Pergamon

Mathl. Comput. Modelling Vol. 24, No. 4, pp. 1–14, 1996

0895-7177/96 $15.00 + 0.00

PII: S0895-7177(96)00103-3

Choosing the Best Model: Level of Detail, Complexity, and Model Performance

R. J. BROOKS* AND A. M. TOBIAS
Operational Research Group
School of Manufacturing and Mechanical Engineering
University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.
r.j.brooks@birmingham.ac.uk

(Received March 1996; accepted May 1996)

Abstract—The lack of a methodology, or at least detailed guidelines, for choosing the best model in a mathematical or computer modelling study stems from a poor understanding of the precise ways in which the success of the study depends upon the particular model used. As a result, the choice of the best model is regarded as more of an art than a science. In order to improve the model selection process, model performance needs to be clearly defined, and suitable model attributes identified that can be used to predict the performance of the alternative candidate models. This paper distinguishes the different aspects of model performance and considers the extent to which they can be measured. The most common attributes used to compare alternative models are level of detail and complexity although these terms are used in a number of different ways. The meanings of these concepts are therefore discussed and the likely relationships with the model performance elements considered. The related area of simplification is reviewed and the areas in which further work is required are set out.

Keywords—Level of detail, Complexity, Model performance, Simplification, Modelling methodology.

1. INTRODUCTION

Mathematical and computer models are widely used in many areas of science and industry, from population genetics to climate modelling and from simulating a factory production line to theories of cosmology, etc. Modelling may be undertaken for a number of reasons, but the most common aim is to predict the behaviour of a system under particular circumstances when it is impossible, or at least undesirable, to experiment with the system itself. The understanding of the system gained by the modeller and the user can also be an important benefit of the project [1], particularly in scientific research when it can be the principal objective. A model may be purely predictive or it may also be part of a decision making process, whereas there are other occasions when a model is just descriptive, simply summarising the modeller's understanding of the system [2]. Equally, a modelling project may have other objectives such as helping to design experiments or identifying research requirements. Despite the great variation in the types of model and their usage, the modelling process itself will take the same form for most projects, and can typically be split into the steps of problem formulation, collection and analysis of data, model formulation, model construction, verification and validation, experimentation, analysis of results, conclusions and implementation. Definitions of verification and validation are not consistent in the literature,

*Author to whom all correspondence should be sent.
We are grateful to B. P. Zeigler and S. C. Ward for their helpful comments on an earlier draft of this paper.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

but here verification means the process of testing whether the model is working as intended, and validation means the process of comparing the model output with historical data [3]. The modelling steps do not form a linear process but one which may involve many loops, and so the choice of model affects each of the steps.

The choice of the best model arises in the model formulation step. This step produces a conceptual model, which is a complete specification of the model to be built, and the aim of this step is to select the conceptual model that will lead to the best overall project outcome. The actual building of the model is a separate step. The range of possible models is usually very broad because a great deal is known about the system relationships on many different spatial and time scales [4]. The choice of the best model is often viewed as the problem of choosing the appropriate level of detail, and this is considered one of the most difficult aspects of the modelling process [5], and one which has a major effect on the successfulness of the project [6–8]. If a unique level of detail can be assigned to each of the possible conceptual models then the selection of the level of detail and the choice of the conceptual model are equivalent steps. In practice, groups of models may be considered to have the same level of detail and so the choice of the level of detail may be just one part of the model formulation step.

By viewing the selection of the conceptual model in this way, the alternative models are effectively being ordered by the characteristic of level of detail, which is the most common characteristic used to compare models. This is done in the hope that there will be similarities with previous studies in the effect of the level of detail on model performance, so that experience from these studies can be applied in the selection of the current model. For example, a model that is too simple will be unrealistic and so its results will be, at best, of little use, and at worst, misleading. On the other hand, considerable resources are usually required to build a complex model and so, if the model is too complex, constraints on resources may prevent the completion of the project (it is assumed here that a more detailed model will be more complex, although the meaning of level of detail and complexity are discussed in detail in Section 3). It is generally harder to understand the relationships contained in a complex model and this makes the interpretation of the results more difficult, possibly leading to incorrect conclusions being drawn. A complex model is probably more likely to contain errors as it is harder to verify that the model is working as intended.

