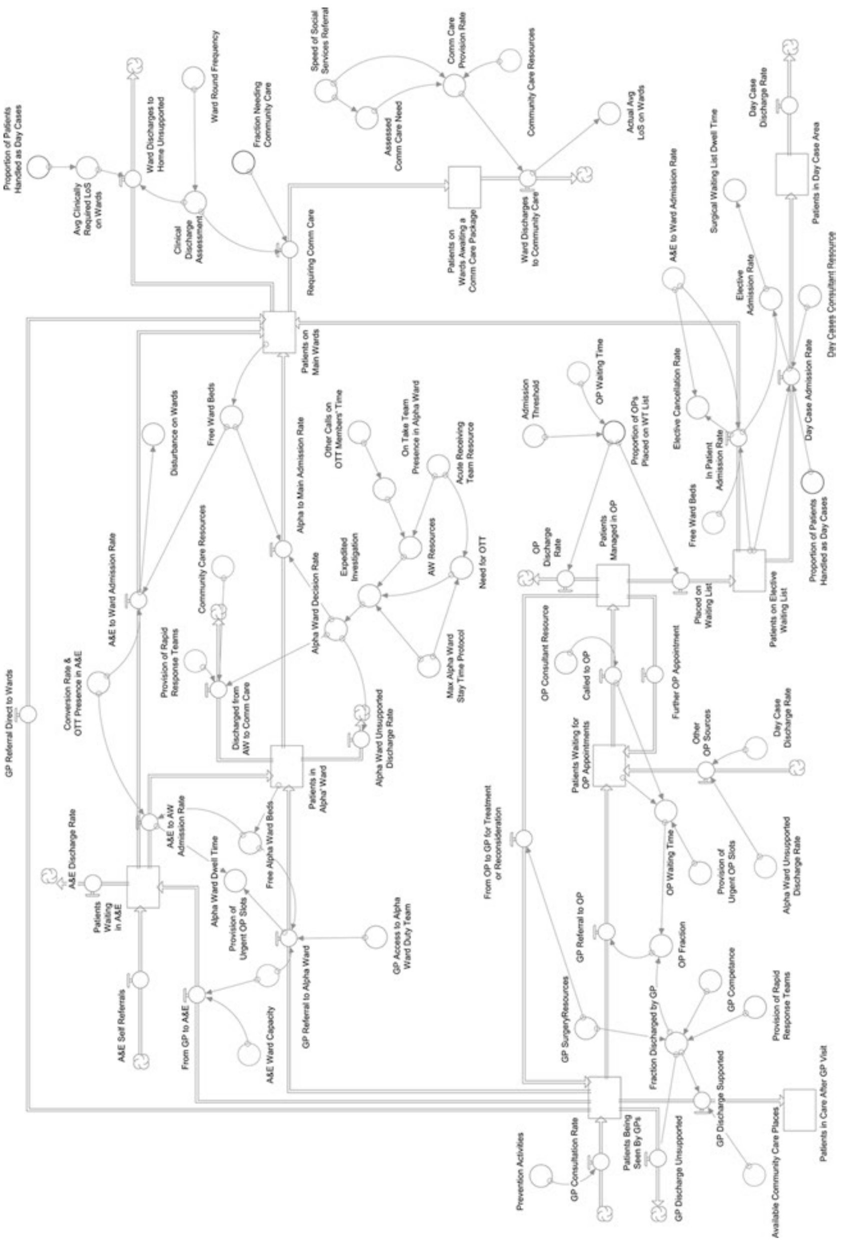


Fig. 4 Core Map of pathways for acute patients

a detailed sub-map of their area with ABCDE points indicated, a task structure listing a set of ABCDE questions concerning the experience of patients flowing through the area of the system. Examples of these last are shown in Fig. 5 and Table 2. Group members were asked to use the sub-map and deal with the questions on the task structure. A designated 'recorder' wrote responses on flipchart pages.

The majority of participants found the framework and the maps self-explanatory. A little encouragement and clarification helped the remaining participants to join in. The groups fairly rapidly adopted the language and symbols of stocks and flows as a way of handling existing and alternatives treatment pathways. Causal links were equally unproblematic. Generally, participants found the provided maps reasonable and useful. When they agreed changes they drew new stocks, flows and links onto the maps. In this way inaccuracies were corrected and important new elements introduced.

The ABCDE question convention worked. In process terms, participants related well to the questions, understanding their main point but feeling able to discuss possibilities in a creative manner while 'operating' within each set question. The simple ABCDE naming made group members remind others of questions still to be asked, or prompt the group to cycle back to overlooked questions. It was also clear that individual questions were being answered in the broader, more systemically aware, context provided by the maps and with consideration of the system complexity that those maps illustrated.

In plenary one group from each task briefly presented their observations and comments from all other participants followed.

After lunch four different groups were formed (Table 1), concentrating on main hospital wards and what had been labelled 'alpha wards' (specialist wards where a concentration of resources allows more rigorous discharge protocols; patients thought to require only short stays are dealt with here). Again, these groups had detailed sub-maps and task structures with ABCDE questions. Another plenary feedback session followed. The day closed with a description of how the project would go forward.

Post-workshop Activities

The workshops generated annotated maps, flipchart responses to the ABCDE questions, tapes of the plenary sessions and contemporaneous notes from the group work. The final stage therefore involved the analysis of these materials (Fig. 2). The first part of this was the finalization of the maps.

System dynamicists aim for maps to have high 'face validity' (Forrester and Senge 1980; Frankfort-Nachmias and Nachmias 1992). In a workshop setting emphasis falls on the subjective understanding that groups have of how a system works now, and how it might be made to work (Eden et al. 1979; Checkland 1995). Workshop participants had questioned and changed the maps thereby increasing the level of confidence that they were accurate and included the most important system elements. Creating final versions was straightforward and did not require major reworking. Emerging from this analysis was the 'Suite Model', a set of SFDs derived from the Core Model and the five detailed submaps. Having been subjected to the comments, changes and corrections of the workshops, this expressed the participants' shared understanding of the stocks, flows and influences of the different parts of the NHS through which acute patients (and others) pass, as well as displaying the understanding of the system which gave rise to the proposed changes.

The second part of the post-workshop analysis involved finding the general ideas and themes that participants had come up with about those

Table 1 Format of the three workshops

9:30	'Meet & greet' period
10:00	Presentation: Introduction and scene-setting
10:30	First pair of parallel group tasks:
2 × GP Activities 2 × A+E Activities	
12:00	First plenary feedback session
12:30	Lunch
13:00	Second pair of parallel group tasks:
2 × Main wards Activities 2 × 'Alpha Wards' Activities	
14:15	Second plenary feedback session
14:45	Presentation: What will/could happen next
15:00	Close

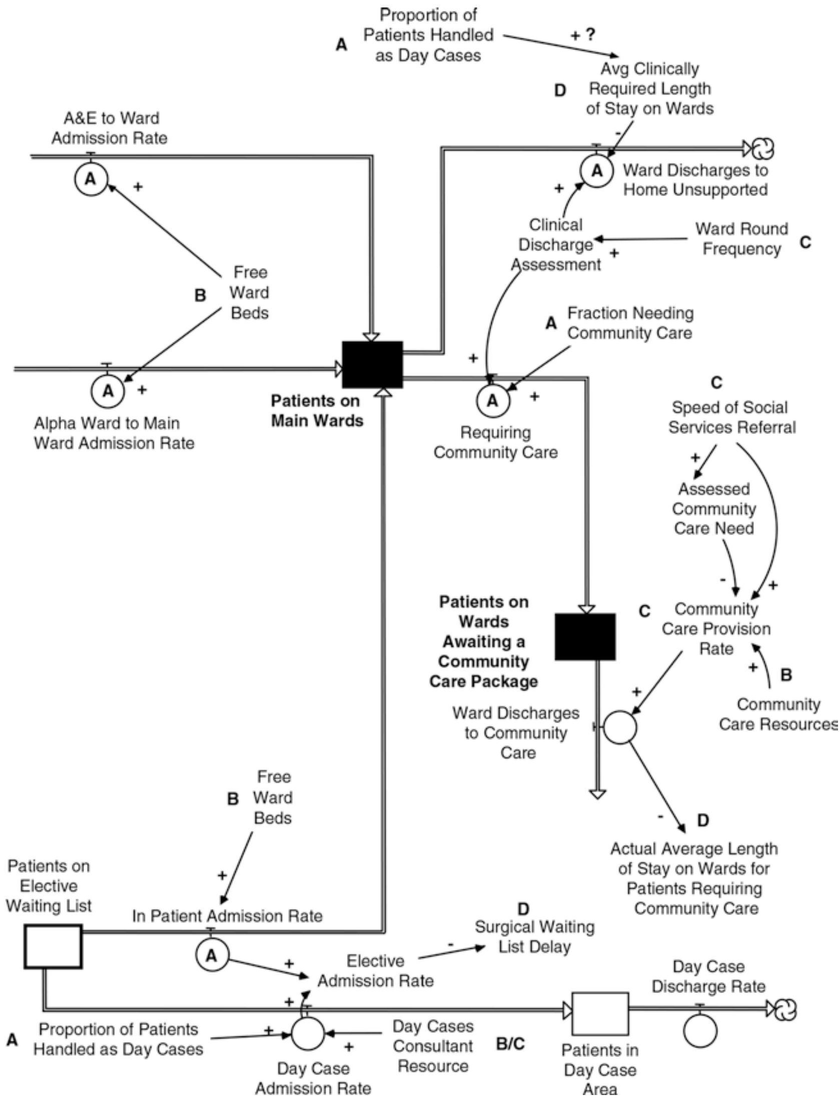


Fig. 5 Sub-map of the acute patient flows into and out of a hospital's main ward

Table 2 Detailed task structure for acute patient flows into and out of a hospital's main wards

Main wards task structure
Please use the following framework to structure your group discussion. Write your comments on the flip-chart, labelling them 'A1', 'A2' etc.
Alternative pathways
(A1) Flowing into and out of the two main rectangles, are there any *additional* important pathways (A)?
(A2) What influences the 'Proportion of Patients Handled as Day Cases'? What are the advantages and disadvantages of increasing this proportion? (E.g. Does the Length of Stay for the remaining in-patients rise?)
(A3) Is the value of the variable 'Fraction Needing Community Care' that you experience too high or too low? In which direction would you like to change this value and how would you do it?
Blocking resources on flows
(B1) For each of the pathways on the map, are there any *additional* Blocking Resources other than 'Free Ward Beds' and 'Community Care Resources' (B) that limit the flow rate to the next stage?
(B2) What interventions could possibly be made to the Blocking Resources to reduce these flow limiting effects?
(B3) Indicate your judgement of the potential importance of these interventions to removing the block using the scale:
1. Slight 3. Important
2. Worthwhile 4. Very Important
Conversion resources on flows
(C1) For each of the pathways flowing out of and into (this activity), are there any *additional* resources that are needed to make the flow to the next stage possible?
(C2) Which Conversion Resources (C) would you wish to increase in order to speed up ('de-bottleneck') each of the flows?
1. Minor 3. Large
2. Appreciable 4. Very Large
Dwell times stages
(D1) Which of the two Dwell Times shown (D) cause you the most concern? Mark them on the following scale:
1. Appropriate duration 3. Longer than necessary
2. A little longer than desirable 4. Far Longer than necessary
(D2) What actions could be taken to reduce the Dwell Times (D)?
Extra comments that you want to make
Any other observations, issues or questions which you would like to record.

possible ways of intervening to improve patient experience. The various materials were analysed to form clusters of related ideas. In the task structures (see Table 2) participants had been asked to rate the importance of their ideas in a way similar to that used in Likert Scales (Frankfort-Nachmias and Nachmias 1992). This, combined with a simple frequency count of related ideas, made it possible to extract a rich set of 'intervention themes' from the material generated by the workshops.

Project Outputs

The innovative features of the project concern the following elements. First, the conceptual framework for mapping acute patient flows, second, the health-specific—yet generic—ABCDE question sets and associated detailed SFDs. However, what should be emphasized is the practical use of these materials by NHS staff in the workshops, during which the different elements were used in conjunction. These features relate to the two intended outputs of the projects: an assessment of the benefits of applying SD with NHS staff not expert in the modelling approach, and the production of an internal report recording the system maps and the suggested interventions.

The Steering Group made a positive assessment of SD as a result of this work. The set of hybrid materials had been created and successfully used. Additing together the workshop participants and the early interviews, the materials had been put to the test by around 50 NHS staff (an improvement on the project target of 20–30). As described, participants found the maps useful and the questions meaningful. The discussions helped communication among participants concerning knowledge of how the system was put together and helped them to think through the opportunities for, and consequences of, any changes using a broader, more systemic understanding. The project produced the Suite Model representing the majority of acute hospitals. These qualitative maps employed the iconography of stocks and flows to show the main patient flows of such hospitals along with more detailed maps of other processes. Although the content of this work was specific, in more general terms it demonstrated that the thinking of SD could be used to illuminate the functioning of healthcare systems. It demonstrated that mapping could be used not only in the PSM style of agreeing a problem definition but as a means of creating, exploring and agreeing actual modifications to the structure of the system in question.

It seems appropriate to quote the comments from two of the senior DoH staff on the project's Steering Committee. One wrote of the workshops, 'Colleagues who were involved in these discussions found the discussions very useful in clarifying their understanding of emergency services'. Assessing the benefits of SD, a second wrote that the project, 'improved the cognition by senior managers, working in the [A&E] services environment, of how the various components of the Acute Health care system were interrelated.'

The various maps were contained within the final report on the project, the production of which was the second intended project output. As a record of the events, all workshop attendees received a copy and its contents were subsequently communicated to ESAT (see below). The report recorded much of the detail of the project, including process maps and participant responses. The workshop responses, cross-referenced with the maps themselves, gave rise to a set of proposals for improvements in the management of acute patients. Some were ideas known already to a few who were present but the workshops gave them an opportunity to explain and explore them with other NHS staff. This helped all to consider the ideas in a broader context, or communicated them to a wider group of NHS staff—along with the reasons why they were thought useful.

The internal report presented these proposed changes in two clusters, or 'intervention themes'. One concerned ways of intervening to get a faster flow of patients along the existing pathways within the acute healthcare system. The ideas in this theme concerned; the more active management of patient progress, increased resource flexibility to cope with demand surges, a wide range of changes in NHS working and career practices to make the service more flexible, the provision of more resources/funding to extend present processes (and/or to create new delivery paths), and movement towards an all day, 7 days a week availability of facilities. The other cluster of proposed interventions concerned ways of changing parts of the systems so that patients are handled in the most appropriate part of that system. These ideas involved: minimization of variations of conditions dealt with at any point in the system (achieved by more filtering of patients and by the concentration of advanced healthcare resources in specific places), and increased availability of information to patients and healthcare workers.

The internal report described all of these in greater detail, however, a sketch of one may be helpful. An interesting aspect of the patient filtering idea was the insight that more patients could be kept away from hospitals by the appropriate redeployments of healthcare assets. Assets could be 'forward deployed' into GP surgeries (to allow a greater range of tests and treatments to be administered) or into NHS Direct (an emergency telephone service from which healthcare advice could be given). These initiatives would have the effect of avoiding admissions. There could also be 'backward deployment' of assets. For example, into home support for post-operative recuperation or into the creation of 'step down wards' which act as a bridge between main wards and a return to life back in the community. These ideas were aimed at enabling patients to be discharged from hospitals as quickly as possible. Such filtering took the view that automatically admitting all sick people to hospitals was old-fashioned and wasteful. As an alternative, many patients could be dealt with closer to home, with less disruption of their social support network and less chance of their becoming 'institutionalized', a common problem with elderly patients as long hospital stays can result in their losing their independence. With lower level healthcare being provided via different services there would then be less variation in the conditions of hospitalized patients; these facilities would be focused on the treatment of severely ill patients.

