Lancaster University Management School
Working Paper
2007/011

Conceptual modelling: framework, principles, and future research

Roger Brooks

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/publications/
LUMS home page: http://www.lums.lancs.ac.uk/

Conceptual modelling: framework, principles, and future research

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

Abstract

The conceptual modelling task in a simulation project is very important and yet is still generally regarded as more of an art than a science. The meaning and nature of conceptual modelling are discussed and a framework set out. The overall aim should be to choose the best model for the project and conceptual modelling can be viewed as a difficult optimisation problem that can be tackled effectively using a creative search process that develops alternative models and predicts their performance throughout the project. An experiment relating model characteristics to some aspects of performance is described and this type of experiment may inform the process of predicting model performance. Based on advice from the literature and my own previous work on conceptual modelling 17 principles of conceptual modelling are suggested. Conceptual modelling research is still at an early stage and ideas for future research are proposed.

Keywords: Conceptual Model, Project Outcome, Complexity, Simplification

Introduction

Of all the tasks involved in a modelling project, conceptual modelling is probably the one that has received the least attention and consequently is the least well understood. Most other tasks such as data analysis, model building, verification and validation, and output analysis have a strong element of mathematics, statistics or logic. This has enabled techniques from other disciplines to be applied so that there are now well-established methods for most typical situations (see, for example, Law and Kelton (2000) Chapter 10 on output analysis for comparing alternative systems). The nature of conceptual modelling is quite different, so much so that it is often described as being an art rather than a science (e.g., Shannon, 1975). Problem formulation is somewhat similar to conceptual modelling in this respect, although here a range of tools and approaches such as soft systems methodology may be useful (e.g., Checkland, 2006). Most textbooks devote only a few pages to conceptual modelling, with one notable exception being Robinson (2004) which includes two chapters on the topic. This paper examines the meaning and nature of conceptual modelling and suggests a framework for the process. An experiment to investigate possible relationships between model characteristics and performance is described. Some principles of conceptual modelling are suggested based partly on various modelling projects I have carried out over the past 12 years. Approaches for future research are also proposed.

Meaning of Conceptual Modelling

The term conceptual modelling itself seems to cause considerable confusion because of its slightly different uses in different areas of science and also because there is no agreed definition within simulation and operational research (O.R.). There is a considerable amount of technical literature in computer science and artificial intelligence on conceptual modelling that focuses on modelling techniques and ontologies (such as UML) for modelling thoughts, knowledge and concepts. In the context of O.R. projects, the usage is rather broader in that it tends to refer to the whole task (however it is carried out) of deciding what to include in the model. Thus typical project tasks are problem formulation, conceptual modelling, model building (coding), verification and validation, experimentation, analysis of results, and implementation of findings. Figure 1 shows these tasks (the arrows) up to the model building stage along with their intended outcomes (the rectangles). The arrows go in both directions to indicate returning to any previous task. Data obtained from available sources or collected as part of the project may be an input to each task to reach the outcome.

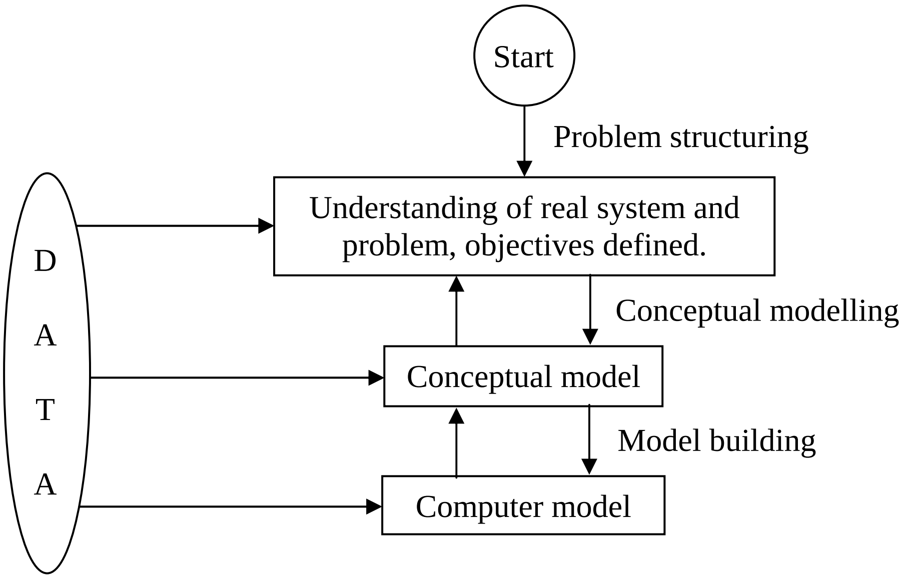

Figure 1. The initial tasks and outcomes in a simulation project.

This paper follows the definition of Brooks and Robinson (2001) that a conceptual model is “a software independent description of the model that is to be constructed”. The conceptual model specifies how the virtual world being simulated should work – the entities that it contains and all the interactions, rules, and equations that determine their behaviour. The specification includes the type of model (e.g., whether the model state changes continuously or discretely), and the scope or boundary of the model.

Assuming the model is intended to represent a real system then a key input into the conceptual modelling process is the modeller’s understanding of how the real system works. This will typically be based on observations, input from experts, and data, but will probably also include various assumptions. The conceptual modelling process will then involve decisions on which modelling simplifications of the real system to make (such as which elements and processes from the real system to include). In principle the conceptual model could therefore also be specified in a top down way by the understanding of the real system including the assumptions made, the model boundary, and the modelling simplifications.