The advice given on selecting the level of detail seems to consist almost entirely of vague principles and general guidelines. A commonly quoted maxim is Ockham's (or Occam's) razor, attributed to the 14$^{\text{th}}$ Century philosopher William of Ockham, and translated (from the Latin) as "entities should not be multiplied without necessity" or "it is vain to do by more what can be done by fewer." In other words, choose the simplest model that meets the modelling objectives. Often, the advice given is to start from a simple model and progressively add detail until sufficient accuracy is obtained. It is important to match the complexity of the model with the modelling objectives and with the available data [9–11]. However, knowledge of these principles is of only limited use to the modeller and the choice of the level of detail seems to be regarded as more of an art than a science.

The scarcity of research in this area is surprising, given the widespread use of mathematical and computer models in science. This can be attributed partly to the difficulty in measuring either the performance or suitable attributes of models and partly due to the fact that building alternative models can be very time consuming. The following section considers the constituent elements of model performance and how they might be measured. The process of choosing the best model can be improved by relating model performance to appropriate model attributes, so that the performance of the alternative models can be predicted from their attributes without having to actually build each one. As already discussed, the level of detail and complexity are often used to compare alternative models and so they are likely to be good attributes to use. The meanings of these two terms are therefore considered and then the current understanding of the way the model performance is related to level of detail and complexity is discussed, along with an approach for improving this understanding. Simplification and other related areas are briefly reviewed, and the paper concludes by setting out the areas in which more work is required.

2. MODEL PERFORMANCE

The evaluation of the performance of a model should cover the impact of the model on all aspects of the project. Such an assessment is dependent on the particular project; a model would give different performance if used for two different projects. The full evaluation should include the following 11 performance elements:

RESULTS.

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

FUTURE USE OF THE MODEL.

4. The portability of the model and the ease with which it can be combined with other models.

VERIFICATION AND VALIDATION.

5. The probability of the model containing errors (i.e., the model constructed does not match the conceptual model).
6. The accuracy with which the model fits the known historical data (validity).
7. The strength of the theoretical basis of the model including the quality of input data (the credibility of the model).

RESOURCES REQUIRED.

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 (e.g., computer memory) of running the model.

An assessment of the modelling project as a whole would compare the benefits of the project, which are the overall project conclusions together with the benefits obtained from any actions taken as a result, with the costs incurred. The performance assessment of the model should consist of the impact of the choice of the model on these costs and benefits. The performance elements, therefore, attempt to focus purely on the effect of the choice of conceptual model, but it is not possible to isolate entirely the effect of this choice. For example, the resources required to build the model not only depend upon the model used, but also on a number of other factors such as the ability and experience of the modeller. A single absolute measure of model performance cannot be obtained, but a meaningful comparison of alternative models in similar circumstances should be possible.

The quality of the conclusions arising from using the model depends upon the quality of the results, which is a combination of their accuracy and the extent to which they address the modelling objectives (elements 1 and 2). In the terms of Zeigler [12], element 1 is the extent to which the model results cover the experimental frame. Clearly, if the results are inaccurate then the decisions taken, and conclusions drawn, are likely to be incorrect. It is also important that the model and the results can be understood (element 3) to facilitate the analysis. This element is particularly important if the main aim of the project is to increase understanding of the system. The use of the model in future projects (element 4) can also be an important benefit.