Space restriction prevents further discussion here. Even so, it merits emphasizing that systems mapping helped participants to look outside their functional specialisms when discussing such ideas. It was noticeable in the discussions that working with the maps allowed participants to judge proposals in a richer context, informed by the system maps that they were all referring to. Consequently, problems with some ideas were uncovered while new possibilities were invented. This is expressed in a comment by one of the workshop participants; 'There is no point focussing on one solution: all areas must be addressed'.

Beyond the Project

This work showed that a version of SD mapping could be used successfully by NHS staff. It added to a number of positive experiences that the DoH's EOR Division had with SD. In the words of a member of the Steering Group, the 'project represented a successful use of SD particularly in a learning context'. The DoH has gone on to conduct other SD studies and this interest is continuing. For example, at the September 2005 launch meeting of MASHnet, a network for modelling and simulation in healthcare the DoH's Head of OR, Geoff Royston, expressed his belief that SD offered a uniquely powerful combination of analytical power and user transparency (Royston 2005).

The specific proposals for improving patient experience were fed back into the DoH and, hence, the NHS. This work was presented to ESAT at BMA House, reporting on both the specific results and the general mapping approach. The report itself, in the words of one of the Steering Committee members, 'has informed the modernisation of [A&E] services initially through an [A&E] task force and subsequently through the Modernisation agency.' Another Committee member completes the story of the project described in this paper: 'In policy terms, [the] work made an important contribution to work on improving A&E departments. [The] report showed how relatively simple modifications in physical arrangements and treatment pathways might improve the delivery of service … [the] report also helped to inform the work of the Accident and Emergency Modernisation Programme, a £15 million capital investment programme overseen by an expert taskforce that was also charged with developing proposals for modernisation of service delivery in A&E. The A&E task force … discussed [the] findings in some depth. I know that they found this work useful in developing models of future service delivery…. The task force recommendations are now being implemented by the National Patients' Access Team, a Government agency working with acute hospital.'

Acknowledgements We are grateful to the Department of Health for initiating this project and to all of the NHS staff who contributed their time to it. This project was funded by the Research and Development Division of the Department of Health.

References

Andersen DF and Richardson GP (1997). Scripts for group model building. *Syst Dyn Rev* **13**: 107–129.

Ashton R, Hague L, Brandreth M, Worthington D and Cropper S (2005). A simulation-based study of a NHS walk-in centre. *J Opl Res Soc* **56**: 153–161.

Bennett JC and Worthington DJ (1998). An example of a good but partially successful OR engagement: Improving outpatient clinic operations. *Interfaces* **28**(5): 56–69.

Bensley DC, Watson PC and Morrison GW (1995). Pathways of coronary care—A computer-simulation model of the potential for health gain. *IMA J Math Appl Med Biol* **12**: 315–328.

Booth-Sweeney L and Sterman JD (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *Syst Dyn Rev* **16**: 249–286.

Bouyssou D, Forder R, Merchant S, Nance R, Pierskalla W, Roper M, Ryan D and van der Duyn Schouten F (2004). *Review of Research Status of Operational Research in the UK*. Swindon, UK: EPSRC/ESRC/ORS.

Brailsford SC, Lattimer VA, Tarnaras P and Turnbull JC (2004). Emergency and on-demand health care: Modelling a large complex system. *J Opl Res Soc* **55**: 34–42.

Checkland PB (1995). Model validation in soft systems practice. *Syst Res* **12**: 47–54.

Checkland PB (1996). *Guidelines for the Use of 'Soft Systems' Methodology*. Winchester: Information Management Group of the NHS Executive.

Davies R and Bensley D (2005). Meeting health challenges with OR (special issue). *J Opl Res Soc* **56**: 123–125.

Davies R and Davies H (1994). Modelling patient flows and resource provision in health systems. *Omega* **22**: 123–131.

Eden C (1987). Problem solving or problem finishing? In: Jackson MC and Keys P (eds). *New Directions in Management Science*. Aldershot, UK: Gower, pp 97–107.

Eden C, Jones S, Sims D and Gunton H (1979). Images into models: The subjective world of the policy maker. *Futures* February: **11**(1): 56–62.

Eden C, Jones S and Sims D (1983). *Messing About in Problems*. Oxford: Pergamon Press.

Forrester JW (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Forrester JW (1968). *Principles of Systems*. Cambridge, MA: MIT Press.

Forrester JW (1992). Policies, decisions and information sources for modelling. *Eur J Opl Res* **59**: 42–63.

Forrester JW (1994). System dynamics, systems thinking, and soft OR. *Syst Dyn Rev* **10**: 245–256.

Forrester JW and Senge PM (1980). Tests for building confidence in system dynamics models. In: Lagasto AA, Forrester JW and Lyneis JM (eds). *System Dynamics: TIMS Studies in the Management Sciences*. Oxford: North-Holland, pp 209–228.

Frankfort-Nachmias C and Nachmias D (1992). *Research Methods in the Social Sciences*, 4th edn. London: Edward Arnold.

Gallivan S, Utley M, Treasure T and Valencia O (2002). Booked inpatient admissions and hospital capacity: Mathematical modelling study. *Br Med J* **324**: 280–282.

Griffiths JD, Price-Lloyd N, Smithies M and Williams JE (2005). Modelling the requirement for supplementary nurses in an intensive care unit. *J Opl Res Soc* **56**: 126–133.

Harper P and Shahani AK (2002). Modelling for the planning and management of bed capacities in hospitals. *J Opl Res Soc* **53**: 11–18.

Homer JB and Hirsch GB (2006). System dynamics modeling for public health: Background and opportunities. *Am J Pub Health* **96**: 452–458.

Kleinmuntz DN (1993). Information processing and misperceptions of the implications of feedback in dynamic decision making. *Syst Dyn Rev* **9**: 223–237.

Lane DC (1992). Modelling as learning: A consultancy methodology for enhancing learning in management teams. *Eur J Opl Res* **59**: 64–84.

Lane DC (1998). Can we have confidence in generic structures? *J Opl Res Soc* **49**: 936–947.

Lane DC (2000). Diagramming conventions in system dynamics. *J Opl Res Soc* **51**: 241–245.

Lane DC (2008). The emergence and use of diagramming in system dynamics: A critical account. *Syst Res Behavior Sci* **25**(1): forthcoming.

Lane DC, Monefeldt C and Husemann E (2003). Client involvement in simulation model building: Hints and insights form a case study in a London hospital. *Health Care Mngt Sci* **6**: 105–116.

Lane DC, Monefeldt C and Rosenhead JV (1998). Emergency—but no accident—a system dynamics study of an accident and emergency department. *OR Insight* **11**(4): 2–10.

Lane DC, Monefeldt C and Rosenhead JV (2000). Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *J Opl Res Soc* **51**: 518–531.

Leischow SJ and Milstein B (2006). Editorial: Systems thinking and modeling for public health practice. *Am J Pub Health* **96**: 403–405.

Millard PH and McClean SI (eds) (1996). *Go with the Flow: A Systems Approach to Healthcare Planning*. London: Royal Society of Medicine Press.

Moxnes E (1998). Not only the tragedy of the commons: Misperceptions of bioeconomics. *Mngt Sci* **44**: 1234–1248.

Pidd M (1992). *Computer Simulation in Management Science*, 3rd ed. Chichester: Wiley.

Pratt J, Gordon P and Plamping D (1999). *Working Whole Systems: Putting Theory into Practice in Organisations*. London: King's Fund.

Richardson GP (1986). Problems with casual-loop diagrams (originally published in 1976). *Syst Dyn Rev* **2**: 158–170.

Richmond B (1994). Systems thinking/system dynamics: Let's just get on with it. *Syst Dyn Rev* **10**: 135–157.

Rosenhead J (ed) (1989). *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. Chichester: Wiley.

Royston G (2005). Modelling and simulation in health: Potential, achievement and challenge. *Presentation at the MASHnet launch event*, WBS, 20 September.

Royston G, Dost A, Townshend J and Turner H (1999). Using system dynamics to help develop and implement policies and programmes in health care in England. *Syst Dyn Rev* **15**: 293–313.

Senge PM (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday/Currency.

Sterman JD (1989). Modelling managerial behaviour: Misperceptions of feedback in a dynamic decision making experiment. *Mngt Sci* **35**: 321–339.

Sterman JD (1994a). Beyond training wheels. In: Senge PM, Roberts C, Ross R, Smith B and Kleiner A (eds). *The Fifth Discipline Fieldbook*. London: Nicholas Brealey Publishing, pp 177–184.

Sterman JD (1994b). Learning in and about complex systems. *Syst Dyn Rev* **10**: 291–330.

Sterman JD (2006). Learning from evidence in a complex world. *Am J Pub Health* **96**: 505–514.

Vennix JAM (1996). *Group Model-building: Facilitating Team Learning Using System Dynamics*. Chichester: Wiley.

Vennix JAM and Gubbels JW (1992). Knowledge elicitation in conceptual model building: A case study in modeling a regional Dutch health care system. *Eur J Opl Res* **59**: 85–101.

Warren K (2004). Why has feedback systems thinking struggled to influence strategy and policy formulation? Suggestive evidence explanations and solutions. *Syst Res Behav Sci* **21**: 331–347.

Wernerfelt B (1984). A resource-based view of the firm. *Strat Mngt J* **5**: 171–180.

Improving the Cost-Effectiveness of Chlamydia Screening with Targeted Screening Strategies

D. Evenden, P.R. Harper, S.C. Brailsford, and V. Harindra

Introduction

Chlamydia trachomatis is the commonest sexually transmitted bacterial infection in the UK, with 89 431 diagnoses in Genito-Urinary Medicine (GUM) clinics in 2003 (Health Protection Agency 2004), and has been widely reported in literature (Hicks et al. 1999; Department of Health

Journal of the Operational Research Society (2006) **57**(12), 1400–1412.
https://doi.org/10.1057/palgrave.jors.2602134
Published online 28 December 2005.

D. Evenden
School of Mathematics, University of Southampton, Southampton, UK

P.R. Harper (✉)
School of Mathematics, Cardiff University, Cardiff, UK

S.C. Brailsford
School of Management, University of Southampton, Southampton, UK

V. Harindra
St Mary's Hospital, Portsmouth, UK

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_14

2000; Hart et al. 2002; Honey et al. 2002) and recently in the media (eg, The Guardian 2004). It constitutes a major public health concern. The majority of infections are asymptomatic, but can lead to serious long-term sequelae including pelvic inflammatory disease, tubal infertility and ectopic pregnancy. Screening programmes have been shown to be effective in Sweden and the United States (Herrmann et al. 1991; Addiss et al. 1993). However, the methodology employed may not be suitable in the UK and there are concerns that blanket screening of the whole population at risk will add extra burden to the overstretched health economy. Recently, the UK Department of Health provided funding to introduce national Chlamydia screening of people between the ages of 16–25 in 10 centres with the view to extend this programme to the rest of the country within the next few years as part of the National Chlamydia Screening Programme (NCSP) (Department of Health 2005).

Portsmouth was one of two UK Department of Health sites chosen for an opportunistic screening trial of Chlamydia, whereby 20,000 persons in the 16–24 age range were screened (Pimenta et al. 2003a, b). Portsmouth is an island city situated on the coast of Southern England and has a population of just under 200 000. Very high levels of population coverage were achieved in the opportunistic screening trial and this was regarded as an important factor in the success of future screening interventions. An infection prevalence of around 10% was observed, with an age peak noted at 18 years. Harindra et al. (2002) provide more detailed insight of the methods and preliminary results from the trial.

This paper presents collaborating work with the University of Southampton and Consultants at the GUM Department, St Mary's Hospital, Portsmouth. The research was timely, given that the Portsmouth opportunistic trial had been completed and that it was felt that findings from this trial could help to inform the NCSP. The work was novel in crossing the boundaries of various disciplines, namely Operational Research, Statistics, Health Services Research and Geography. The methodologies adopted, as discussed in this paper, combined geomapping, statistical clustering methods, and System Dynamics (SD) modelling. The geomapping work, using the software MapInfo, allowed for the spreading patterns and infection clusters to be observed, and provided a critically important contribution to screening intervention planning. The analysis of socioeconomic indicators, using regression models and tree-based classification trees, identified high-risk groups within the population for screening intervention targeting. The SD model, built using the software Vensim, captured the infection dynamics and cost-effectiveness of screening using strategies informed by the previous two components. Overall the multiple approach that was adopted, utilizing OR and statistical methods in combination with geomapping tools, facilitated a holistic view of the problem. Thus, the recommendations that emerged to help inform health policy were considered to be well founded. They were put forward to help form local policy to support the Chlamydia screening programme and to provide more general guidance and recommendations on a cost-effective screening methodology. Although we analyse and present data from the Portsmouth Chlamydia opportunistic screening trial, the methodology adopted here could be readily applied to other geographical areas, and indeed for other sexually transmitted infections or diseases.

Although there is an extensive literature on modelling infectious disease, there is relatively little published work on modelling Chlamydia infection. A review by Honey et al. (2002) describes studies that show screening to be more cost-effective than just testing symptomatic women. The role of male partners, and the fact that men seem to be forgotten in the infection and treatment equation was recognized by Hart et al. (2002), particularly where women were screened opportunistically. Previous papers to analyse the cost-effectiveness of screening have included Haddix et al. (1995), Genc and Mardh (1996) and Buhaug et al. (1989). The analyses reported in these papers ignore risk-groups within the population and the impact of screening on prevalence. Townshend and Turner (2000) developed a SD model that overcame many of these previous concerns, and was an excellent source of guidance for this work. Gove (1997) used a Discrete Event Simulation (DES) model to evaluate screening options for Chlamydia. The research presented here is novel in that we combine geomapping, risk groupings and computer simulation techniques, allowing for each component of the work to be informed by the other components. For example, within the SD model we have included different risk groups within the population based on sexual behaviour that were identified during the statistical clustering work. Furthermore, by considering the spatial prevalence of Chlamydia over the geographical region using geomapping techniques, we have been able to determining the relationship between socio-economic indicators and prevalence by postcode, which in turn informed the parameters for the SD model.

To summarize the key research objectives:

- Geographical mapping of Chlamydia prevalence in the Portsmouth region.
- Statistical determination of relationships between socioeconomic indicators and prevalence by postcode.
- Identification of high-risk populations.
- Determination of those factors which help to plan and target screening and inform health education methods, in order to reduce the national incidence and prevalence of Chlamydia.

In order to meet the research objectives, the following methodologies were adopted and are discussed in subsequent sections of this paper:

- Preliminary analysis of the opportunistic screening trial data (Section "Analysis of the Opportunistic Screening Data").
- Geomapping analysis of the screening data (Section "Geomapping Analysis").
- Statistical analysis of prevalence and socio-economic indicators (Section "Socio-Economic Indicators and Prediction").
- A comprehensive SD model of the Chlamydia infection process, to promote an understanding and justification of the targeted screening intervention (Section "Simulation Modelling of Infection Dynamics and Costs-Effectiveness").

Analysis of the Opportunistic Screening Data

Data were collected during the opportunistic screening trial held in the Portsmouth area from October 1999 to September 2000. A number of usable subsets of data were made available for analysis, including opportunistic screening data, GUM patients and partners tracing records, and referrals from positive results in the opportunistic screen.

Over 25,000 Chlamydia tests were carried out during this trial, though these included repeat tests (eg to check that the infection had been cured) and visits by the same person to different medical facilities (and thus getting a new patient number). These raw data were recorded in an Excel file consisting of 25,553 records. Even though this particular data contained no test results, it was useful in that it contained records of patient sex, ethnic group, and postcode, all referenced by a patient index P-number. These data also included the partner tracing details where this was possible. Although these data had patient name fields, these were deleted to ensure patient anonymity. Harindra et al. (2002), and Pimenta et al. (2003a, b), provide more insight into the design of the opportunistic screening trial in Portsmouth, particularly in the context of screening for co-infections and presumptive treatment. These initial findings suggested that the screening trial, and the heavy reliance on patient cooperation, was well received in this instance. The trial itself was seen to be highly effective.