The definition of the conceptual model used here deliberately distinguishes the decisions regarding model content from the technical construction of the model. There are many ways of building the same model by using different software or different code and the conceptual model could be built using any suitable software or programming language. In practice, the choice of conceptual model may be affected by limitations in the available software and the knowledge and skills of the modeller in using the software.

Documentation of the conceptual model can vary considerably from none at all (where the actual model built gives the only evidence) to detailed specifications. Possible formats for recording aspects of the conceptual modelling information include diagrams (such as a process diagram or activity cycle diagram), a list (such as the elements or the assumptions and simplifications), or just a written description. A lack of documentation will tend to make model verification and validation more difficult, and could cause problems if the model is handed over to other users or modellers or if the modelling decisions need to be justified to third parties.

The modelling process is, of course, not generally a simple linear process of completing one task and then moving onto the next. Conceptual modelling may take place in several phases and may happen in parallel with other tasks, particularly data collection and model building. The initial conceptual model may also be reconsidered and amended at various points during the project.

In particular, the ease of use of modern software means that sometimes the model may be coded as it is defined so that conceptual modelling and model building take place along side each other in an iterative manner (i.e., decide on the way part of the model should work, code it, then move onto the next part). Therefore, in some projects the model may be constructed in the simulation software with little or no prior planning and with no separate documentation of the conceptual model. Even so, the decisions taken during this process regarding the content of the model (as opposed to the specific computer code) constitute the conceptual modelling process in such a project.

This paper concentrates on conceptual modelling in simulation, and the flexible nature of simulation means that there can be a very wide choice of possible models. This makes conceptual modelling a particularly difficult task in simulation. Nevertheless, many of the issues considered here will be relevant to other types of models.

Advice on conceptual modelling

Advice given in the literature on the process of deciding what to include in the model tends to be fairly general and brief. For example, even though Law (1991) states “I now strongly feel that the most difficult aspect of a study is that of determining the appropriate level of model detail”, Law and Kelton (2000) devote only two pages out of over 700 pages in their textbook to this issue and their advice is to consider the following: the study objectives, the outputs required, the advice of those who know about the real system, the model credibility required, the data available and the budget for the study. They also suggest using sensitivity analysis (although, in fact, this can only be done once the initial model is built), and they warn against the temptation of including too much detail.

Pidd (1998) considers the modelling method and the experimental frame of Zeigler (1976) as factors affecting the conceptual model. The modelling method (e.g., discrete event or system dynamics) will certainly affect the elements and relationships in the model, but I would consider the choice of method as one aspect of the conceptual modelling decisions themselves in that it specifies the nature of the rules in the model. The experimental frame is the circumstances to be investigated using the model and so is similar to advice to consider the objectives and required outputs. In Pidd (2003), six principles of modelling in management science are set out, several of which focus on different aspects of keeping the model simple.

Robinson (2004) gives much more emphasis to conceptual modelling, devoting two chapters to it. He provides a very good discussion of many of the issues including a definition, concepts for evaluating the conceptual model, documentation and model simplification. A framework for developing a conceptual model is also presented consisting of a process with stages of understanding the problem, deciding on the objectives of the project, specifying the inputs and outputs of the model, and finally determining the model content. This is broadly similar to relating the conceptual model to the objectives and experimental frame. However, presenting this as a step by step logical process and providing an in-depth discussion with examples makes this a much more useful practical guide than previous literature.

A very common piece of modelling advice is to keep the model as simple as possible, even relating this back to Ockham's (or Occam's) razor from the 14th Century that can be translated from the Latin as "entities should not be multiplied without necessity". The other term often used in the literature is level of detail, with complexity and level of detail appearing to be synonymous in most cases. Various authors have set out the advantages of a simple model compared to a complex model including being easier to understand, easier to change and update, quicker to run and analyse, requiring less data and fewer resources, and having a greater chance of acceptance by the client (e.g. Ward, 1989; Salt, 1993). A simple model is identified as a key factor in O.R. project success in the survey by Tilanus (1985).

Nevertheless, although we may have an intuitive idea of the level of detail and complexity of a model (e.g., we may be confident in a comparative ranking of the complexity of certain models) there appears to be no precise definition of model complexity or agreed way of measuring it. Indeed it is quite a broad concept and narrower attributes may be more useful. For example, Brooks and Tobias (1996) suggested size, connectedness and calculational complexity as sub-components of complexity although, even then, there are various ways of measuring these.

More detailed reviews of relevant literature for conceptual modelling are given in Brooks and Tobias (1996) and Robinson (2006).

Conceptual modelling framework

In examining the nature of conceptual modelling, a good starting point is to consider what the objective of this task should be. Conceptual modelling specifies the content of the model to be used in the project and so it has a big impact on all the other modelling tasks and on the success of the project as a whole. The conceptual modelling objective should therefore be to choose the model that will result in the most successful project. This involves considering how the model will affect the different aspects of model performance, in particular the following eleven elements specified in Brooks and Tobias (1996):

<u>Results</u>

1. The extent to which the model output describes the behaviour of interest (whether it has adequate scope and detail).
2. The accuracy of the model's results.
3. The ease with which the model and its results can be understood.