In most cases, the model is predictive, and so elements 1 and 2 cannot be assessed until some time after the modelling project has been completed. Acceptance of the conclusions from the modelling and implementation of the recommendations requires that the user has confidence in the model, and the user's confidence should be based on elements 5–7. It is therefore important not only that (with the benefit of hindsight) the model produced realistic results (element 2), but also that the model was seen to be realistic at the time the project was carried out (elements 5–7). It is possible for a very unrealistic model to produce accurate results (for example, due to compensating errors, particularly if the results are a single value). Even if the results of such a model were accepted, and these led to the correct decisions being taken, the understanding gained of the system is likely to be incorrect and this may have serious consequences in the future. Elements 5–7 take this into account by giving an assessment of the underlying quality of the model. Successful reuse of the model also requires the model to have a sound basis, as well as requiring the model to be portable (element 4).

Element 5 relates to the process of verification and so is concerned with errors occurring in the model construction step (as opposed to elements 6 and 7 which relate to the model formulation step). It is not possible, for most models, to ensure that the model constructed will operate as intended in all circumstances (i.e., to fully verify it) and so the model may contain errors [3,13]. The probability of the model containing errors is the product of the probability of the initial model containing errors, which partly depends upon the choice of the model (as well as the quality of the model building), and the probability of the errors not being discovered in the verification process, which also partly depends upon the choice of the model (as well as the quality of the verification process). There is a trade-off here between the performance elements; the model is less likely to contain errors if more resources are put into building and verifying the model.

Elements 6 and 7 assess how well the conceptual model matches the real system. It is not sufficient for the model output simply to fit the historical data; confidence in the model structure, either on a theoretical basis or on the basis of successful previous experience, is also important. It is possible for a model to fit the historical data well, but to give poor predictions if the basis of the model is incorrect, particularly if the conditions for the period being predicted are very different to those in the past. If the system being modelled is one that does not currently exist, validation can consist only of an assessment of model credibility [13], i.e., element 7.

The effect of the choice of the model on the costs of the project is addressed by elements 8–11. In some projects, a model already exists and the aim of the project is to simplify it or to make it more complex. In this case, the time and cost of building the model becomes the time and cost of modifying the model. Considerable effort can be required to simplify an existing model [14].

An assessment of model performance requires a measurement to be made of each of the performance elements and this is far from straightforward. It should be possible to evaluate elements 2, 6, 9, and 11 easily in most cases, although care is required in the interpretation of the measures used. However, the remaining elements are hard to quantify, and a subjective qualitative assessment may be all that is possible. The ease of understanding and the probability of errors both contain a human element which makes a numerical evaluation difficult (elements 3 and 5). Similarly, the strength of the theory behind the model is subjective (element 7). The resources required to build and analyse the alternative models should be those required if the model is built from scratch with no prior knowledge of the system (elements 8 and 10). Such an assessment therefore, ought ideally to consist of measuring the resources used by independent modelling teams of equal modelling experience and ability, but this will not be feasible in most instances. Meaningful measures for the extent to which the scope and detail of the results matches the problem requirements, and particularly the model portability, are also likely to be difficult to derive (elements 1 and 4). An overall assessment of model performance requires a relative weighting to be given to each of the elements and such a weighting will be subjective and will vary considerably from study to study. It may be possible, for a particular study, to ignore some of the elements as insignificant in terms of the overall performance. However, if a number of studies attempt to measure at least some of the performance elements, the measurement procedure is likely to improve.

3. THE MEANING OF LEVEL OF DETAIL AND COMPLEXITY

There is no single accepted definition of either level of detail or complexity, and in fact, a formal definition is rarely attempted, but, when applied to a model, the level (or amount) of detail usually means an assessment of the extent to which the observable system elements and the assumed system relationships are included in the model. It is usually assessed qualitatively and is most often used just to rank alternative models. Level of detail tends to refer to the system that the model represents (for example, in the case of a model of a production line, the number of machines, parts, etc., included in the model) rather than to the precise way in which the model is implemented (such as the number of variables used). Models are often described as being detailed, meaning that the model contains most of the elements and interactions thought to exist in the system being modelled. For example, see [15–18].