The results data for the opportunistic trial consisted of 17,342 patient records. This was provided as an Excel file and consisted of patient number, the test date, date of birth, testing locations, laboratory specimen number, and test result. To find patient postcode, the patient number (eg 2234) was reformatted to the fixed format patient index P-number (eg P02 234) using Excel text manipulation functions. This could then be cross-referenced to the 25,553 records of the raw data to extract postcode and patient details. When records were sorted by postcode in the Portsmouth area, there were just over 11,100 records pertaining to a valid postcode district (eg PO1), and just under this number pertaining to a valid postcode sector (eg PO1 1). Figure 1 shows how the various datasets were extracted and manipulated to obtain the final data set to be analysed.

The screened percentage in each postcode district is shown in Fig. 2. In order to calculate these figures, we needed first to obtain census population statistics for 30,202 women in the 16- to 24-year-age group for the PO postcode areas, which were extracted from the UK Office of National Statistics (ONS) census data (http://www.statistics.gov.uk). It was then necessary to convert the Census data from electoral ward to postcode district in order to permit comparison with the number of tests by postcode district. As shown in Fig. 2, the values achieved were large (typically

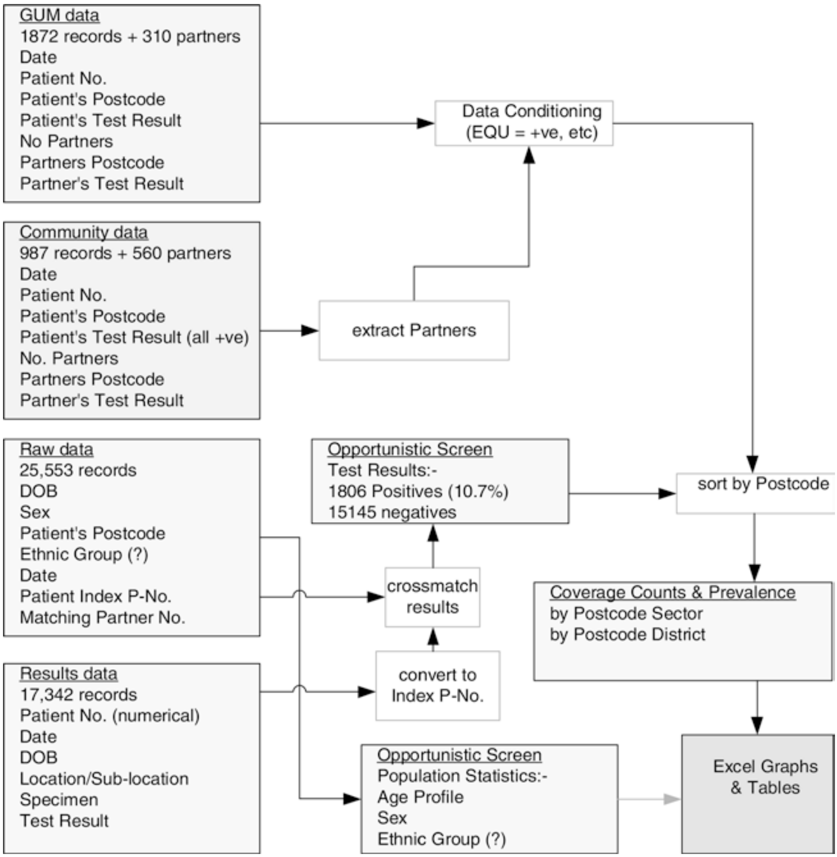


Fig. 1 Screening data extraction and manipulation

30–40%) and this provided confidence that further calculations based on these data gave results that were representative of the overall population.

Infection prevalence was calculated by patient type and is shown in Table 1. When the number of positive results or the sample size was small the exact method based on the binomial distribution was used. Eqs. (1) and (2) provided an exact solution, where N_D was the number of positives in a group of N tests, with prevalence p, and where *k* was the summing variable.

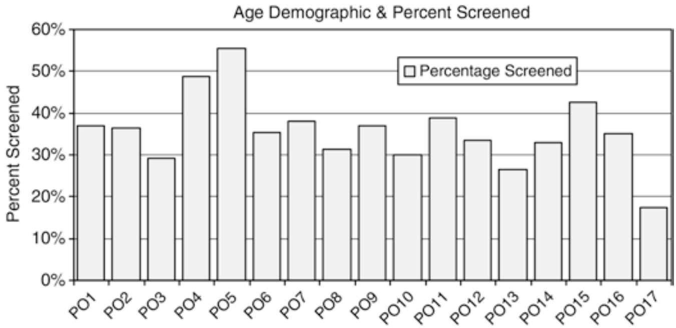


Fig. 2 Screening penetration: percentage of 16- to 24-year- old population screened by postcode district

Table 1 Prevalence by patient type

Patient group	Prevalence (%)	95% CI (%)	Sample
GUM patients	17.8	1.8	1632
Partners of all GUM patients	23.0	5.1	254
Partners of positive GUM patients	51.9	7.1	77
Partners of community referrals	45.9	4.4	488
Opportunistic screen—PO only	9.07	0.53	11,140
Opportunistic screen—raw data	10.41	0.45	17,342

For the lower CI limit (1−α)100%:

$$\begin{aligned}\text{Binom}(k; N, p) &= 0.975 \\ &= \sum_{k=0}^{N_{D-1}} \binom{N}{K} \times p_L^k \times (1 - p_L)^{N-k} \\ &= 1 - \alpha / 2 \end{aligned} \tag{1}$$

For the upper CI limit (1−α)100%:

$$\begin{aligned}\text{Binom}(k; N-1, p) &= 0.025 \\ &= \sum_{k=0}^{N_D} \binom{N}{K} \times p_U^k \times (1 - p_U)^{N-k} \\ &= \alpha / 2 \end{aligned} \tag{2}$$

This shows that prevalence among GUM patients was significantly higher than the opportunistic screen prevalence, suggesting that GUM users constitute a higher risk group. The prevalence among partners of GUM patients and partners of community-screened patients was consistent.

Of 17,342 records with patient sex and a test result available 12,653 were cross-matched. Of these 12,454 (98.43%) were female of which 8.97% tested positive. Only 199 (1.57%) records were male of which 16 (8.04%) tested positive. These results are shown in Fig. 3, along with 95% confidence intervals.

Since the confidence intervals overlap it suggested that prevalence among male and female groups was not significantly different. This was confirmed using a χ^2 test for homogeneity where $\chi^2 = 0.2$, df = 1, and P-value = 0.65, suggesting homogeneity between male and female results (Yates' correction to this method provides the same conclusions). For this reason, and since the number of men in the sample was small, no further distinction was made between male and female records.

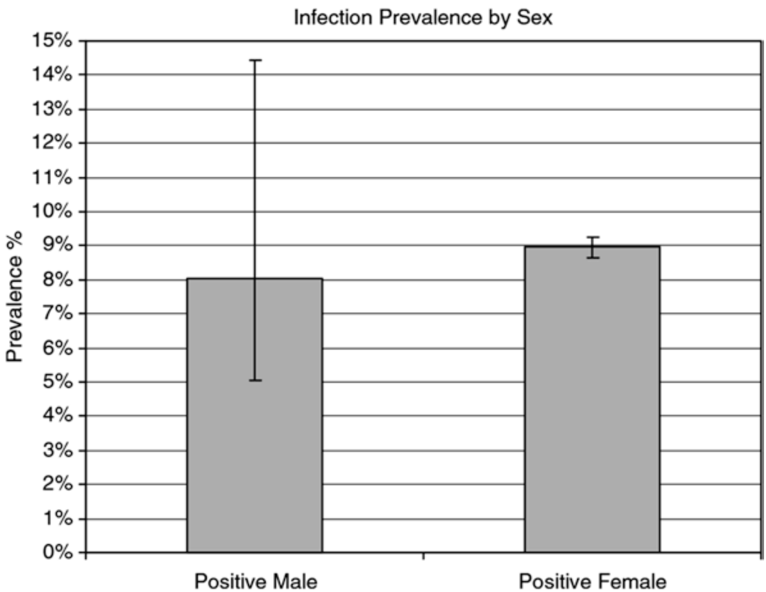


Fig. 3 Prevalence by sex

Figure 4, based on 11,140 patient records partitioned by postcode district, shows the prevalence of infection, with confidence intervals calculated for each district. These confidence intervals varied in width because the partitioned sample size within each district was different. In particular PO17 has a very large confidence interval due to the small sample.

The age profile of patients tested, their test results, and prevalence within each age group were investigated, using 16,411 records for persons aged between 11 and 40 years. Figure 5 shows the prevalence calculated from the above data with confidence intervals for each estimate. A peak in infection prevalence is shown at the age of 18 years.

GUM patients and community referral patients were asked to provide the postcode of the most recent partners. Where these data were available it was possible to assess the pattern of relationships across the postcode structure. The number of relationships which occurred in the same postcode were counted, and also within the same postcode sector and district. Using a VBA program it was possible to search for the number of partnerships in adjacent districts. This largest proportion of partnerships was found to be within the immediate geographic area of the patient, with a significant proportion in the same postcode (which includes the same address). Figure 6 shows partnership location for both community referrals and GUM patients. Partnerships within the same postcode sector, district or adjacent district are proportionally similar for both patient types, and these satisfied a homogeneity test ($P = 0.995$). This

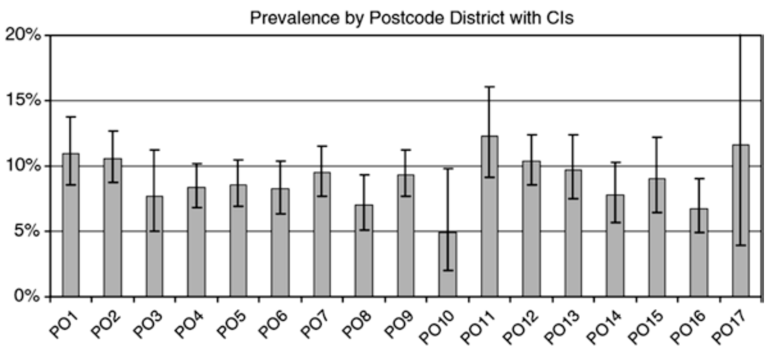


Fig. 4 Opportunistic screen infection prevalence by postcode district

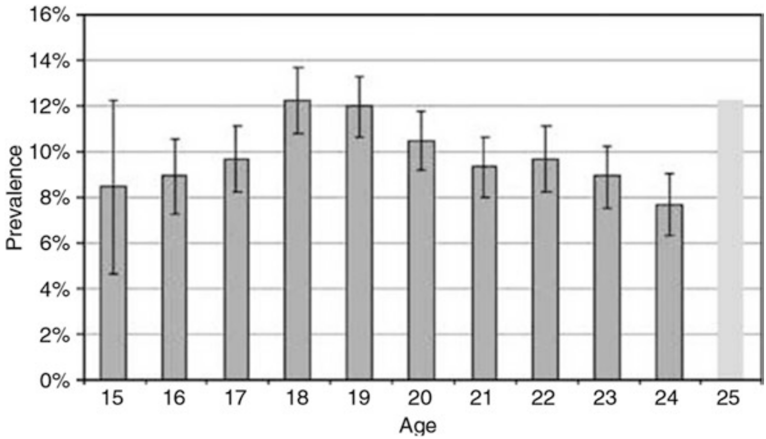


Fig. 5 Age and prevalence distribution

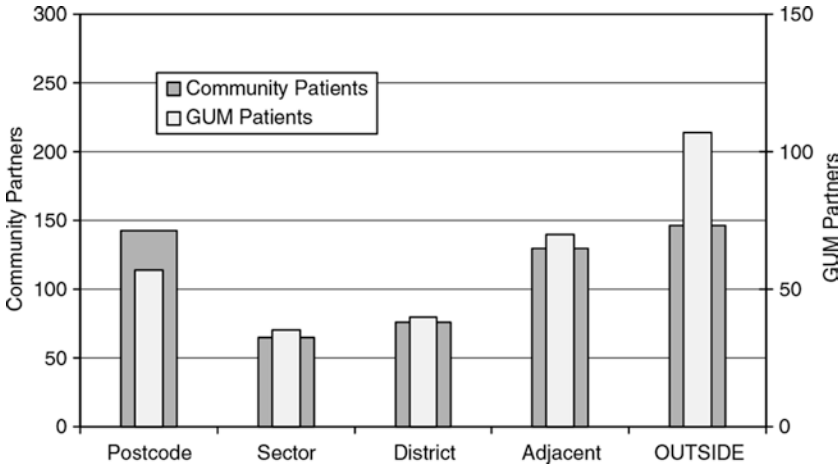


Fig. 6 Partner locations by patient type

suggested that the behaviour of the GUM patients and community patients was similar in the distribution of partnerships in the sectors, districts and adjacent districts. GUM patients have proportionally fewer partners in the same postcode and proportionally more partners outside these regions. An independence test on these two categories shows that this difference is significant (P = 0.002). Therefore, we conclude that GUM patients behave differently compared to the rest of the community, whose partners tended to be closer to home.

The GUM data and community referrals data sets allowed partnership change to be assessed. From the GUM and community data 1872 patients and 987 patients, respectively, indicated the number of partners in the last 3 months, shown in Table 2. One GUM patient was recorded with 12 partners and another with 20 partners.

To summarize key findings from the preliminary data analysis:

- The 17,553 screening trial results constituted a huge sample of data. The conclusions drawn from the analysis can be regarded as representative of the Portsmouth region.
- Screening rates were high. Typically 30–40% of the 16- to 24-year-age group was included. Over the 1-year screening trial a monthly average of nearly of 1500 patients were processed.
- Measured prevalence rates were high. For all Chlamydia tests 10.41% were positive and 12.37% were positive or equivocal (where a positive result could not be confirmed).
- Prevalence among males was not statistically different to that among females. Consequently, male partners should not be ignored in a screening programme of female patients. In addition, infection rate among partners was high in both the GUM and community referrals data sets. Partner tracing and treatment is important in the screening strategy to mitigate the risk of re-infection.
- At the postcode district level (eg PO1) prevalence varies between 4.9 and 12.3%. At the postcode sector level (eg PO1 1) prevalence varies

Table 2 Partnership frequencies (number of partners over a 3-month period)

Number of partners	GUM patients	Community referrals
0	111	49
1	1423	824
2	294	97
3	33	13
4+	12	5

between 3.3 and 18.1%. This analysis forms the basis for the geomapping work (Section "Geomapping Analysis").
- The age profile showed a clear prevalence peak at the age of 18 years.
- GUM patients have been shown to exhibit a different behaviour pattern with a greater proportion of partners outside of their immediate vicinity. It was felt that increased mobility does have significant effects on the risk dynamics, and this may be related to higher levels of disposable income.
- The partnership change frequency was almost identical between GUM patients and community referrals.

Geomapping Analysis

As part of the preliminary data analysis, as described above, results by postcode district were reported and presented in tables and bar chart format to Consultants at Portsmouth. This approach had two main disadvantages. The first was that postcodes were found to be fairly anonymous descriptors unless one happens to be very familiar with the area. The other problem was that there were many tens of postcode sectors. At this greater level of detail it was difficult to produce meaningful, easily interpreted bar charts. This provided the motivation for geomapping.

Geomapping is the art of representing data or measurements superimposed on a map of the area to which it relates. It is a powerful technique since the data values are placed in location context. The maps themselves may be conventional street maps, ordnance survey maps, or other representations, which may even include terrain elevation with 3D perspective views. For the purpose of this research project, it was sufficient to use two approaches; ordnance survey maps, which place the data in an effective and familiar context; and postcode polygon maps, to clearly display the relationships between the data. The MapInfo software tool (www.mapinfo.com) was used to manage, manipulate, and display the mapping data. Plotting prevalence data in MapInfo was, however, was nontrivial, and required various data sources and conversion programs in order to obtain sufficiently detailed plots. Support for this work was provided by the Geodata Institute (within the School of Geography) at the University of Southampton. However, despite the initial efforts, once accomplished geomapping analysis permitted:

- A clear indication of the geographical prevalence and numbers of patients in the 16- to 24-year-age group across the opportunistic screening region.
- Clustering of areas of high numbers of positive test results and high levels of test prevalence.
- A presentation of the infection clusters with their relative magnitudes, allowing for future screening and intervention strategies to be focused in the important infection hotspots and to facilitate intervention planning.
- Analysis and geomapping of the confidence intervals and prevalence uncertainties, providing necessary checks and balances to ensure that any intervention policy is robust.