<u>Future use of the model</u>

4. The portability of the model and the ease with which parts of the model can be reused in future models.

<u>Confidence in the model (verification, validation and credibility)</u>

5. The probability of the model containing errors.
6. The accuracy with which the model output fits the historical data.
7. The strength of the theoretical basis of the model including the quality of input data.

<u>Resources required</u>

8. The time and cost to build the model (including data collection, verification and validation).
9. The time and cost to run the model.
10. The time and cost to analyse the results of the model.
11. The hardware requirements of running the model.

Although the aim should be to find the best model, as a minimum, conceptual modelling needs to find a model that is predicted to lead to a successful project outcome. Minimum performance standards to achieve this could be identified for each of the above elements. If all the conceptual models identified are likely to result in failure then the project as a whole needs to be re-assessed - see Robinson and Pidd (1998) for a discussion of simulation project success based on interviews with providers and customers, and Tilanus (1985) for survey data of the key factors causing success or failure on O.R. projects.

Therefore conceptual modelling is essentially an optimising (or at least satisficing) task that could be formulated in principle as:

Objective function:	Maximise overall predicted project performance, P: $P = f\left(k_1(m,R,u_1),...,k_n(m,R,u_n)\right)$ where the $k_i(m,R,u_i)$ for $i = 1, 2,\ldots, n$ are measures of each of the n aspects of project performance.
Subject to:	The resources allocated to the different project tasks are less than or equal to the maximum available resources, A: $r_c + \sum_{i=1}^{l} r_i \leq A$ where r_c is the resource allocated to conceptual modelling and r_i for $i = 1, 2,\ldots, l$ are the resources allocated to the other tasks.
Decision variables:	Conceptual model chosen, m; allocation of resources to the different tasks, $R = (r_c, r_1,..., r_l)$
Uncontrollable variables:	Other factors affecting each performance aspect, u_i

There are several interactions and uncertainties that make this a very difficult problem. The overall performance is a function *f* of the measures of each aspect of model performance (such as the eleven aspects listed above), which in principle could be a weighted sum of the measures. The relative importance (and hence the weights) of these performance aspects will differ substantially between different projects. For example, improved understanding could be the main aim of a research project rather than making accurate predictions. A short timescale, strict deadline and small budget would mean that the resources required are particularly important. The measures of predicted performance of the different aspects will, to an extent, be subjective (as discussed in Brooks and Tobias, 1996). In fact, even after the project has been completed an objective quantitative measure of some of the aspects is probably not possible. Therefore, in practice an overall performance assessment may just consist of considering all the different aspects bearing in mind their relative importance for the project, rather than combining them in any formal function.

The performance measure or assessment for each aspect depends on the conceptual model chosen (m), but also depends on the way time and other resources are allocated (R) to the different tasks throughout the project. Both the model chosen and the resource allocation depend on how much resource is spent on the conceptual modelling process – for example, spending more means that the conceptual model chosen is likely to be better, but this leaves less for the remaining project tasks (coincidentally this type of trade-off is similar to the "commercials" exercise used in Willemain's (1995) experiment described later). The performance of each aspect also depends on various other factors (represented in a combined variable u_i) such as the nature of the problem, the knowledge and skills of the modeller and the input provided by the client. There will probably also be various trade-offs between the allocation of resources to the other simulation tasks apart from conceptual modelling but this is considered to be outside the scope of the conceptual modelling problem.

The further complicating factor is that there is a very wide range of alternative models (theoretically covering all possible different types and sizes of model) and so the task is not simply a search through a small well-defined set of alternative models. Identifying the models is, to some extent, a creative process. The extent of the choices available will depend on the project, in particular on the complexity of (the modeller's perceptions of) the system and the problem. A more complex system is likely to mean a greater choice of plausible conceptual models as there are more combinations of factors that could be included.

Therefore a suitable heuristic solution approach to this problem is an iterative process of developing new models, predicting their performance and deciding whether to stop. Figure 2 shows a framework for this selection process. The assessment of the predicted performance of the best model (or models if several are to be used) needs to consider whether the best model will produce a successful project and also whether searching for an even better model might be worthwhile (comparing the search time and effort with the potential improved performance).