The term complexity is much more common than level of detail, although it is used in many different ways [19–22], such as the difficulty in computing a function (computational complexity), a structural attribute of a piece of software (software complexity), the difficulty experienced by people in perceiving information or solving problems in a particular environment (behavioural complexity) and the complexity of terms, sentences and theories (logical and semantic complexity). Weaver [23] categorised the problems tackled by science as problems of simplicity, organised complexity (the most difficult), and disorganised complexity, and Chaitin [24] equated the complexity of a number with its randomness. Recently, the phrase "science of complexity" has been used to describe a scientific discipline covering the study of complex systems that give rise to emergent properties [25,26].

Systems theory has perhaps seen the greatest discussion of the concept of complexity, although this has resulted in a number of different meanings. Flood and Carson [27] in their extensive discussion of complexity considered that complexity meant "anything that we find difficult to understand." They therefore viewed complexity as a combination of the structure of an object (particularly the number of elements and relationships) and of the nature of people and the way in which they interact with the object in the particular context. Golay *et al.* [28] also equated complexity with the difficulty in understanding the system, whereas Simon [29] took a complex system to be "one made up of a large number of parts that interact in a non simple way." Casti [20] referred to the difficulty in understanding a system's structure as static complexity which he distinguished from the dynamic complexity of the system output, noting that a simple structure can produce complex output (for example, a simple system can be chaotic). Rosen [30] defined a complex system "as one with which we can interact in many different kinds of ways, each requiring a different mode of system description" which varies according to the point of view of the observer, whereas George [31] considered a system to be complex when it contained sufficient redundancy to be able to function despite the presence of many defects.

Dictionaries usually define complex as being something consisting of many parts, as well as often defining it as something that is difficult to understanding (see also Ward's [7] discussion of the meaning of simple). For example, the Collins English dictionary's definition of complex is "1. Made up of various interconnected parts; composite. 2. (of thoughts, writing, etc.) Intricate or involved" and Chamber's English dictionary gives "Composed of more than one, or of many parts: not simple: intricate: difficult." Clearly, objects that have many interacting parts do tend to be difficult to understand and vice versa. The complexity of a model is, therefore, sometimes used to mean the difficulty in understanding the model or the difficulty (in terms of resources required) in generating model behaviour [12,32,33]. However, these are performance measures rather than model attributes and so, in the first case, for example, the commonly asserted disadvantage of complex models that they are difficult to understand is just a tautology [7].

Complexity is commonly used to refer to a model when comparing the output of alternative models (for example, [34–38]). In such cases, complexity is usually not defined. However, it is the structure of the alternative models that is described with the difficulty in understanding being rarely mentioned. It therefore appears that the complexity of a model usually refers to a structural property of the model, and this is certainly the appropriate usage when comparing the characteristics and performance of different models.

Complexity is, therefore, being used in a very similar way to level of detail and, in comparing models, a number of authors appear to equate the terms complex and detailed when referring to the models (for example, [7,17,37,39]). Certainly a simple model is generally considered to be the opposite of both a detailed model and a complex model. Applying the dictionary definition, complexity would be a measure of the number of constituent parts and relationships in the model, and so complexity should differ from level of detail in referring to the actual model elements rather than the system elements, and this is assumed here. However, the level of detail largely determines the complexity and, in most cases, the ordering of alternative models by level of detail or complexity will be the same. If the system being modelled is extremely complex, then it is possible to build a model that has many parts but which omits many system elements, so that such a model would be complex but not detailed. In comparing two models, occasionally the more detailed model may be less complex if, for example, an approximation to a detailed system relationship requires more model elements and connections than the actual relationship, although such an approximation would be unlikely to have any advantages. There may also be several ways in which a given modelling assumption can be implemented (such as alternative algorithms for generating pseudorandom numbers), so that models of the same level of detail may have different complexity.