Numerous maps were produced for the study. Only a small sample is presented here to illustrate the use of geomapping. Spatial prevalence was plotted at both postcode district and sector levels. Figure 7 shows the

Fig. 7 Trial coverage by postcode sector

geographical extent of the opportunistic trial based on the 11,140 patients tested in the Portsmouth area. The postcode sector polygons have been omitted for clarity, although the bars are located at the centres of each postcode sector. The numbers are also omitted but the numbers of tests carried out in each postcode sector is indicted by the relative height of each bar. The largest count (501) occurred in PO4 0 near the Havelock area of Portsmouth, with other large counts in the populated areas around Fareham, Gosport, Portsmouth and Havant.

Figure 8 plots test results at the district level. The left-hand bar (darkest) shows the total number of tests; the middle bar shows the number of positives; the right-hand bar (lightest) shows other results (equivocal, etc). Note that to show these results alongside each other a non-linear (logarithmic) scale was used.

Similar plots were obtained for prevalence, with and without equivocal results, by postcode district. Furthermore, the same maps were produced at the more detailed postcode sector level. For example, Figure 9 shows

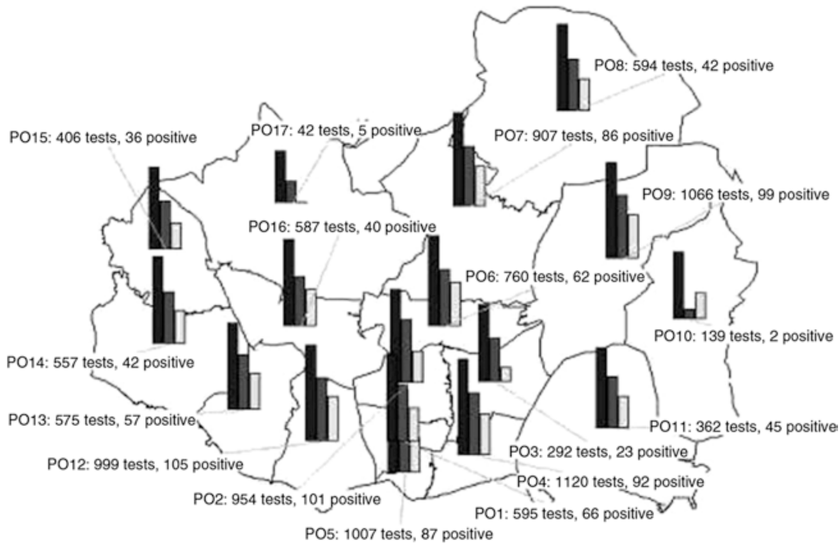


Fig. 8 Total and positive test results by postcode district

Fig. 9 Prevalence by postcode sector

prevalence by sector. Here, prevalence is indicated by the disc diameter (larger the diameter, higher the prevalence). Actual values are not shown on the map for clarity, but relative visual comparison can be made here. The size of the square at the centre of each disk indicates confidence interval (larger the square, larger the CI).

To target intervention strategies, we split the postcode sectors into groups to reflect the distribution of prevalence values, as shown in Fig. 10. The horizontal axis shows increasing levels of prevalence. In conjunction with analysis of the actual values testing positive and equivocal, we decided to group the sectors into four categories of prevalence ranges. This process allowed the top nine sectors to be identified as a distinct group, followed by a second group of seven sectors. Both of these have above average prevalence levels. The remaining sectors were split into two larger groups with middling and low prevalence ranges. Table 3 shows these groupings.

The key tasks and findings of the geomapping work may be summarized as:

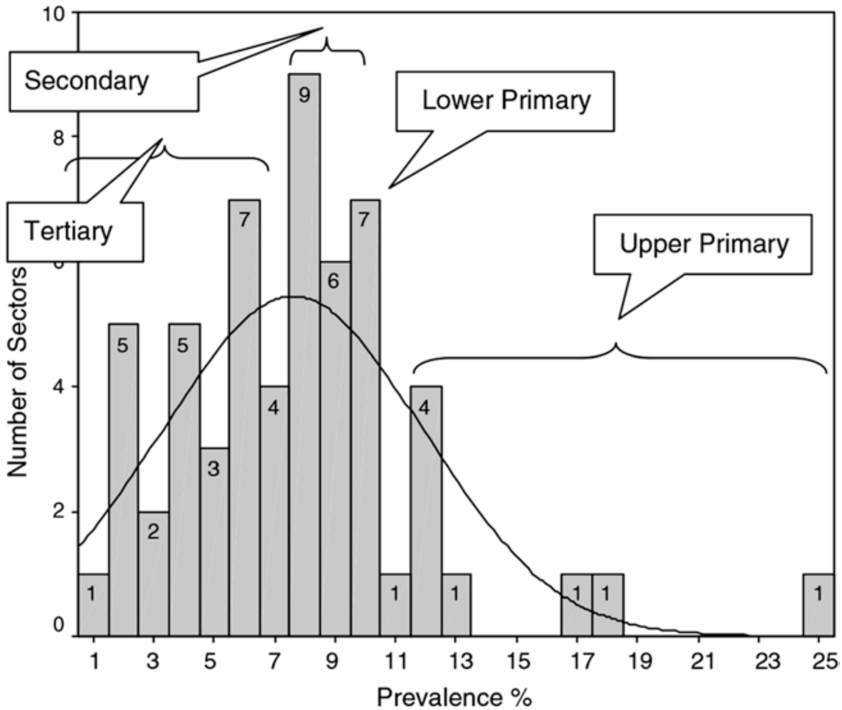


Fig. 10 Distribution of prevalence

- The extent of the screening trial coverage has been plotted provided a clear indication of the numbers of patients in the 16- to 24-year-age group across the opportunistic screening trial region.
- Test result data and prevalence levels have been plotted. This shows the clustering of areas of high numbers of positive test results and high levels of test prevalence.
- Presentation of the infection clusters with their relative magnitudes allows future screening and intervention strategies to be focussed in the important infection 'hotspots'.
- By prioritising prevalence values, four categories of infection prevalence ranges have been identified. These have been plotted as maps to facilitate intervention planning.

Table 3 Allocation of sectors to target groups

Category	Target sector		Category	Target sector	
Upper primary	PO9 6	PO12 1	Lower Primary	PO13 0	PO7 7
Top 9	PO1 3	PO1 4	Second 7	PO6 4	PO6 3
Prevalence	PO17 6	PO1 1	Prevalence	PO12 4	PO9 3
Average 17.3%	PO11 9	PO11 0	Average 13.51%	PO7 8	
SD 4.0%	PO15 6		SD 0.2%		
Secondary	PO2 0	PO13 9	Tertiary	PO7 5	PO15 7
Middle 20	PO5 4	PO4 8	Lower 23	PO10 8	PO9 1
Prevalence	PO9 2	PO3 6	Prevalence	PO4 9	PO6 2
Average 11.62%	PO2 8	PO9 5	Average 7.87%	PO8 8	PO5 2
SD 0.7%	PO12 3	PO13 8	SD 1.7%	PO9 4	PO16 7
	PO2 7	PO5 3		PO4 0	PO17 5
	PO12 2	PO16 0		PO7 6	PO14 2
	PO14 4	PO6 1		PO8 9	PO3 5
	PO16 9	PO14 1		PO1 5	PO8 0
	PO1 2	PO5 1		PO16 8	PO10 7
				PO2 9	PO15 5
				PO14 3	

Socio-economic Indicators and Prediction

Building on the geomapping work and identification of high prevalence sectors, this section describes analysis to ascertain whether of socio-economic variables could act as indicators of infection prevalence. Statistical analysis was carried out using SPSS (www.spss.com) and Minitab (www.minitab. com) and consisted of multivariate regression analysis, and tree-based regression analyses, CART and CHAID (Breiman et al. 1984). Multivariate regression analysis allowed for the primary determinants of prevalence to be identified. Since the collinear variables were removed in the regression process, only the important variables remained that best explained the variation in prevalence. CART and CHAID analyses were then used for clustering of prevalence groups, in order to identify higher risk predictors.

A number of types of socio-economic indicator variable were investigated. The primary source was the Office of National Statistics (ONS). ONS provided census data in a predefined format and content. Other census data were available from Manchester University's Casweb facility via MIMAS (www.mimas.ac.uk). Casweb provided a much higher degree of detail and flexibility, available in large and complex datasets that allowed user definable file format and content.

We used seven indices provided by the ONS to capture aspects of deprivation, plus an overall index. Various factors are used to calculate an index for each electoral ward. In addition, the wards are ordered according to its index to provide a rank position for that ward. Rank position 1 was the most deprived, and 8414 the least deprived. Correlation analysis was conducted on both index and rank. Deprivation indices are determined in the following domains:

- Income deprivation.
- Employment deprivation.
- Health deprivation and disability.
- Education, skills and training deprivation.
- Housing deprivation.
- Geographical access to services deprivation.
- Child poverty (a subset of income deprivation).

Correlation with infection prevalence, as measured at the electoral ward level, gave some very significant correlations. There were potentially important differences of emphasis in the GUM data. This is shown in Table 4, where significant correlations at the 99% level are indicated by double asterisks, and at the 95% level are indicated by single asterisks. Lower *P*-values indicate a greater correlation than higher *P*-values. With one exception, all deprivation indices showed that worse deprivation was associated with higher prevalence. The exception was Geographical Access, where worse deprivation was associated with reduced prevalence, and this was the only socio-economic indicator in both opportunistic screen and GUM datasets to be correlated in this way. Positive GUM patients appeared to be from less deprived wards, since Access and Education deprivation were the only variables to be strongly correlated with GUM patient's prevalence levels.

Other Casweb indictors were also tested for correlation with prevalence. These included deprivation, age, ethnic origin, car ownership, social class, income support and jobseekers allowance, and vital statistics (birth and death

Table 4 Indices of deprivation at ward level

	Significance	
Deprivation index	Trial data P-value	GU data P-value
Multiple rank position	0.000**	0.141
	0.000**	0.023*
Income rank position	0.001**	0.136
	0.000**	0.028*
Employment rank position	0.002**	0.189
	0.001**	0.038*
Health rank position	0.001**	0.017*
	0.001**	0.026*
Education rank position	0.000**	0.002**
	0.000**	0.000**
Housing rank position	0.004**	0.057
	0.001**	0.042*
Access to services rank position	0.000**	0.005**
	0.000**	0.001**
Child poverty rank position	0.001**	0.263
	0.001**	0.111

*Significant at 0.05
**Significant at 0.01

rates). All indicators, except vital statistics, were found to be statistically significantly correlated (at the 95% level) with prevalence. Results were as might be expected, for example higher prevalence was associated with higher income support levels, lower car ownership and lower-skilled professions.

Multiple regression was carried out to reduce the number of previously observed correlated variables to a minimum and useful (practical) subset that best-described infection prevalence. It was found that most of the variables were collinear, that is, they generally exhibited similar behaviour, and thus duplicated the true underlying mechanisms. Regression analysis was used to identify an independent and non-collinear set of variables. Kolmogorov-Smirnov and Shapiro-Wilk's tests were used to make informed judgment on necessary transformations, which included squaring prevalence to satisfy normality assumptions.

Selecting a range of candidate variables based on the correlation analysis gave a useful regression equation to describe or predict prevalence. The key variables found using Minitab's best subsets function were education deprivation rank, ratio of 20–24 years old to ward population, child poverty rank, and number in the ward of age 16–17 years old. The R^2 value was 0.650 which indicated that 65% of the variation was described by these variables. The ANOVA result showed that the regression was statistically significant, since P-value <0.05. The coefficient table showed that all the variables were significant, with the possible exception of the 16–17 years age group. This could have been excluded with a small reduction in R^2 value to 0.621. The regression equation has been plotted in Fig. 11 below, and seems to explain the prevalence variation reasonably well.

CART and CHAID analysis was undertaken using AnswerTree in SPSS. These methods utilize splitting rules and variance reduction techniques. A trade-off is made between misclassification cost and tree complexity to prune the tree structure to its optimal and simplest structure. Only the CART tree is presented here (Fig. 12) as CHAID gave almost identical results. Reassuringly, CART and CHAID suggested the same key variables as identified in the regression analysis. CART analysis found education deprivation as the primary determinant, whereby low deprivation rank (i.e. worse deprivation) provides a group of higher mean prevalence. In the left-hand branch, this was in turn split by the variable describing the population ratio in the 20–24 years age group. Wards with

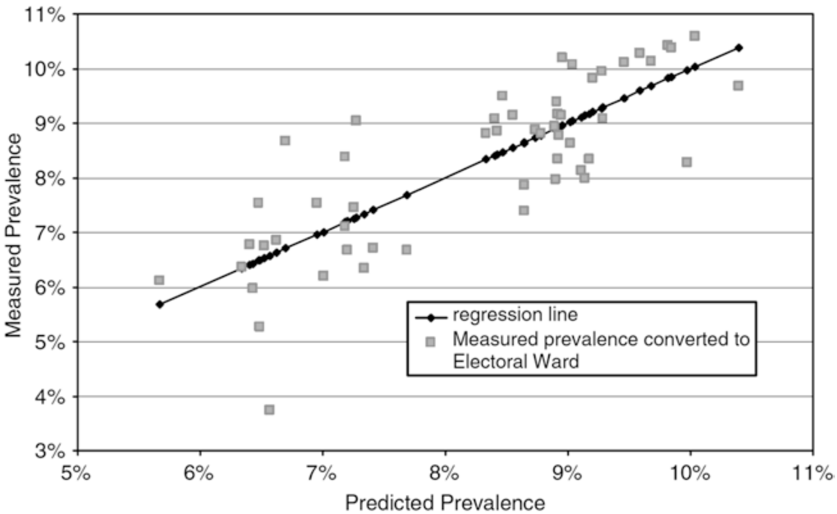


Fig. 11 Opportunistic screen prevalence regression plot

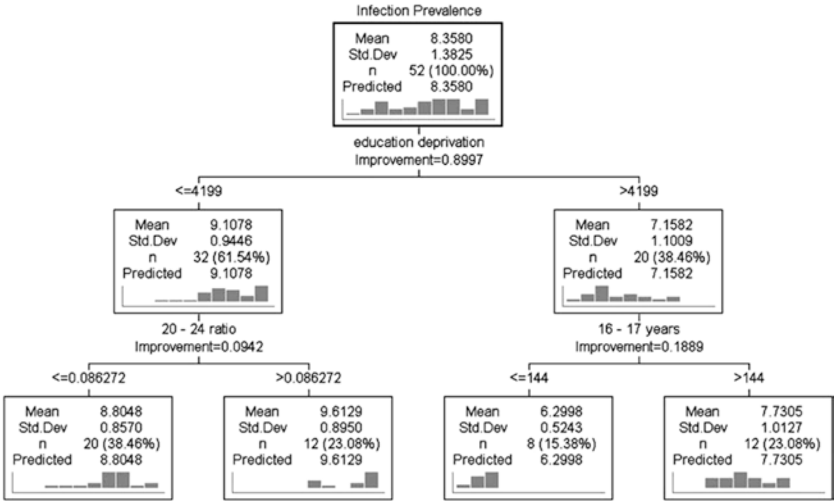


Fig. 12 Opportunistic screen prevalence: results from CART

a proportion of 20–24 years old greater than 8.63% of the total population had the highest prevalence. In the right-hand branch, wards where the number of 16–17 years old was less than 144 have the lowest overall prevalence. Education deprivation became a continual theme in exploratory analyses, with some binary trees branching more than once using this variable. This was perhaps explained by CHAID, which allows multiple splits from each node, and showed opportunistic screen prevalence split into three clear groups, based primarily on education deprivation.