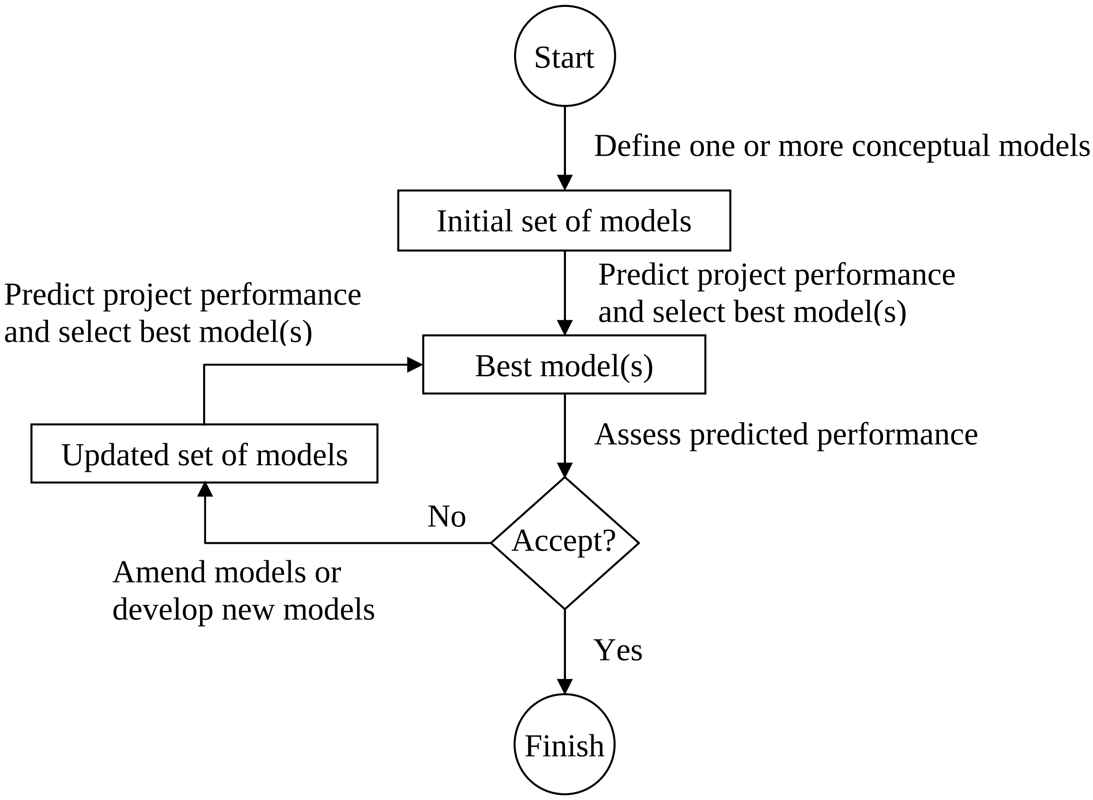

Figure 2. Conceptual model selection process

Willemain (1995) is one of the very few researchers to have collected empirical evidence of the modelling process and there is evidence from his experiment that expert modellers do go through such a process of iteratively developing and mentally evaluating alternative conceptual models. In Willemain's experiment experts spoke aloud their thoughts for the first hour of tackling an artificial problem. Tape recordings were made and transcripts produced and analysed, which provided very valuable insights into the nature of the experts' modelling approaches. The transcripts were broken into "chunks" (from a phrase to a couple of sentences) and each one categorised as one of five topics in Willemain's stages in the modelling process. There was a lot of switching between different modelling topics, particularly between "structure" (conceptual modelling) and "assessment" (verification and validation) as modellers evaluated and revised the conceptual model. Willemain comments that even though only the initial stage of the modelling process was studied the experts did "conduct thought experiments involving the other stages as an integral part of the formulation stage". Therefore the experts were probably predicting the performance of the conceptual models through these thought experiments as a basis for the choice of conceptual model. A follow-up study was recently carried out analysing transcripts of novice modellers tackling similar types of problems (Powell and Willemain, 2007; Willemain and Powell, 2007). One of the important differences that Powell and Willemain identified between novices and experts was that the novices spent much less time on "assessment" and so did not appear to evaluate and revise the proposed models to the extent that the experts did.

Willemain (1994) also carried out a survey of the modellers who participated in his experiment, and the findings included a strong sense of the importance of creativity in the modelling process. The words and phrases given by the experts for important qualities of an effective modeller included creative (2 experts), creativity (2 experts), imaginative, insightful, inspired, willingness not to pigeon hole problems, and open-minded. There were also 7 point Likert scale questions with a strong overall response under "the way the experts model" for several questions including "look for analogies" rather than "start from scratch", "start small and add" rather than "start big and subtract", and "always draw / doodle" rather than "never draw / doodle".

A key part of this solution approach for the conceptual modelling problem is to relate the choice of model to the different performance aspects. Experts can use past experience as a basis for predicting performance, but studies that investigate the effect of different model attributes on performance would also provide useful information. The next section describes such a study.

Experiment on model characteristics and performance

Amongst the 11 performance elements set out above are the time taken to build the model, the likelihood of errors and the ease of understanding of the model and the results. These were compared for four discrete event simulation models of production lines by analysing the performance of the 33 students on the M. Sc. Operational Research course at the University of Birmingham in answering questions on and in building the models. The reason for this approach is that, in order to assess the effects of the differences between the models on these performance elements, the models should be built by different people of roughly equal ability and experience. Otherwise, if more than one model is built by the same person, building and analysing the first model helps with the next.

As mentioned above, the complexity of a model is the most common model characteristic related to performance in the literature, and yet it is not defined clearly. Brooks and Tobias (1996) considered the overall complexity of a model to be a combination of its size (the number of nodes or elements), its connectedness (the average number of connections per element) and its calculational complexity (the complexity of the calculations making up the connections). The aim of the experiment was to examine the effects of these characteristics and so the models were devised to differ in these three aspects.

The models used are shown in Figure 3. Since they represent production lines, the natural definition for the elements is machines and buffers with the connections being the routes taken by the parts. Models A and B both have eight machines and eight buffers in the same basic layout with model A having more part routes (23 compared to 19) and hence higher connectedness. Model C has five machines and five buffers laid out in the same way as a portion of model A and differs from A mainly in size. Model D has only three machines and three buffers but has the most complex calculations to determine the part routes. Model D has high connectedness and calculational complexity.