Assuming that the model can be considered as a number of interconnected parts, or components, the overall complexity of the model is taken here to be a combination of three elements: the number of components, the pattern of the connections (which components are related), and the nature of the connections (the complexity of the calculations determining the relationships). The aim here is to identify invariant structural attributes of the conceptual model, and so the frequency of occurrence of each connection has not been included as an element of complexity as this may depend on the particular model runs carried out. We term the three elements: size, connectedness, and calculational complexity, respectively.

4. MEASURING MODEL COMPLEXITY

If the model can be specified as a number of connected components, then it can be represented as a graph, with the nodes of the graph representing the components, and the edges representing the connections (i.e., the relationships) between the components, and graph theory measures used to measure the first two of these elements. However, for any given model there are likely to be several possible graphs and many alternative measures.

Models implemented as computer programs can, for most programming languages, be graphically represented by the program control graph in which the nodes represent blocks of code in which control is sequential and the edges represent branches in the program. The complexity measure proposed by McCabe [40] was the number of edges, less the number of nodes, plus twice the number of connected components in the program control graph. For a single program (so the number of connected components is one), this is equal to the cyclomatic number (which is the number of nodes less the number of edges, plus the number of connected components) of the program control graph with an additional edge added to join the last component to the first. This additional edge strongly connects the graph so that the cyclomatic number is equal to the maximum number of linearly independent circuits in the graph. It, therefore, represents the number of basic paths through the program, which McCabe [40] equated to complexity.

A graph of a discrete event simulation model is also possible by depicting the events as the nodes and the relationships between the events as the edges [41]. Events are activities that alter the state of the model, and two events are related if the occurrence of one of the events can cause the other to occur (or can cancel the occurrence of the other). A directed edge from event A to event B indicates that if event A occurs and certain conditions hold then, after a specified time, event B will occur. Schruben and Yücesan [33] suggested several graph theory measures, including the cyclomatic number, which could be applied to event graphs to measure the complexity of a model. It is usually not necessary to explicitly model all the events occurring in the system and the event graph can be used to identify events not required. This means, however, that several graphs are possible for the same conceptual model. For consistency, the graph with the minimum number of events should be used for the complexity measure (although it is not clear whether there is only one such graph). Activity cycle diagrams, which connect the possible states of each entity in the model, are a further way of representing a discrete event simulation model as a graph [42].

A graph can also be obtained by assigning each possible model state to a node, and representing possible transitions between the states by the edges, or by letting the nodes represent the state variables and the edges interactions between the variables [12]. There can be many choices for the state variables. Graphs of the interaction between the state variables were used to measure the complexity of alternative fate models of toxic substances in a lake by Halfon [43,44]. He used Bosserman's [45] $\bar{c}$ measure which is the proportion of paths of length $\leq n$ (where n is the number of nodes) that exist in the graph. The measure can be obtained using the adjacency matrix A which has $a_{ij} = 1$ if there is a connection from node i to node j and $a_{ij} = 0$ otherwise. The matrix A^k, obtained by k multiplications of matrix A by itself using Boolean algebra, has $a_{ij} = 1$, if and only if, there exists a path from node i to node j, of length k. The $\bar{c}$ measure is then given by the sum of all the elements in the matrices $A, A^2, \ldots, A^n$ divided by n^3 (the total number of elements).

A graph theory measure may not always be static. Neural networks are typically defined in terms of nodes and connections, and many other adaptive systems models can be represented in this way [46], which gives a natural graph of the model structure. In these models a complexity measure based on such a graph would change as the model runs as a result of the connections changing.