In this section, three main techniques have been employed to gain a multiple-perspective insight into the relationships between socio-economic indicators and infection prevalence. Key findings are summarized as:

- Indices of deprivation were found to be very important, with significant correlations with both screening trial data and GUM data.
- A worse level of education deprivation was associated with higher prevalence. The converse was true for access to services deprivation.
- Other variables were investigated with a view to finding a better set of indicators or surrogates for deprivation, but the regression analysis

confirmed that infection prevalence at the ward level was best explained by deprivation indices and age.

- CART analysis provided a simple sorting rule set, easier to interpret than a regression equation, and provided a graphical explanation for the non-statistician. This confirmed the importance of education deprivation. Education deprivation was also identified as the primary determinant of prevalence using CHAID.

Simulation Modelling of Infection Dynamics and Costs-Effectiveness

To model the dynamics of infection recovery and sequelae, and in order to quantify cost-effectiveness of various screening strategies, a SD model was developed using the Vensim software (www.vensim.com). SD is ideally suited to modelling infections and large population movements. It was particularly relevant in modelling Chlamydia, in that the repeat reinfection mechanism was captured, along with the increased risk of sequelae given repeat infection. These time-dependent effects cannot easily be captured decision analysis models.

Adjustable parameters were included for screening and treatment options, and the model allowed interactions to be assessed and cost-effectiveness to be estimated. More detailed information on the model structure, parameters and results have already been reported in Evenden et al. (2005). In this paper, we present limited results in order to demonstrate how previous components of the research (geomapping and risk grouping) are combined with the SD model.

With the agreement of Consultants at St Mary's Hospital, we developed a simplified, efficient, and user-friendly model with the practical needs of policy makers in mind. Evenden et al. (2005) describes why an SD approach was adopted together with the novel aspects of our model compared to the Townshend and Turner (2000) SD model, and other models in the literature, and is not repeated here. The basis of the chosen model is presented in Fig. 13 and shows the causal loop diagram (CLD) in conjunction with the corresponding simplified stock and flow diagram (SFD).

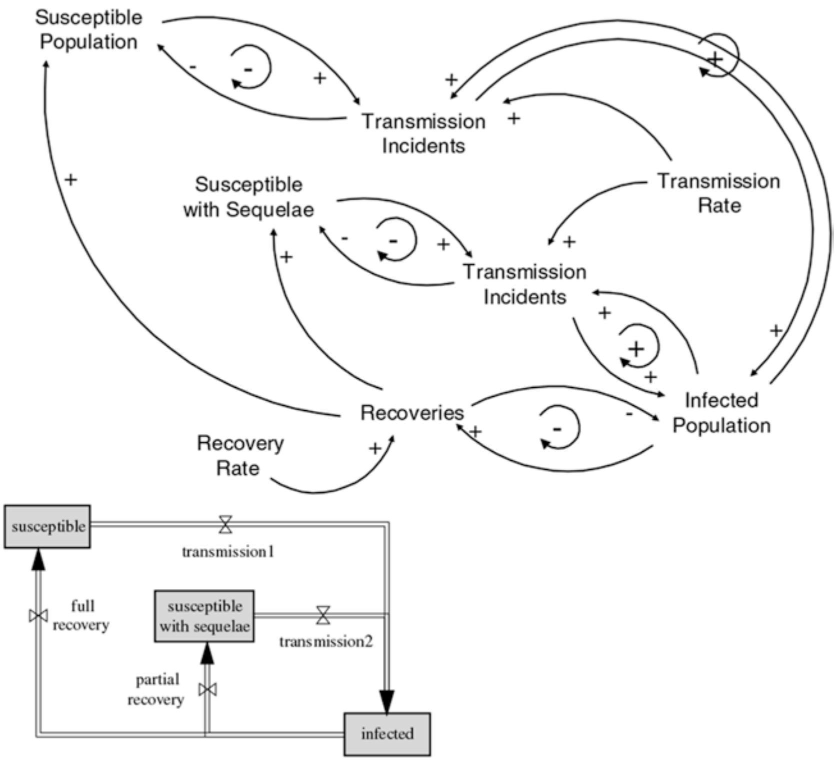


Fig. 13 SD model structure

Two risk groupings were used to capture a higher and lower level of sexual activity and infection. Parameters, such as frequency of partners and size of population, were informed by the geomapping and statistical analyses. Experimentation with the model showed that one of the most important factors was infection from the high-risk infected group into the low-risk susceptible group. The fraction of Chlamydia infections resulting in sequelae was set to 20%. This value was based on a range of sources investigated (Genc and Mardh 1996; Magid et al. 1996; Howell et al. 1998; Townshend and Turner 2000; van Valkengoed et al. 2001; Gift et al. 2002; Skaza and Erzen 2002; Yeh et al. 2003), whence a number of other key parameters including treatment and screening costs were extracted and confirmed by expert opinion from Consultants in GUM at

St Mary's Hospital. A probability tree was built to represent the various sequelae possibilities, including pelvic inflammatory disease (PID), infertility, ectopic pregnancy and chronic pelvic pain. Table 5 shows a selection of the base-case parameters.

In addition to these base-case parameters, infection prevalence was modelled at three levels of infection prevalence (5, 8 and 10%). In order to validate the model, parameters were set to those from other published studies in the literature, such as Howell et al. (1998), and the results compared to those published. Statistical tests indicated that results from our model, for simple base-line scenarios, were not significantly different. Various sensitivity analyses were also performed, for example by adjusting the values of screening rates for both low- and high-risk groups. Sensitivity analysis showed that rate and cost of sequelae, low-risk partnership rate, and mixing rate were the dominant variables. This was as anticipated, suggesting that the model was not exhibiting any unexpected behaviour.

To illustrate cost-effectiveness results, we present Fig. 14, which shows overall costs as a function of the screening rate of the low-risk group. This is shown for six cases of screening of the high-risk group indicated in the

Table 5 Base-case modelling parameters

Parameter	Value
Population	10,000
Percentage of population initially in high-risk group	2.5%
Percentage of population initially in low-risk group	97.5%
Risk group mixing probability (within-group)	90%
Risk group mixing probability (between-group)	10%
High- and low-risk group sequelae fraction	20%
Low-risk group partnership frequency	1 every 5 years
High-risk group partnership frequency 1	every 2 months
High- to low-risk group mixing ratio	10%
Infection mean recovery time	30 months
Modelling period	24 months
Screening cost	£10
Treatment cost	£10
Sequelae cost	£2000

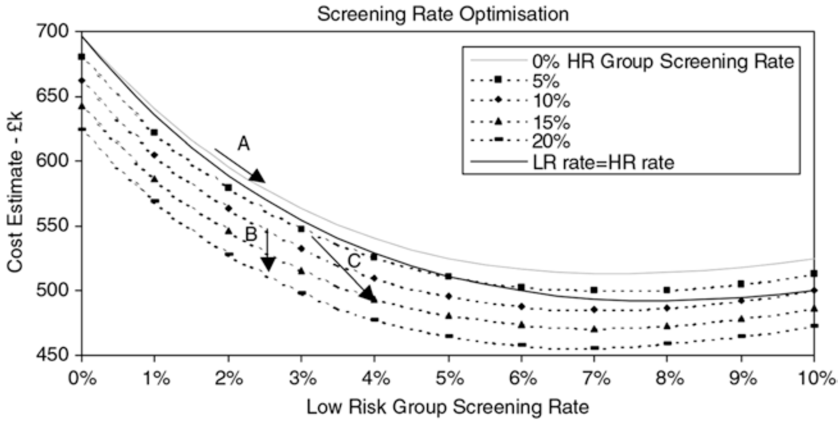


Fig. 14 Cost savings for infection prevalence at 10%

legend. An idealized intervention is described using the labelled arrows. Various other scenarios were also simulated and are discussed in Evenden et al. (2005).

The intersection of the topmost curve with the vertical axis shows the cost of not screening. Arrow A along the solid line shows the direction and overall cost reduction by increasing screening within the general population (i.e. both high- and low-risk groups). Arrow B shows the benefits of further reducing costs by increased targeted screening of the high-risk group, where cost savings can be accrued more rapidly.

Clearly this is an ideal situation since it would not necessarily be possible to target the high-risk group so precisely without also screening more from the low-risk group. For this reason most practical interventions will probably consist of a mix with a predominance of high-risk persons, as indicated by arrow C. Here we make reference to the geomapping work and statistical analysis, which will inform staff at St Mary's Hospital, Portsmouth, on the location of these high-risk groups and their likely socioeconomic characteristics. This combined geomapping and modelling approach is a key benefit of this research over other published studies.

The key findings of the SD work were:

- SD modelling can provide key insights to allow the infection dynamics to be better understood. Key parameters can be varied and the effect on the dependent variables can be observed.
- Since prior infection does not convey immunity, reinfection is one of the main characteristics, which makes its prevalence such a long-term problem. It is a behavioural as well as a medical problem.
- The role of the high-risk groups in the infection dynamic is very important, as it provides a key source of new infection into the low-risk groups. Within the high-risk group itself the prevalence stabilises at a high level, with the majority infected.
- Screening provides immediate cost benefits—costs may be reduced by up to a half at high levels of infection prevalence. At reasonably achievable level, of screening, say 1–2% of the overall population, cost savings are worthwhile.
- To achieve optimal cost savings, a larger proportion of the high-risk groups need to be screened. It will be easier to screen a large proportion of the smaller high-risk groups than it will be to screen a smaller proportion of the larger low risk groups.
- For every high-risk person screened per month, around £1500 can be saved.
- For every low-risk person screened per month, around £200 can be saved.
- These high-risk groups have been identified in geographic location terms and socio-economic terms (Sections "Geomapping Analysis" and "Socio-Economic Indicators and Prediction") to enable this critical strategy to be planned, implemented and to succeed.

Conclusions and Policy Implications

In this study, we have conducted a detailed analysis of 17,553 screening trial results from the UK Department of Health Chlamydia opportunistic trial in Portsmouth. We developed a system dynamics model, with parameter values informed by the analysis of the trial data, which has shown that a high-risk sub-group of the general population, despite being relatively small in size but with a high number of sexual partnerships per case, is critical in the infection dynamics of Chlamydia. Such a group has the largest proportion of sequelae, and provides a major source of infection into the low-risk population sub-group. While the benefits of a universal screening approach have been calculated, it is clear that greater benefits accrue from a higher level of screening of the high-risk group. Thus, blanket screening of the entire at risk population, as proposed in the Department of Health's national screening programme, might be seen simply as to add extra burden to the overstretched health economy, whereas improved targeting of high-risk populations has been shown here to achieve greater cost-effectiveness and to control the recent alarming rise in cases of Chlamydia.

A modelling framework combining computer simulation, geomapping and risk-group clustering techniques, has facilitated a holistic view of the problem. Thus, it has been possible to find the indicators that determine high-risk and high-prevalence within the Portsmouth population, as well as geographically displaying their location across the region by postcode district and sector. Age and indices of deprivation were found to be good predictors, especially education deprivation and proportion of young people within the resident postcode sector. CART, CHAID and multiple regression approaches all gave consistently similar results.

Staff in the GUM department, St Mary's Hospital, are now able to evaluate their screening intervention planning and to re-organize their services in order to target the high-risk groups. For example, they now plan to utilize resources more effectively by visiting schools and public places within the identified primary target postcode sectors in order to increase awareness of Chlamydia infection. It will be possible, using the same modelling framework, to analyse other trial data when this comes available from other geographical regions of the Department of Health screening programme (NCSP). This will enable us to see if the same indicators of high-risk behaviour in Portsmouth are consistent with other regions. Clearly this level of information, coupled with geomapping, will assist regional planning of the screening programme. Furthermore, although we present how the methodology has been used to consider the spatial prevalence and screening interventions for Chlamydia, this approach could be readily applied to other sexually transmitted infections or infectious diseases, such as HIV/AIDS surveillance and intervention planning.

References

Addiss DG et al (1993). Decreased prevalence of *Chlamydia trachomatis* infection associated with a selective screening programme in family planning clinics in Wisconsin. *Sex Transm Dis* **20**: 28–34.

Breiman L et al (1984). *Classification and Regression Trees*. London: Chapman & Hall.

Buhaug H, Skjeldestad FE, Backe B and Dalen A (1989). Cost effectiveness of testing for Chlamydial infections in asymptomatic women. *Med Care* **27**: 833–841.

Department of Health (2000). *Summary Report: A Pilot Study of Opportunistic Screening for Genital Chlamydia trachomatis Infection in England*. London: The Sexual Health and Substance Misuse Team, Department of Health.

Department of Health (2005). *National Chlamydia Screening Programme (NCSP)*. www.dh.gov.uk

Evenden D, Harper PR, Brailsford SC and Harindra V (2005). System dynamics modelling of Chlamydia infection for screening intervention planning and cost benefit estimation. *IMA J Man Math* **16**: 265–279.

Genc M and Mardh PA (1996). A Cost-effectiveness analysis of screening and treatment for *Chlamydia trachomatis* infection in asymptomatic women. *Ann Intern Med* **124**: 1–7.

Gove DJ (1997). *Simulation modelling of infectious diseases*. PhD theses, University of Southampton, Southampton, UK.

Gift TL, Walsh C, Haddix A and Irwin KL (2002). A cost-effectiveness evaluation of testing and treatment of *Chlamydia trachomatis* infection among asymptomatic women infected with Neisseria gonorrhoeae. *Sex Trans Dis* **29**: 542–551.

The Guardian (2004). *Just say no to avoid sex infections, young told*. 25 November, 2004. http://www.guardian.co.uk/uk_news/story/0,,1358905,00.html

Haddix AC, Hillis SD and Kassler WJ (1995). The cost effectiveness of azithromycin for *Chlamydia trachomatis* infections in women. *Sex Trans Dis* **22**: 274.

Harindra V, Tobin JM and Underhill G (2002). Opportunistic chlamydia screening: should positive patients be screened for co-infections? *Int J STD AIDS* **13**: 821–825.

Hart GJ, Duncan B and Fenton KA (2002). Chlamydia screening and sexual health, Editorial. *Sex Transm Infect* **78**: 396–397.

Health Protection Agency (2004). HIV and other Sexually Transmitted Infections in the United Kingdom in 2003. *Annual Report*, November 2004.

http://www.hpa.org.uk/infections/topics_az/hiv_and_sti/sti-chlamydia/chlamydia.htm

Herrmann B, Johnsson A and Mardh PA (1991). A retrospective study of efforts to diagnose infections by *Chlamydia trachomatis* in a Swedish county. *Sex Transm Dis* **18**: 233–237.

Hicks NR et al (1999). Chlamydia infection in general practice. *BMJ* **318**: 790–792.

Honey E et al (2002). Cost effectiveness of screening for *Chlamydia trachomatis*: a review of published studies. *Sex Transm Infect* **78**: 406–412.

Howell MR, Quinn TC and Gaydos CA (1998). Screening for *Chlamydia trachomatis* in asymptomatic women attending family planning clinics. *Ann Intern Med* **128**: 277–284.

Magid D, Douglas JM and Schwartz JS (1996). Doxycycline compared with azithromycin for treating women with genital *Chlamydia trachomatis* infections: an incremental cost-effectiveness analysis. *Ann Intern Med* **124**: 389–399.

Pimenta JM et al (2003a). Opportunistic screening for genital *Chlamydial infection*. I: Acceptability of urine testing in primary and secondary healthcare settings. *Sex Transm Dis* **79**: 16–21.

Pimenta JM et al (2003b). Opportunistic screening for genital Chlamydial infection. II: Prevalence among healthcare attenders, outcome, and evaluation of positive cases. *Sex Transm Dis* **79**: 22–27.

Skaza A and Erzen I (2002). Cost-effectiveness of screening for *Chlamydia trachomatis* in adolescent females in Slovenia. *Dermatovenerologica* **11**. www.mf.uni-lj.si/acta-apa/

Townshend JRP and Turner HS (2000). Analysing the effectiveness of chlamydia screening. *J Oper Res Soc* **51**: 812–824.

van Valkengoed IGM et al (2001). Cost effectiveness analysis of a population based screening program for asymptomatic *Chlamydia trachomatis* infections in women by means of home obtained urine specimens. *Sex Transm Inf* **77**: 276–282.