The models were assigned at random to the students. The students were quite inexperienced modellers having received between 14 and 16 hours of tuition, mainly consisting of hands on experience together with some formal teaching and demonstrations. The first stage of the experiment aimed to compare how easy the models were to understand. The students were each asked the same four written questions on aspects of the behaviour of the particular model assigned to them, and were provided with the model description and selected model output. The second stage focused on model building and the students were each timed as they built their model using the WITNESS software. The number of errors in each model was subsequently determined (the students were instructed to build the model as quickly as they could but not to test it). The results are shown in Table 1.

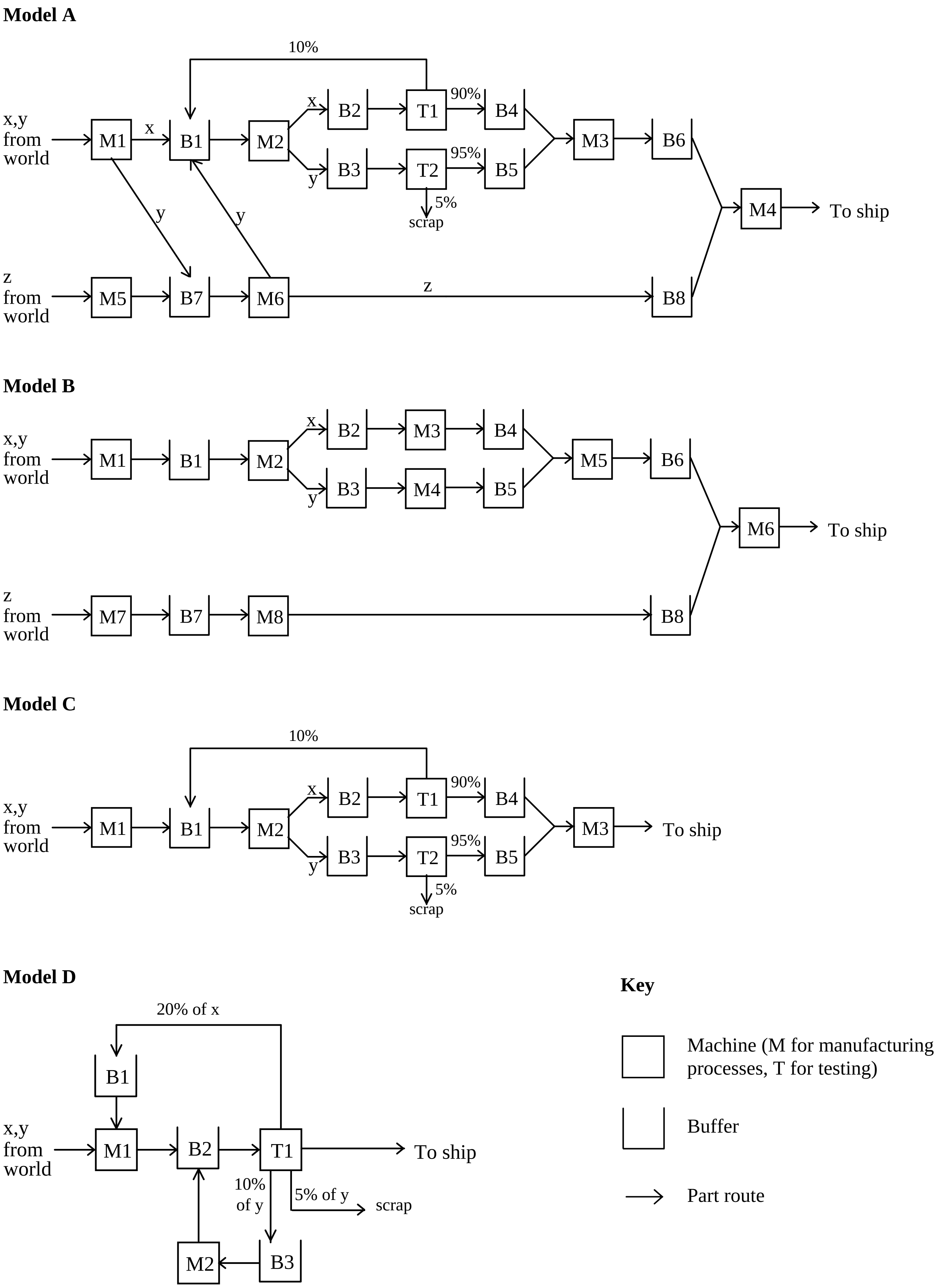

Figure 3. Process diagrams of the models used in the experiment.

Table 1 Results from the experiment.

Model	Aspects of complexity			Performance		
	Size	Connectedness	Calculational complexity	Understanding	Build time	Errors
	No. elements	Av. connections per element		Average marks	Average no. minutes	Average no. per model
A	16	1.43	Medium	33%	52.5	2.63
B	16	1.19	Low	45%	44.4	2.25
C	10	1.40	Medium	39%	46.3	1.63
D	6	1.67	High	57%	(1) 62.4	(1) 2.89

(1) Includes two students who had not quite finished within the maximum time. Six of the 26 errors on this model are omissions in these two models which are probably due to lack of time.

Using analysis of variance (ANOVA), the differences between the models are statistically significant at the 5% level for build time ($P = 0.032$), question 2 ($P = 0.014$) and question 3 ($P = 0.022$), but not for question 1, question 4, the average mark for all questions and the number of errors.