Yeh JM, Hook EW and Goldie SJ (2003). A refined estimate of the average lifetime cost of pelvic inflammatory disease. *Sex Transm Dis* **30**: 369–378.

Competitive Dynamics in Pharmaceutical Markets: A Case Study in the Chronic Cardiac Disease Market

M. Kunc and R. Kazakov

Introduction

Bulgaria has implemented the European Union pharmaceutical legislation regarding drug regulation and control in several phases prior to its EU accession in 2007. The political idea of a single and competitive EU market providing timely access to market for innovative and generic medicines will continue to fuel healthcare and pricing policies' harmonisation across Europe. Nevertheless, the area of pharmaceutical pricing and reimbursement policy remains still fragmented and incoherent, subjected primarily to national budgetary constraints and institutional inconsistence.

Journal of the Operational Research Society (2013) **64**(12), 1790–1799.
https://doi.org/10.1057/jors.2012.150
Published online 30 January 2013.

M. Kunc (✉)
Warwick Business School, University of Warwick, Coventry, UK

R. Kazakov
BG PharmA, Sofia, Bulgaria

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1_15

An important consideration for Bulgarian and CEE pharmaceutical markets is that they are traditionally branded generic in nature, in contrast to the west and north EU countries, for example United Kingdom, where International Non-proprietary Name (INN) generics dominate. Branded generic markets are those in which generic medicines (bio-equivalent versions of the off-patent original drug) have trade names which, apart from being high-valued intellectual property assets, serve for building customer (doctors, patients, pharmacists) loyalty and market differentiation.

The objective of the system dynamics modelling study presented in this paper is to analyse the impact of different aspects of a drug regulation policy like providing timely access to market, influencing prescribing of generic medicines and implementing programmes for increasing the percentage of diagnosed patients on the dynamics of the pharmaceutical market of one chronic disease.

Reforming Healthcare Policies—An Evolutionary Process

The price of a medicine is one of the major market attributes of pharmaceutical products, and as such it is under a tight administrative scrutiny and control through local regulation. Price regulation itself is viewed as a necessary measure, but only if it relates to medicines that are reimbursed by the state and healthcare funds (EC Pharmaceutical Forum Report 2008). However, the administrative processes associated with price and reimbursement status approval appear to have a far significant effect on the market access to newly authorised medicines, and hence on the patients' access to alternative new generic and cost-effective therapies (Kazakov 2007a; Stiglitz and Jayadev 2010).

Since the fall of the Berlin wall in 1989 and the start of the democratisation of the CEE countries, the healthcare system in Bulgaria has been under continuous changes, proving that the public policy reformulation is more an evolutionary than a revolutionary process. With a new amendment of the national drug legislation from the summer of 2011, it took about 22 years for the healthcare regulation to take into account the timing of the drug access of newly authorised generic medicines in relation to cost savings and patients' compliance effects it can have for the healthcare system and the society.

Timing of pricing and reimbursement has an impact on cost and is critical for timely access to market and in enhancing competition, and provides significant value to patient groups and public funds. However, most of the CEE countries overlooked the timing factor by sticking to cost containment measures like therapeutic class reference pricing and cross-border reference pricing, constrained reimbursement lists, price cuts, co-payment increase, and reduction of margins of whole traders and retailers. Broadly speaking, market inefficiencies should be addressed by the policymakers in the most efficient way, so as to ensure that patients have timely access to innovative and cost-effective treatment therapies. Markets for healthcare and insurance services are marked by 'significant degrees of asymmetric information and agency relationships', which lead to the adverse selection phenomenon of paying less for more (Arrow 1963; Stiglitz 2003; Folland et al. 2004; Kazakov 2007b). This situation results in dynamically complex problems where the long-term impact of policies is difficult to visualise, requiring simulation approaches.

Using System Dynamics Modelling in Health Care

While many modelling approaches can be employed in health care, System Dynamics has been applied predominantly within the frames of traditionally well-developed healthcare systems and public healthcare administrations (Katsaliaki and Mustafee 2011). Some of the themes addressed were disease management in diabetics (Homer et al. 2004a, b; Jones et al. 2006), HIV/AIDS (Roberts and Dangerfield 1990; Dangerfield et al. 2001), and the impact of smoking (Roberts et al. 1982; Lane et al. 2000; Tengs et al. 2001). Other issues modelled were related to modelling patient flows (Wolstenholme 1996, 1999; Hirsch 2004; McDonnell et al. 2004), performance assessment (Hoard et al. 2005) and public health emergencies (Hirsch and Immediato 1998, 1999), and high-level policymaking in human resources planning in health care

(Birch et al. 2007; Murphy et al. 2009). Some scholars have concentrated on public health planning and healthcare reform (Homer and Milstein 2004; Hirsch et al. 2005; Homer and Hirsch 2006; Milstein 2008).

In the area of cardiac diseases, Cooper et al. (2008) developed a discrete simulation model to quantify the costs, benefits and cost-effectiveness of increasing the levels of secondary prevention drug usage for the prevention of heart disease.

Discovering and implementing key policy drivers towards achieving a public policy goal is critical for the advancement of the healthcare reform everywhere (Birch et al. 2007; Murphy et al. 2009). It is especially critical in developing countries where budgetary constraints are severe like Bulgaria and other CEE countries, to mention but a few, and access is a key policy for the government (Mcintyre et al. 2009). For this reason, the modelling project presented here is concentrated on the analysis of one driver—timing of access to market for new generic medicines within the pharmaceutical market for chronic cardiac diseases.

The modelling process followed the general system dynamics modelling framework (Sterman 2000; Kunc and Morecroft 2007; Morecroft 2007) of five stages: Problem articulation; Dynamic hypothesis; Formulation; Testing and Policy formulation and evaluation.

Problem Articulation and Dynamic Hypothesis

The modelling project was conducted focusing on one widely employed drug molecule therapy with the following Non-proprietary Names (INN): bisoprolol. Initially, there were only patented branded products on the market. After their patent expired, they started to face gradual competition with the introduction of generic versions to the originator reference products. The first branded generic version of bisoprolol was introduced in 2005, the second in 2007, then the third and the fourth in 2008 and 2009, respectively. In 2010, there was a total of 5 branded generic drugs. Figure 1 illustrates the volume dynamics of original brand and generics shares in the local market for the drug.

The system dynamics model was designed to capture the initial monopolistic situation on the market and the introduction of the generic drug versions at certain points of time. The idea was to map the effects of

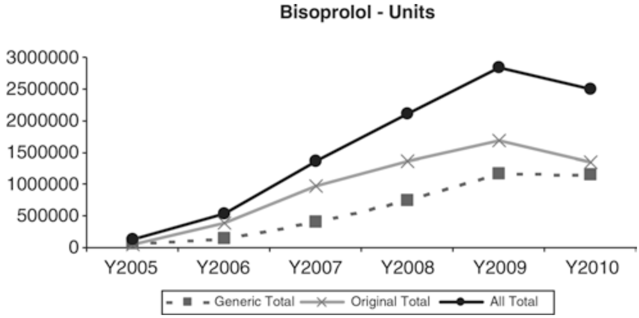


Fig. 1 Volume of bisoprolol in Bulgaria. Source: IMS Health Report (2010)

the two situations (only original drugs compared to original and generic drugs) on healthcare costs for the National Health Insurance Fund (NHIF) in Bulgaria and on the patient population access to treatment, including non-compliant and returning patients. The objective of the model was not to replicate the exact pattern followed by the market, since the model was employed to generate insights (Morecroft 2007) about the drivers of the market.

The initial working hypothesis was focused on testing the effect of timing of generic entry on the market. Timing of generic entry is viewed as one of the critical drivers that can help governments improve market competition, which consequently leads to rapid cost containment process for the public healthcare funds, as generic competition is documented to drive down prices from 20 to 80% in all therapeutic areas (EGA Report 2009; Generics Bulletin 2011). Additional effects, as informed by different stakeholders, are believed to be related to increasing the number of treated patients, as therapies become less expensive and more affordable, and decreasing future hospital care costs, associated with decreasing the number of non-compliant patients. Non-compliant patients can be viewed as a future problem to the healthcare system as they increase future healthcare costs due to hospitalisation. Non-compliance in developing countries is mainly associated with high treatment costs. Figure 2 shows a causal loop diagram[1] of the initial working hypothesis.

A market with only an original drug is mainly driven by a reinforcing feedback loop (R1). Higher numbers of nonpersistent patients, due to lack of affordability based on high prices, imply future hospital costs as they

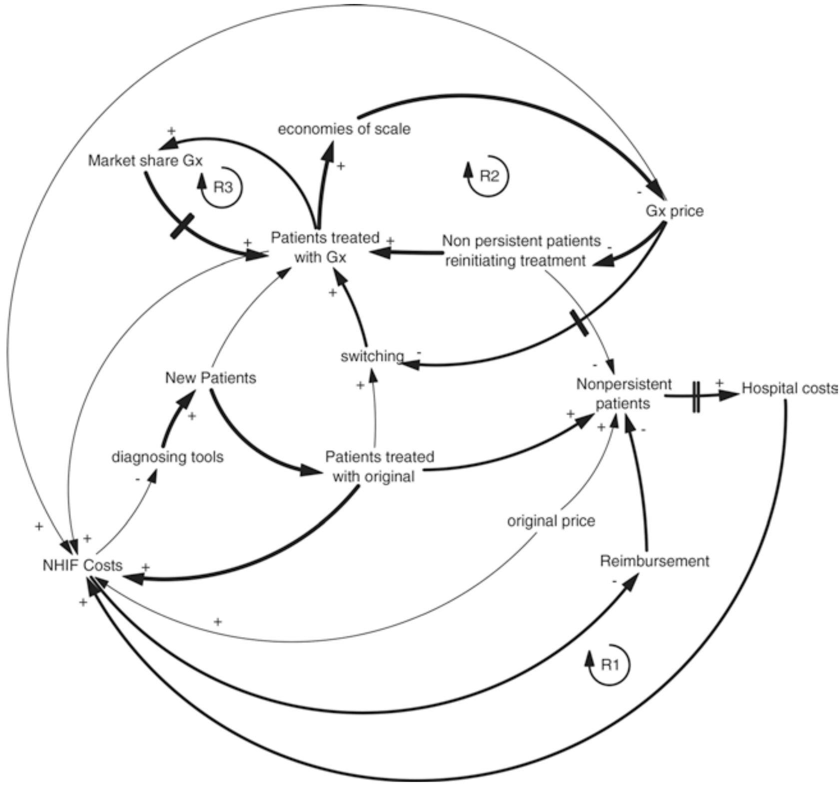


Fig. 2 Initial working hypothesis

may need to be treated for more serious health problems, which increase the total healthcare costs of the NHIF. Higher hospital costs for the NHIF will imply in the future lower reimbursement capacity and even more non-persistent patients reinforcing the growth of hospital costs due to the lack of prevention. When a market includes generic competition, it also generates a reinforcing process (R2). More patients treated with generic drugs will generate economies of scale in the generic industry, reducing their prices even more, since their competitive strategy is based on low prices. Lower prices reduce the number of nonpersistent patients, thus reducing hospital costs and NHIF costs. Finally, a reinforcing feedback process also drives a developed generic market (R3). More patients increase the market share of the generics drugs, which leads to an increase in doctors prescribing

Table 1 Qualitative analysis of the main relationships in the feedback processes

Key relationships existing in the reinforcing processes	Characteristics (Strength and time delay)	Comment
R1. Non-persistent patients → Hospital costs → NHIF costs → Reimbursement	Medium strength feedback process and long term	Original drugs are very expensive leading to patients not to persist with their treatment (especially if the drug is not fully reimbursed), which affects their health in the long term and potentially takes them to hospital for more serious treatment (and augmenting hospital costs). An increase in hospital costs increases NHIF costs further eroding its ability to reimburse expensive drugs
R2. Patients treated with Gx → Economies of scale → Gx price → non-persistent patients reinitiating treatment	Strong feedback process and very fast	Generic drugs' costs are mainly manufacturing costs so higher volume implies lower costs, which are usually translated into lower prices to increase market share
R3. Patients treated with Gx → Market share Gx	Medium strength feedback process and medium term	Increasing adoption of Generic drugs (Gx) by patients leads to higher visibility and higher market share of Gx but it takes time to change doctors, pharmacies and patients' behaviour

generics drugs due to their acceptance. An increase in market share also allows generic firms to invest more in drug development improving treatment attractiveness thus attracting even more patients. Table 1 provides a qualitative analysis of the reinforcing feedback processes shown in Fig. 2.

After analysing the feedback loops in the industry and to reduce the complexity of the model, it was decided to include in the model only drivers like the percentage of reimbursement, price reduction per year required to drugs and prescription support through incentives to doctors in order to capture the effects of the main feedback processes. Changing the percentage of reimbursement (the fraction of the market price that the public funds pay for a prescribed drug with the lowest price in a therapeutic group) is related to current practice in the country, and is based on decisions by a pricing and reimbursement committee (Commission on

Pricing and Reimbursement—Ministry of Health—MoH) and by the introduction of new generic versions that bring reference price levels down. In addition, when follow-up generic drugs come to market, the reductions in price are generated due to the dominant competition rationale in generic drug manufacturers. On the other hand, generics prescription support through incentives to doctors is not a developed practice in Bulgaria, though it is widely accepted in other EU countries. Modelling the effect of prescription support was viewed as important to start the reinforcing feedback processes driving the generic medicines.

After defining the scope of the model, we analysed market data from IMS Health, a well-known market research firm in the pharmaceutical industry. A clear and immediate insight from the IMS Health data is that cost containment effects for the NHIF increase when generic drug competition intensifies. It is also clear that generic competition affects positively the number of treated patients, as sales volume indicates. However, it is not immediately clear what happens to undiagnosed and non-persistent patients, and how key policy drivers like timing of entry, reference price, generics prescription support and price reductions can affect healthcare public cost containment and the number of patients reaching treatment. These effects and the relevant policy design measures that can positively support them are additionally revealed and highlighted by the system dynamic modelling results presented in the next section.

Formulation and Testing

We modelled the flows of new patients with cardiac afflictions, their accumulation into undiagnosed patients until they are diagnosed, patients currently treated with original drugs and with generic drugs, and non-persistent patients. Our analytical framework followed a needs-based approach (Birch et al. 2007) where the main drivers of the market dynamics are the rate of new patients and the distribution of patients across different conditions in their treatment. The inflow rates accounted for new people with cardiac affliction per year, diagnosed patients taking generic drugs, diagnosed patients taking original drugs, discontinuing treatment with generic or original drugs, and reinitiating treatment on original or generic drugs. See Fig. 3.

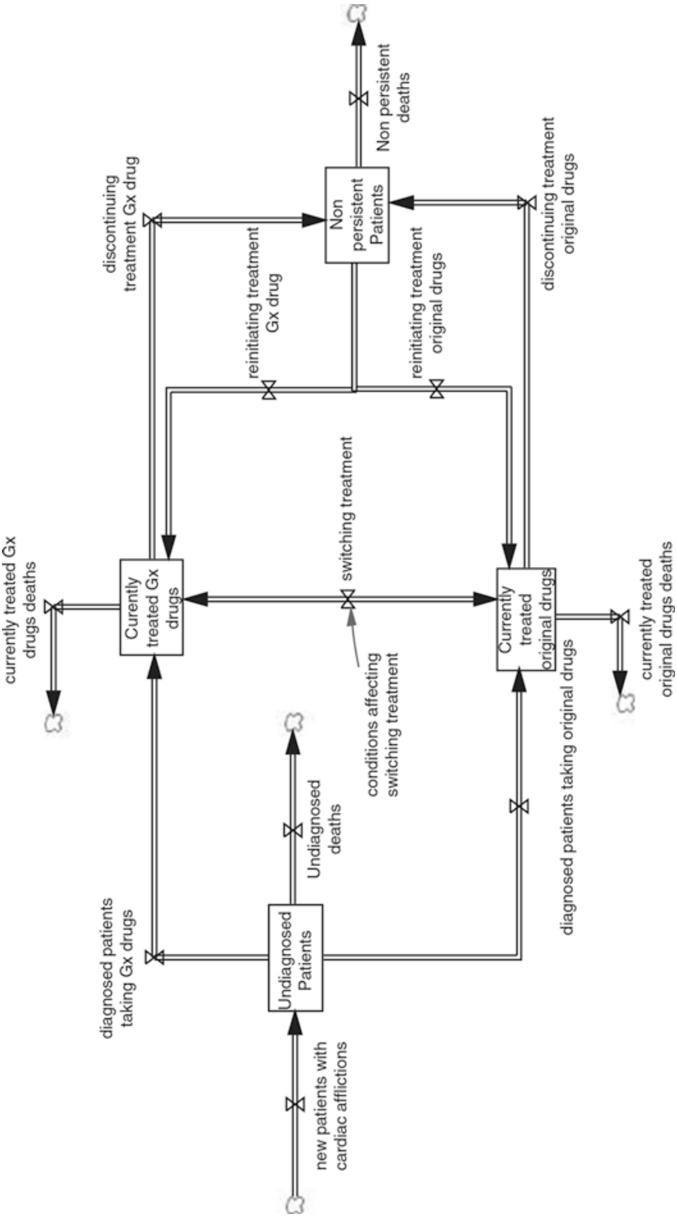


Fig. 3 Stock and flow diagram for chronic disease patients

Any healthy person may become a potential afflicted person given certain lifestyle and other factors generating an inflow of patients into the stock of undiagnosed patients. We used statistics to identify the number of potential afflicted people per year. In chronic diseases, a patient (stock undiagnosed patients) develops a disease physiologically before being diagnosed (ie diagnosed incidence) by a doctor and may die without being diagnosed (flow undiagnosed deaths). In chronic diseases, a diagnosed patient is treated for life, for example a cardiac or a diabetic person (stocks currently treated Gx drugs and currently treated original drugs), until the patient dies (flow currently treated Gx drug deaths and currently treated original drug deaths). The patient may stop the treatment (flows discontinuing treatment original drugs and discontinuing treatment Gx drug) due to certain behavioural factors, such as feeling better, as well as other factors such as lack of adequate treatment and unaffordability. Non-persistent patients may have shorter lives and die or they may return to being treated (flows reinitiating treatment Gx drug and reinitiating treatment original drugs), after a problem that implies hospital treatment or because the doctor is persuasive or the patient finds cheaper options. A currently treated patient may not take all pills prescribed due to their costs or other factors like secondary (adverse) effects, so this patient's true compliance will be lower than the prescription from the doctor, affecting his health.

Table 2 presents the data and assumptions employed for the Stock Flow Model for Chronic Cardiac Problems. The data were based on discussion with medical and industry specialists, and consulting with MoH and the National Health Care Insurance Fund's yearly reports on chronic cardiac disease situation in the country. The results obtained from the model, mainly total sales in monetary units and volume, were validated using IMS Health historical statistics for sales and volume.

Policy Formulation and Evaluation

The model was built to reflect some policy drivers regarded by policymakers and industry experts as key to the market regulation, patients' access to treatment and intra-industry competition, see Table 3.

Table 2 Variables and data employed for the model

Variable in the model	Data	Comment
New patients with cardiac afflictions	66,000 people	Disease reports by MoH, NHIF and National Centre for Public Health and Healthcare Analysis
Undiagnosed patients	0 people	No data available
Death rate from chronic cardiac disease for undiagnosed or non-persistent patients	4% per year	Incidence report, National Centre for Public Health Analysis
Death rate for patients under treatment (Gx and original drugs)	2.2%	Incidence report, National Centre for Public Health Analysis
Average price Gx and original drug	5 and 9.5 BGN (Bulgarian currency)	IMS Health statistics
Detection rate of people that develops a chronic cardiac problem	90%	Incidence report, National Centre for Public Health Analysis, Society of Cardiologists in Bulgaria
Percentage of diagnosed patients taking either Gx or original drugs	Market share	Information provided by industry specialists
Patients discontinuing treatment with Gx drugs	0%	Information provided by industry specialist
Patients discontinuing treatment with original drugs	30% if reimbursement covers 75% of price 50% if reimbursement covers 25% of price	Assumption provided by the clinical specialists based on their practice and patient flow observations Assumption provided by the clinical specialists based on their practice and patient flow observations
Non-persistent patients reinitiating treatment with original drugs.	0%	Assumption provided by the clinical specialists based on their practice and patient flow observations
Non-persistent patients reinitiating treatment with Gx drugs	5–40% as price decreases	Assumption provided by the clinical specialists based on their practice and patient flow observations
Switching treatment	0%	Neither data nor assumptions to use

Table 3 Policy drivers included in the model and used in the scenario simulation

Policy drivers	Value range	Notes
Percentage of reimbursement	From 25 to 100%	Per therapeutic group based on drugs with same International Nonproprietary Name (INN)
Time to enter the market	From 0 to 5 years delay	Time for getting Marketing Authorization (MA) and pricing and reimbursement approval
Price reductions	From 0 to 10%	Price reductions in Gx due to market competition and potential negotiations with NHIF
Prescription support	From 0 to 1 (index)	Incentives to doctors to prescribe generic medicines

Figure 4 presents a base case with only the original drug being on the market. About 60,000 people are being afflicted per year with a diagnosis rate of 90%, and the number of accumulated people under treatment reaches 180,000 in 5 years. However, there are 80,000 people who are non-persistent given the co-payment level (25% of 9.5 monetary units, which is the price of the original drug). This situation may imply future healthcare costs (take the pills *versus* going to hospital due to health problems). Finally, the number of undiagnosed patients, who may need future hospital treatment, is 80,000. Almost 50% (80,000 non-persistent and 80,000 undiagnosed) of the people with the chronic disease are out of treatment. The sensitivity analysis, see Fig. 4, shows that reimbursement policies on original drugs (varying from 0 to 100%) can generate differences of approximately 50,000 patients treated (or 33% of the total number of patients treated).

The results of a second policy are presented in Fig. 5. The second policy shows the presence of original and generic drugs in the market simultaneously. A similar number of people are being afflicted per year with a diagnosis rate of 90%. The number of patients under original drugs reaches 100,000, which indicates an important reduction compared with the previous simulation, but it is not a complete substitution of original drugs. The number of accumulated people under treatment with generic reaches 150,000 in 5 years, given the important market share achieved due to its low price. In this policy, the total number of patients is 250,000,

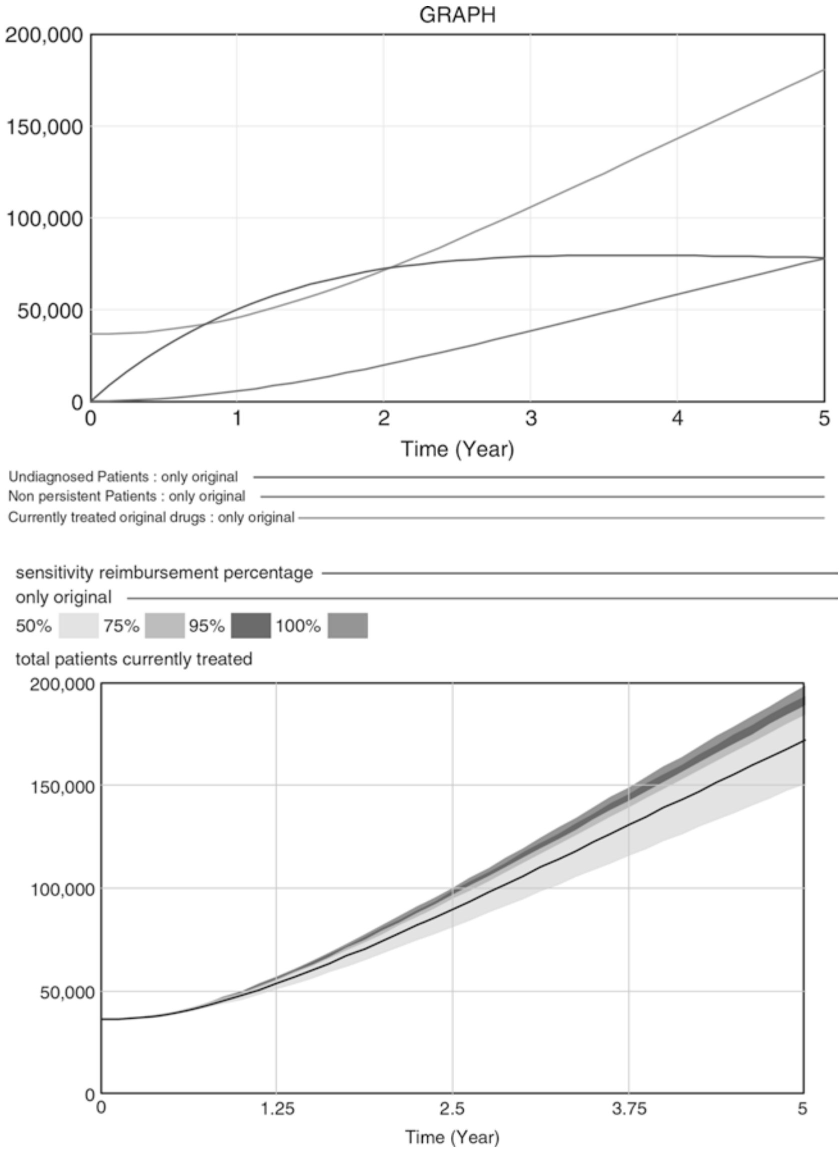


Fig. 4 Patient flows treated only with original drug and its sensitivity analysis

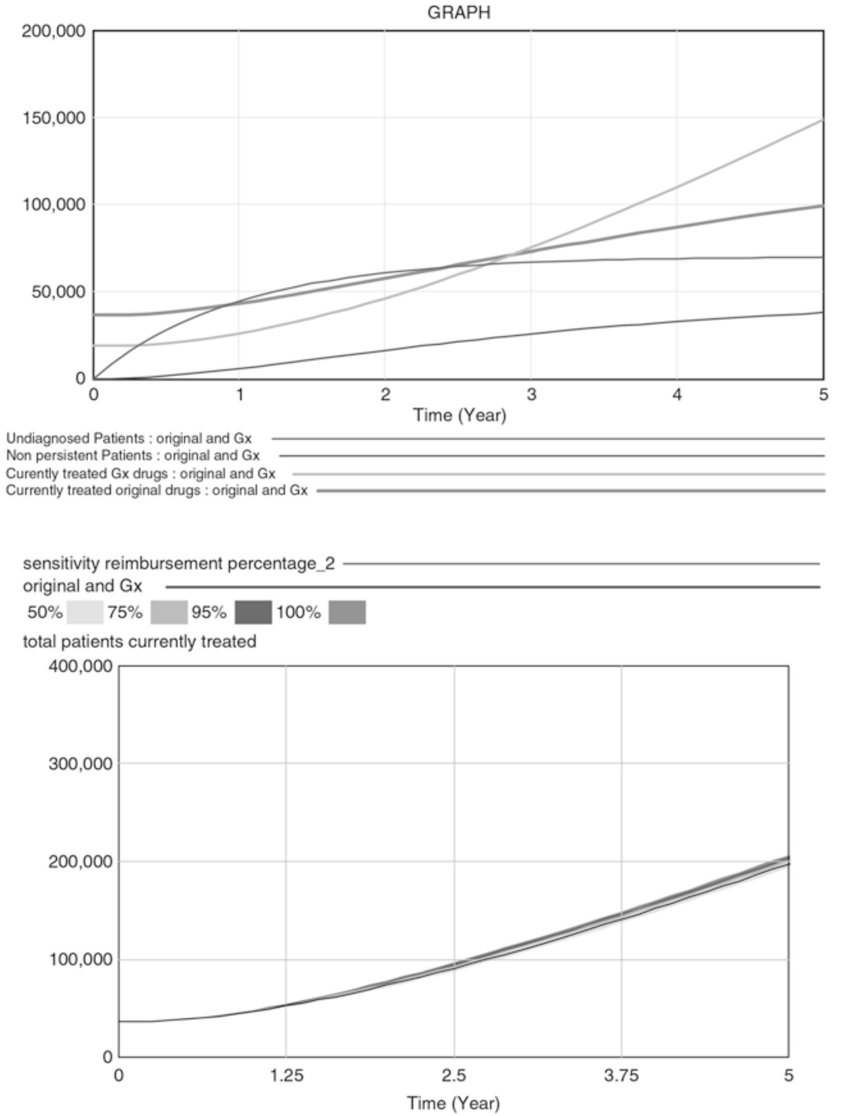


Fig. 5 Patient flows on a market with original and generic drugs (and its sensitivity analysis)

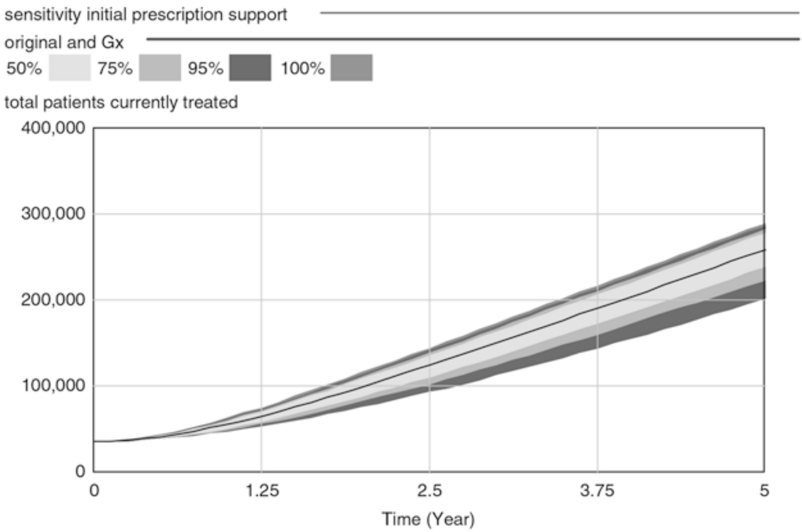


Fig. 5 (continued)

which is 70,000 more than the previous simulation. Now there are only 40,000 people who are non-persistent, so future healthcare costs will be reduced. Finally, the number of undiagnosed patients, who may need future hospital treatment, is 60,000. Now only 30% (40,000 non-persistent and 60,000 non-persistent) of the afflicted people are without treatment. Additional sensitivity analysis performed with the model varying the policy drivers in Table 3, such as reimbursement percentages and initial prescription support, shows that initial prescription support for Gx drugs—bottom graph in Fig. 5—provides the largest impact in terms of patients currently treated (approximately 50% more patients than previous scenarios).

The total costs for NHIF related to drug reimbursement decline substantially given the new co-payment level (25% of 5 monetary units, which is the price of the generic drug). There are more people being treated with a lower level of expenditure for the NHIF, as Fig. 6 shows. Different simulations on time to market entry can be performed to identify the impact of savings for NHIF.

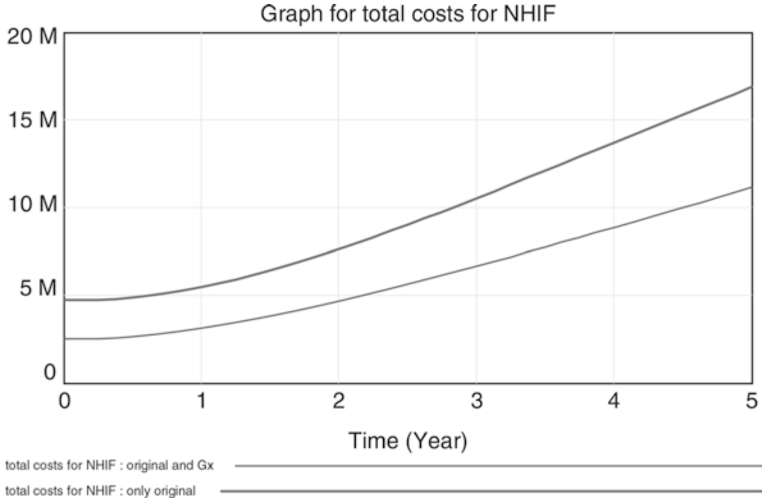


Fig. 6 NHIF costs: Original and generic drugs on the market

Table 4 presents the impact of the sensitivity analysis on the total costs for NHIF and number of total patients treated. The table compares the average, maximum and minimum values at the end of year 5. The combination of different percentage of reimbursement and the presence of Gx is the best policy to achieve lowest costs per patient on average, as well as the minimum cost. However, the maximum coverage is achieved with strong prescription support actions to promote the switch of patients from original to Gx drugs.