For build time calculational complexity appears to have the most effect with model D taking considerably longer to build than the other models. With a package like WITNESS which is user friendly and already contains many of the constructs required, thinking time is the most important component of the build time and so it is the complex and less familiar commands that are the most important. Observations also indicated that the aspects of the model that were easy to code were completed very quickly by the students.

The questions were analysed both by comparing the marks and by considering the reasoning process required to answer each question, which is discussed in detail in Brooks (1996). Students were asked to give a reason for their answer and considerable importance was given to this since the aim was to assess understanding. Both the correct answer and correct reason were required to score 1 mark. If either the answer or reason were only partially correct then ½ mark was awarded. An incorrect answer or the correct answer with an incorrect reason scored 0 marks. As stated above, the significant differences between the models were on questions 2 and 3. Question 2 asked “Which machine(s) is the bottleneck?” and the average mark was much higher for the model D participants (72%) than for the other models (19%, 31% and 25% respectively). The small size of model D made this question easier to answer because there are fewer elements to compare to identify the bottleneck. In fact machine M2 is rarely in operation and so this question only required comparing 2 machines. This also meant that there were more acceptable reasons for the correct answer than for the other models. Question 3 asked “Which buffer(s) were full at some time during the period?” and could be answered by identifying blocked machines from the output statistics. The average marks were much higher for models B and D (81% and 72% respectively) than for models A and C (25% and 44% respectively). Again this reflects the question being inherently easier for models B and D since the blocked machines only sent parts to one buffer whereas in models A and C they sent parts to several buffers. Therefore, the difference in marks seems to be a result of lower connectedness in the critical section of the models.

The marks were not statistically significant at the 5% level for questions 1 and 4. Question 1 (“How many parts were sent to SHIP in the period?”) was expected to be harder for model D since the calculation is more complex but, in fact, the average mark was similar to that for models A and B perhaps again reflecting that the small size means that it is easier to identify the correct part of the model to focus on. Question 4 (“Estimate the % increase in output if [a given machine] cycle time is reduced to 10” where the given machine was chosen not to be the bottleneck) was expected to be easier for model D but the marks were only slightly higher than for the other models.

Overall the indication is that the difficulty in understanding is mainly affected by size and connectedness with calculational complexity being much less important, although this of course depends on the specific question being considered. This is probably because the fine details can often be ignored in understanding the system with just an appreciation of which elements influence each other being required.

Most of the model building errors for models A, B and C occurred in the input and output rules for assembly machines, which were relatively complex commands which the students were less familiar with. The number of errors therefore reflects the comparative occurrence of these commands in the models, with models A and B having two assembly machines and model C one (each error was counted separately including repeated errors). Most of the errors for model D were either omissions or occurred in a complex command unique to model D. Generally, the majority of errors are likely to occur in the more complex aspects of the model, and so the number of errors is expected to be most closely related to calculational complexity.

The sample sizes here for each model (8 or 9) are small and the results will depend to some extent on the type of models used and the questions asked. The results can therefore only suggest possible relationships between model attributes and performance and more work is required to investigate this further.

Other modelling studies

One approach to studying the conceptual modelling problem is to compare alternative models for a system and these alternative models can be developed by taking a complex model and simplifying it in various ways. This approach was applied to models of a groundwater system, poppy population genetics and a manufacturing system with one of the aims being to provide insights into the choice of model (Brooks, 1996; Brooks and Tobias, 1999). This can also be a good strategy in a modelling project, particularly where the main aim is increased understanding as was experienced in a wheat simulation project (Brooks et al., 2001). Techniques used in identifying good model simplifications included sensitivity analysis and examining the detailed workings of the model and the results. This resulted in interesting insights and in several cases a new analytical model was developed. The main simplifications were developing a probability based model of the plant population genetics simulation, using simplified parameter structures in the groundwater model, using aggregate variables in the wheat model, and bottleneck analysis of the manufacturing system simulation. The result in each case was a hierarchy of models where the simple model was easier to use and understand but where the confidence in the simple model came from a comparison with the more detailed model. Greater confidence in the detailed models was due to their stronger theoretical basis as there is a direct correspondence with observed elements and known relationships. For several of the models the output of interest was an average and this was considered to be one of the reasons that good simplifications could be found (Brooks and Tobias, 1999). The simplified models also generated some additional theoretical results and enabled easier experimentation. However, simplification is time consuming (Rexstad and Innis, 1985) and risky as there is no guarantee of success. The simplified model may also have a reduced experimental frame (i.e., a more limited range of possible experiments).

Principles of conceptual modelling

Simulation modelling covers an extremely wide range of applications and situations and the resulting models vary enormously in scale. Therefore, there are very unlikely to be any formal rules that will apply in all or most situations, Nevertheless, there may be some general principles that can act as a useful guide. The following list sets out some suggested principles, some of which come from the literature whilst others are based on the conceptual modelling framework and the modelling studies previously described. Many of these may be well known to experienced modellers but hopefully will be useful for those with less experience.

1) As often mentioned in the literature, the project objectives should drive all the tasks including conceptual modelling. Different purposes for the same system (e.g., safe vehicle movement through a plant layout / choice of machines to improve plant throughput) usually require quite different models.

2) The conceptual model will be based on the knowledge and belief about the real system. Clearly better knowledge should mean a better conceptual model and so taking time to find out about the way the system works (e.g., from those familiar with the system) will be very beneficial (e.g., Law and Kelton, 2000).