Facilitating the Development of Drug Regulation Through Modelling and Experimentation

The two drug cases were presented at a conference workshop with the national healthcare authorities like MoH NHIF and the Parliamentary Commission on Health. Academic groups from the Faculty of Pharmacy at the Sofia Medical University and patients, pharmacist and doctors associations were also present. During the workshop, the participants

Table 4 Comparison between sensitivity results in terms of average, maximum and minimum values

	Average	Max	Min
Percentage of reimbursement only original			
Total patients currently treated	179,810	198,176	155,520
Total costs for NHIF (in Bulgarian currency)	13,671,736	22,592,000	5,318,780
Percentage of reimbursement original and Gx			
Total patients currently treated	199,257	206,668	194,426
Total costs for NHIF (in Bulgarian currency)	7,831,264	12,400,100	3,499,670
Time to enter the market			
Total patients currently treated	194,002	200,825	186,327
Total costs for NHIF (in Bulgarian currency)	9,808,120	15,930,900	8,384,710
Prescription support			
Total patients currently treated	258,166	288,701	200,825
Total costs for NHIF (in Bulgarian currency)	11,617,448	12,991,500	9,037,130
Best policy (Cost per patient treated)	Percentage of reimbursement original and Gx	Prescription support	Percentage of reimbursement original and Gx

Note: See Table 3 for range of values

were involved in working out an optimal regulatory decision set of key driver configurations. The exercise was conducted using a system dynamics interactive learning environment (see Fig. 7). The main goal of the workshop was to help the healthcare authorities to understand the impact of adequate changes to the current pricing and reimbursement drug regulation in the country, while allowing all stakeholder groups to take part in the process.

Working with the Policy Design Simulator provided a virtual learning environment for all the participants. Using the simulator helped all the stakeholders understand the importance of considering healthcare system as a complex system with feedback processes and delays rather than

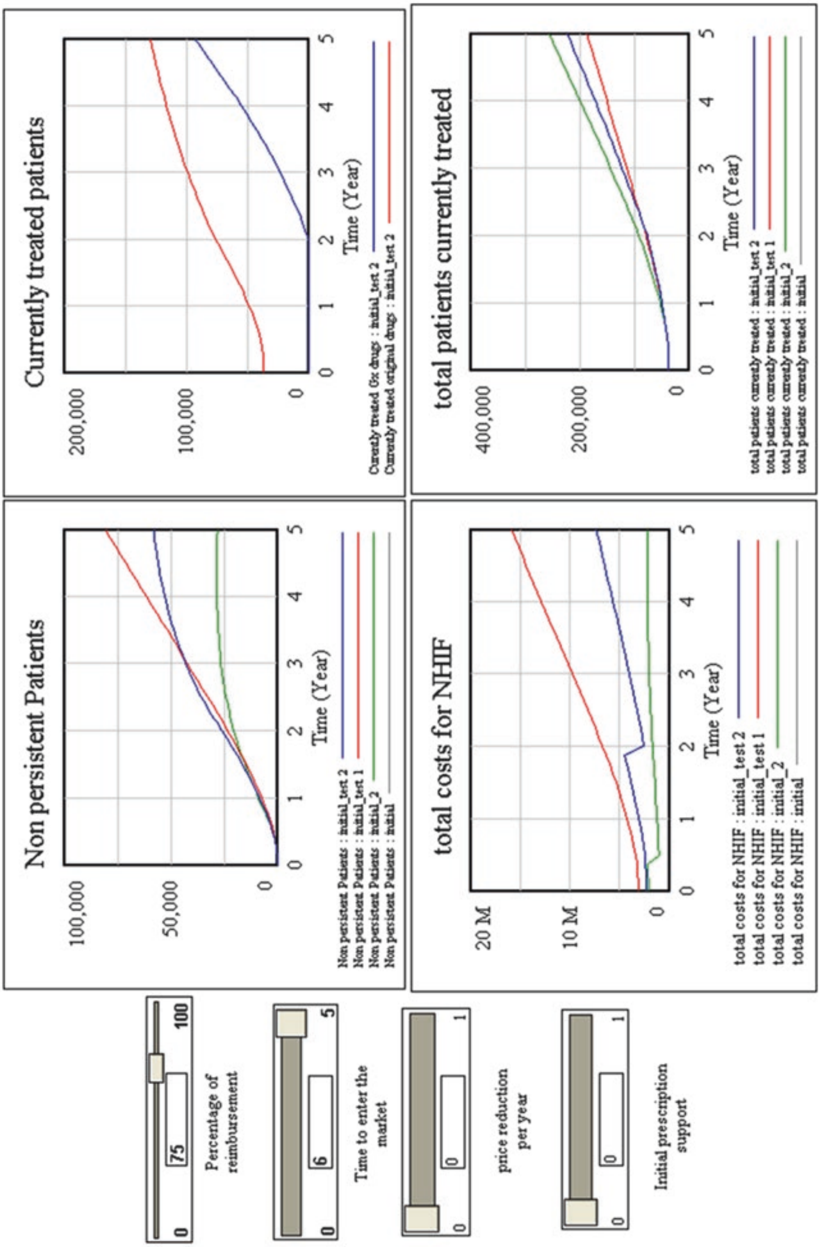


Fig. 7 Interactive learning environment: Policy Design Simulator

independent components with hidden costs like non-persistent patient effect on hospital resources. It was made clear that generic medicines' timely entry coupled with prescription support incentives can be powerful policy drivers towards achieving healthcare policy goals. Furthermore, a drug policy design oriented towards a developed generic market can generate economies of scale and affordable prices for treatment coverage of a larger part of the afflicted population. Participants recognised that the barriers to generic entry like administrative delays and patent protection strategies need to be addressed to ensure effectiveness of the policy measures.

System dynamics proved to be a useful and transparent tool, grounded in market research, which can aid decision making for pricing and reimbursement authorities by exploration of what-if scenarios and plausible paths for improvement in drugs pricing and reimbursement regulation and access to treatment, healthcare coverage and insurance costs.

Conclusions

Our paper contributes to the rich experience of system dynamic modelling employed in healthcare issues. System dynamic models for the treatment of socially significant diseases can help the process of drug pricing and reimbursement regulatory policy redesign for improving total healthcare system performance. Such models can be powerful tools for health and drugs policymakers, including legislative and administrative authorities in developing countries.

The system dynamics study presented shows how important it is to understand and consider the pharmaceutical market as a complex dynamic system. It was made clear that generic medicines' timely entry coupled with prescription support incentives can be powerful policy drivers towards achieving healthcare policy goals. Furthermore, a drug policy design oriented towards a developed generic market can generate economies of scale and affordable prices for treatment coverage of a larger part of the afflicted population.

The study clearly emphasised the importance of the prescription support policy driver, that is, providing incentives to doctors to prescribe generic medicines, which facilitated the development of the generics market and increased the number of treated patients (both newly diagnosed and reinitiating treatment), improving patient access through affordable drugs (McIntyre et al. 2009). A very important third policy driver, coupled with the timely generics access to market and prescription support to doctors, was the percentage of reimbursement, as this supported the decrease of the patients' non-compliance rate and the increase of their treatment reinitiating rate. This brought more patients being treated, that is, widened access to treatment, and provided grounds for long-term savings in the healthcare system due to decreasing future patient hospitalisation rates.

However, we believe that additional research on doctors and patients' behaviour has to be performed through in-depth study in the design of healthcare policies. The difference between the policies shows the importance of treating every situation individually but always using the same framework to learn key lessons from each case in relation to the emerging market behaviour of the system key resources and their optimal configuration for the improvement of the public healthcare systems efficiency and performance.

Acknowledgement We are very grateful for the comments received from our reviewers as they help us to improve our paper substantially.

Notes

1. Causal loop diagrams are used to illustrate feedback systems (Sterman 2000). There are four main components: arrows, polarity, delays and feedback processes. Arrows indicate the direction of causality between two variables. Signs (' + ' or '—') at arrow heads indicate the polarity of relationships between two variables: a ' + ' indicates that an increase (decrease) in a variable causes an increase (decrease) in the related variable, *ceteris paribus.* If the sign is '—', an increase (decrease) in a variable will cause a decrease (increase) in the related variable. However, changes between variables may take some time before they occur. Delays between two variables are represented using a line crossing an arrow, for example the link between

market share and patients treated with Gx in Fig. 2. The nature of the feedback processes is represented using 'loop identifiers', such as R1, which indicate the type of feedback process. 'R' denotes a positive (self-reinforcing) feedback (Sterman 2000).

References

Arrow KJ (1963). Uncertainty and the welfare economics of medical care. *America Economic Review* 53(5): 951–954.

Birch S, Kephart G, Tomblin-Murphy G, O'Brien-Pallas L, Alder R and MacKenzie A (2007). Human resources planning and the production of health: A needs-based analytical framework. *Canadian Public Policy* 33(Supplement 1): S1–S16.

Cooper K, Davies R, Raftery J and Roderick P (2008). Use of a coronary heart disease simulation model to evaluate the costs and effectiveness of drugs for the prevention of heart disease. *Journal of the Operational Research Society* 59(9): 1173–1181.

Dangerfield BC, Fang Y and Roberts C (2001). Model-based scenarios for the epidemiology of HIV/AIDS: The consequences of highly active antiretroviral therapy. *System Dynamics Review* 17(2): 119–150.

EC Pharmaceutical Forum Report. (2008). Working group on pricing and reimbursement. http://ec.europa.eu/pharmaforum/docs/ev_20081002_frep_en.pdf, accessed December 2011.

EGA Report. (2009). How to increase patient access to generic medicines in European healthcare systems. http://egagenerics.com/images/publication/PDF/Market_Barriers_Report_FINAL_ update.pdf

Folland, S., Goodman, A. and Stano, M. (2004). *The Economics of Health and Health Care*. New Jersey: Pearson Education, pp. 187–225.

Generics Bulletin. (2011). Issues 5 August 2011 and 24 January 2011 www.generics-bulletin.com

Hirsch, G. (2004). Modeling the consequences of major incidents for health care systems. *22nd International Conference of the System Dynamics Society*. Oxford, England, 25–29 July.

Hirsch, G., & Immediato, C.S. (1998). Design of simulators to enhance learning: Examples from a health care microworld. *International System Dynamics Conference*. Quebec City, July.

Hirsch, G. and Immediato, C. S. (1999). Microworlds and generic structures as resources for integrating care and improving health. *System Dynamics Review* 15(3): 16–25.

Hirsch, G., Homer, J., McDonnell, G., & Milstein, B. (2005). Achieving health care reform in the United States: Toward a whole-system understanding. *23rd International Conference of the System Dynamics Society*. Boston, USA.

Hoard, M., Homer, J., Manley, W., Furbee, P., Haque, A., & Helmkamp, J. (2005). Systems modeling in support of evidence- based disaster planning for rural areas. *International Journal of Hygiene and Environmental Health* 208(6): 117–125.

Homer J & Hirsch GB (2006). System dynamics modeling for public health: Background and opportunities. *American Journal of Public Health* 96(3): 452–458.

Homer, J., & Milstein, B. (2004). Optimal decision making in a dynamic model of poor community health. *37th Hawaii International Conference on System Science*. Big Island, HI, 5–8 January.

Homer J, Hirsch G, Minniti M and Pierson M (2004a). Models for collaboration: How system dynamics helped a community organize cost-effective care for chronic illness. *System Dynamics Review* 20(3): 199–222.

Homer, J., Jones, A., Seville, D., Essien, J., Milstein, B., & Murphy, D. (2004b). The CDC diabetes system modeling project: Developing a new tool for chronic disease prevention and control. *22nd International Conference of the System Dynamics Society*. Oxford, England, 25–29 July.

IMS Health Report. (2010). Generic Medicines: Essential contributors to the Long-term health of society. Sector sustainability challenges in Europe. http://egagenerics.com/images/publication/PDF/IMS.pdf

Jones AP, Homer JB, Murphy DL, Essien JDK, Milstein B and Seville DA (2006). Understanding diabetes population dynamics through simulation modeling and experimentation. *American Journal of Public Health* 96(3): 488–494.

Katsaliaki K and Mustafee N (2011). Applications of simulation within the healthcare context. *Journal of the Operational Research Society* 62(8): 1431–1451.

Kazakov R (2007a). Pricing and reimbursement policies in new EU accession countries. *Journal of Generic Medicines* 4(4): 249–258.

Kazakov R (2007b). Between strategy and change: Reformulating the medicines industry in an enlarged Europe. *Journal of Medical Marketing* 7(3): 245–253.

Kunc MH and Morecroft JD (2007). System dynamics modeling for strategic development. In F. O'Brien & R. Dyson (Eds.), *Supporting strategy: Frameworks, methods and models*. Chichester, UK: Wiley.

Lane D, Monefeldt C and Rosenhead JV (2000). Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *Journal of the Operational Research Society* 51(5): 518–531.

McDonnell, G., Heffernan, M., & Faulkner, A. (2004). Using system dynamics to analyse health system performance within the WHO framework. *22nd International Conference of the System Dynamics Society*. Oxford, England, 25–29 July.

McIntyre D, Thiede M and Birch S (2009). Access as a policy-relevant concept in low- and middle-income countries. *Health Economics, Policy and Law* 4(2): 179–193.

Milstein, B. (2008). Hygeia's constellation: Navigating health futures in a dynamic and democratic world. The Center for Disease Control and Prevention. Syndemics Prevention Network, Archived version available at http://web.archive.org/web/20100307002849/http://www.cdc.gov/syndemics/monograph/index.htm, accessed 25 January 2013

Morecroft J (2007). *Strategic Modelling and Business Dynamics: A Feedback Systems Approach.* John Wiley and Sons: England.

Murphy, G., et al. (2009). Tested solutions for eliminating Canada's registered nurse shortage. Canadian Nurses Association www2.cna-aiic.ca/CAN, accessed 13 June 2012.

Roberts C and Dangerfield B (1990). Modelling the epidemiological consequences of HIV infection and AIDS: A contribution from operational research. *Journal of the Operational Research Society* 41(4): 273–289.

Roberts EB, Homer J, Kasabian A and Varrel M (1982). A systems view of the smoking problem: Perspective and limitations of the role of science in decision-making. *International Journal of Biomedical Computing* 13(1): 69–86.

Sterman JD (2000). *Business dynamics. Systems thinking and modelling for a complex world.* New York: McGraw-Hill.

Stiglitz G (2003). Information and the change in the paradigm in economics. *American Economist* 47(2): 15–16.

Stiglitz JE and Jayadev A (2010). Medicine for tomorrow: Some alternative proposals to promote socially beneficial research and development in pharmaceuticals. *Journal of Generic Medicines* 7(3): 217–226.

Tengs TO, Osgood ND and Chen LL (2001). The cost-effectiveness of intensive national school-based anti-tobacco education: Results from the tobacco policy model. *Preventive Medicine* 33(6): 558–570.

Wolstenholme, E. F. (1996). A management flight simulator for community care. In S. Cropper & P. Forte (Eds.), *Enhancing Decision Making in the NHS*. Milton Keynes, UK: Open University Press.

Wolstenholme EF (1999). A patient flow perspective of UK health services: Exploring the case for new 'intermediate care' initiatives. *System Dynamics Review* 15(3): 253–271.

Index[1]

[1] Note: Page numbers followed by 'n' refers to notes.

M. Kunc (ed.), *System Dynamics*, OR Essentials,
https://doi.org/10.1057/978-1-349-95257-1

Q

R

S

T