3) The aim of conceptual modelling is to choose the model that will give the best overall project performance. As discussed in the conceptual modelling framework section, this requires balancing all the different aspects of performance.

4) The conceptual modelling problem can be tackled as a creative search process that develops alternative models and predicts their performance throughout the project. This requires thinking through all the different project tasks. From Willemain and Powell’s work, the approach of assessing and revising proposed models in this way may be one of the characteristics of experts that is not generally found in novices.

5) Understanding the relationships between model characteristics and performance will help with the assessment of the models. In particular, general advice from the literature is that a simpler model will be easier to understand, and easier and quicker to build, verify, run and analyse but will be less accurate and have lower validity. A simpler model may also have a more restrictive set of possible experiments. However, there will

sometimes be exceptions to such general rules (Brooks and Tobias, 1996). Considering the relationships between more specific model characteristics (such as size and connectedness) and performance is likely to enhance the assessment process.

6) The model must meet certain minimum criteria for the performance aspects, otherwise the project will not be a success. For example, it must produce the required outputs and have sufficient validity. If there are no such models then the project is not feasible and needs to be re-considered.

7) The needs of the client should be an integral part of the performance assessment so as to ensure project success (Ward, 1989; Robinson and Pidd, 1998).

8) Key criteria that can provide a basis for ideas for the conceptual model are that the inputs need to include all the options to be considered in the project and the outputs need to show the variables that will be used to assess which option is best (Robinson, 2004). The internal content of the conceptual model then connects these inputs and outputs (Robinson, 2004).

9) In some circumstances a model with a simpler structure may be more accurate, typically when suitable data is not available for the more complex model. For example, arrivals of customers at a bank would usually be modelled using a distribution from observed arrivals. Modelling the movements of the customers around the city would be a considerably more complex structure but good data is unlikely to be available and so it would be much less accurate.

10) Data acts as a constraint. General advice is that the model should drive the data and there are various ways of overcoming lack of data (e.g., Brooks and Robinson, 2001). Nevertheless, lack of data or poor quality data will limit the accuracy of and confidence in the model.

11) Conceptual modelling is a creative process. This is evident from Willemain's work, both from the nature of the transcripts of experts tackling the different problems (Willemain, 1995) and from the emphasis given by the experts that this is a desirable characteristic for modellers (Willemain, 1994). Having a mindset that the process has this nature and providing an environment to stimulate creativity may therefore be beneficial; for example, accepting a wide range of ideas initially before gradually

narrowing these down, and being prepared to change the conceptual model throughout the project.

12) A top down view of the conceptual model (meaning of conceptual model section) is conceptual model = knowledge of real system (including assumptions) + boundary + simplifications. Producing a list of the boundary, simplifications and assumptions can help to identify subconscious modelling decisions that may need to be challenged. This can also be a good record of the conceptual model.

13) Simplifications can be considered to be of two types. Type 1 are those simplifications that are apparent from a good knowledge of the real system. These simplifications can be applied when choosing the initial conceptual model. Type 2 are those that are derived from analysis of the model such as sensitivity analysis or detailed examination of the model behaviour and could not be seen easily without such analysis. Type 2 simplifications can therefore only be applied after the model has been built and so require revising the initial conceptual model.

14) The common advice of starting with the simplest model and adding detail really means apply as many as possible of the type 1 simplifications from principle 13. In practice, it may be difficult to define the simplest model straightaway. Instead, the easiest candidate model to start with in the conceptual modelling search process will usually be one where most of the model elements correspond directly to observed elements in the real system. This will also tend to have high face validity (Law and Kelton, 2000). Simplifications during the conceptual modelling search process could therefore start with looking for observed elements or rules that can be combined or left out of the initial model, before considering more advanced simplifications.

15) Initial simplifications can frequently be made by considering the scale of the phenomena of interest. Often factors that vary on a much smaller time or spatial scale can be averaged or aggregated, with factors that vary on a much larger scale treated as a constant or ignored (Courtois, 1985).

16) Consider building more than one model. A hierarchy of models may be the best solution to the choice of conceptual models. In particular, if type 2 simplifications can be applied then the resulting model may provide improved understanding and speed. However,

confidence in the model itself may not be high but comparison with the original model (with high validity) may give sufficient confidence.

17) Documenting the conceptual model is very valuable in providing a reference for the other project tasks (e.g., verification) and for communication with the client or other modellers. There are various ways of documenting the conceptual model such as a component list, process flow diagram, logic flow diagram or activity cycle diagram (Robinson, 2004).

Future research

Research into conceptual modelling is still in its infancy hindered to date by a view that is an art rather than a science. Common advice is to learn from experience and empirical evidence of how experts approach conceptual modelling would provide a good foundation. Research could include extending Willemain's (1995) experiment by using other participants and other problems (such as problems where simulation is the appropriate technique). Evidence could also be obtained for real simulation projects by shadowing experts, asking experts to record their experiences, or by conducting a survey or structured interviews. Similar data from novice modellers would enable a comparison of modelling styles, may indicate the specific areas in which novices lack skill or knowledge (which could be used to improve simulation training), and may also enable some evaluation of the effectiveness of different teaching methods. A study in this area has recently been carried out (Powell and Willemain, 2007; Willemain and Powell, 2007). Data on real projects for an expert and several novice groups was obtained by Wang and Brooks (2007). Experiments to investigate the relationships between model characteristics and performance (e.g., similar to experiments described in this paper and in Wang and Brooks, 2007) would inform the prediction of performance during conceptual model selection. Alternative models could also be built and evaluated for specific projects, again similar to work outlined in this paper. Looking at the applicability and benefits of ideas from research on creativity could also be fruitful. Alternative ways of documenting the conceptual model could be investigated and more consistent use of a particular method would facilitate the comparison of models. If a large body of conceptual modelling research over a range of applications could be produced then this would help to identify and evaluate general principles as well as giving insight into the effects of application type. Related areas are simplification and meta-modelling and although ways of simplifying a model have been discussed in the literature (e.g., Zeigler, 1976; Innis and Rexstad, 1983; Robinson, 2004) there is no established methodology. A first step could be developing a clearer idea as to when simplification is likely to be beneficial.

Conclusions

Conceptual modelling is the stage in the simulation process that has received the least amount of attention. The conceptual model is a software independent model description and the aim is to choose the model that will give the best overall performance for the project. The conceptual modelling task can be viewed as a difficult optimisation problem that can be tackled effectively using a creative search process that develops alternative models and predicts their performance throughout the project. A better understanding of how certain model characteristics affect performance would help this process. For example, the small scale experiment described indicates that increased size and connectedness may reduce understanding while model building time is mainly affected by the occurrence of complex rules. Based on the literature and my modelling experiences 17 principles of conceptual modelling were suggested which may provide some useful guidance, particularly to inexperienced modellers.

However, much more research is needed in this area, including research on the relationships between model attributes and performance, on how experts and novices approach conceptual modelling and on methodologies for model simplification. A better understanding of conceptual modelling would be particularly valuable in training novice modellers but may also help experts to improve their modelling skills leading to even greater success in simulation projects.

References

Brooks R J (1996). A Framework for Choosing the Best Model in Mathematical Modelling and Simulation. Ph. D. Thesis, University of Birmingham.

Brooks R J and Robinson S (2001). Simulation, with Inventory Control (author C. Lewis), Operational Research Series. Palgrave: Basingstoke.

Brooks R J and Tobias A M (1996). Choosing the Best Model: Level of Detail, Complexity and Model Performance. Mathematical and Computer Modelling 24(4): 1-14.

Brooks R J and Tobias A M (1999). Methods and Benefits of Simplification in Simulation. In Al-Dabass D and Cheng R (eds). Proceedings of the U.K. Simulation Society (UKSIM 99): U.K. Simulation Society, pp. 88-92,

Brooks R J, Semenov M A and Jamieson P D (2001). Simplifying Sirius: Sensitivity Analysis and Development of a Meta-Model for Wheat Yield Prediction. European Journal of Agronomy 14(1): 43 - 60.

Checkland P (2006). Learning for Action: A Short Definitive Account of Soft Systems Methodology, and Its Use Practitioners, Teachers and Students. John Wiley and Sons: Chichester.

Courtois P-J (1985). On time and space decomposition of complex structures. Communications of the ACM 28(6): 590-603.

Innis G S and Rexstad E (1983). Simulation model simplification techniques. Simulation 41(1): 7-15.

Law A M (1991). Simulation model's level of detail determines effectiveness. Industrial Engineering 23(10): 16-18.

Law A M and Kelton W D (2000), Simulation Modeling and Analysis, 3rd edition. McGraw Hill: New York.

Pidd M (1998). Computer Simulation in Management Science, 4th edition. John Wiley and Sons: Chichester.

Pidd M (2003). Tools for Thinking, Modelling in Management Science, 2nd edition. John Wiley and Sons: Chichester.

Powell S G and Willemain T R (2007). How novices formulate models. Part I: qualitative insights and implications for teaching. Journal of the Operational Research Society: in press.

Rexstad E and Innis G S (1985). Model simplification - three applications. Ecological Modelling 27(1-2): 1-13.

Robinson S (2004). Simulation: The Practice of Model Development and Use, John Wiley and Sons: Chichester.

Robinson S (2006). Issues in conceptual modelling for simulation: setting a research agenda. In Garnett J *et al* (eds). Proceedings of the 2006 Operational Research Society Simulation Workshop (SW06): O. R. Society, pp. 165-174.

Robinson S and Pidd M (1998). Provider and customer expectations of successful simulation projects, Journal of the Operational Research Society 49: 200-209.

Salt J D (1993). Keynote address: simulation should be easy and fun!. In Evans G W *et al* (eds). Proceedings of the 1993 Winter Simulation Conference: IEEE, New York, pp. 1-5.

Shannon R E (1975). Systems Simulation: The Art and Science, John Wiley: New York.

Tilanus C B (1985). Failures and successes of quantitative methods in management. European Journal of. Operational Research 19: 170-175.

Wang W and Brooks R J (2007). Improving the understanding of conceptual modelling. Submitted to Journal of Simulation.

Ward S C (1989). Arguments for constructively simple models. Journal of Operational Research Society 40(2): 141-153.

Willemain T R (1994). Insights on modeling from a dozen experts. Operations Research 42(2): 213-222.

Willemain T R (1995). Model formulation: what experts think about and when. Operations Research 43(6): 916-932.

Willemain T R and Powell S G (2007). How novices formulate models. Part II: a quantitative description of behaviour. Journal of the Operational Research Society: in press.

Zeigler B P (1976). Theory of Modelling and Simulation. John Wiley: New York.