In comparing models, differences in the complexity of the models may be due to differences in the complexity of the calculations, and so the graph theory measures may be inappropriate or may need to be combined with other measures. For example, in a previous study, we compared 16 alternative population genetics models of the self-incompatibility alleles of a poppy population [47]. Each new generation was produced by the model by random mating from the previous generations. For half the models (those with seed dormancy), the parents can come from any of the previous 10 generations, and for the other half the parents can only come from the previous generation. Within each of these two classes, the models have the same structure and so would have the same program control and event graphs. Each class can be further subdivided into two classes of four models (depending on whether or not a single plant size is assumed) in which the possible states of the model are the same. The difference between the models consists of differences in the way the parents for a new plant are selected (for example, whether the parents are chosen completely at random or whether realistic functions for the distance travelled by the seed and pollen are used). The difference is, therefore, in the complexity of the selection process. A possible measure for this is to count the number of logical and mathematical steps in each case. It may be possible to regard the models as being just variants of the same single model with different parameter values but we consider the alternatives to be different models since the inclusion of the various additional factors are effectively based on different conceptual models.

An alternative approach to graph theory may be to use concepts from information theory. Golay *et al.* [28] used the following information entropy measure as a complexity measure:

$$H = -\sum_{i=1}^{n} p_i \log_2 (p_i),$$

where

H = information entropy,
n = number of system states, and
p_i = probability of the i^{th} state.

They justified the use of information entropy as a complexity measure by arguing that entropy measures the amount of uncertainty, and that a more complex model is more difficult to understand, and therefore, more uncertain. This measure can only be used in the very limited cases when it is practical to estimate the probability of each system state. In the case of the population genetics models, each population had 3840 plants with 120 possible genotypes, and so all the models had at least 120^{3840} states. Golay *et al.* [28] applied the measure to systems in which each component had only two states. The use of the entropy measure can also be argued on the basis that it measures the complexity of the behaviour of the model [48], in terms of both the number of systems states in total and relative proportion of time spent at each state. Again this indicates a likely correlation with the difficulty in understanding the model and its results. This measure, as it stands, does not measure our concept of the complexity of the conceptual model, as it is not a direct measure of a structural property of the model, but rather a measure of the complexity of the model behaviour for a particular run. However, it may be possible to use similar concepts to measure the amount of information contained in the model, although it is not clear how to do this at present.

In computer science, many measures (usually termed metrics) of the size and complexity of the code have been proposed, and these can be used to measure the complexity of a model taking the form of software. Most of the metrics are based on counting the occurrence of particular items in the code, such as the number of decision points or just the number of lines of code. An alternative approach developed by Henry and Kafura [49] is to identify the flows of information between separate program procedures, and to incorporate the number of different flows into the metric. Many of the software metrics, however, are partly dependent on the programming language used, whereas our aim is to measure the complexity of the conceptual model which should be independent of the specific implementation of the model.

The purpose of a complexity measure is to characterise the model so that this information can aid the choice of model by predicting model performance. Ideally we would like to have a single, system independent definition and measure of complexity covering all the aspects of the level of detail of a model and applicable to all conceptual models. However, no such definition or measure exists and as a result the term complexity itself is a source of confusion due to its usage in many different contexts. The best approach would seem to be to identify more specific model attributes, such as the attributes of size, connectedness, and calculational complexity discussed earlier, and to devise measures for these. It is important that such measures and the type of measurement scale should match our intuitive notion of the nature of the attribute, which, for example, is not always the case with software metrics [50]. In addition, the earlier in the modelling process in which a measure can be obtained, the more useful it is.

5. THE RELATIONSHIP BETWEEN MODEL PERFORMANCE AND THE LEVEL OF DETAIL OR COMPLEXITY OF THE MODEL

The level of detail and complexity of a model are widely recognised as having a very important effect on model performance, and the relationships between either level of detail or complexity and model performance has been discussed in general terms in a number of places (for example, [5,51–54]). A more complex model is expected to have greater validity and to give more detailed and accurate results, but to use more resources, and to be more likely to contain errors, more difficult to understand, and less portable (i.e., better for performance elements 1–3, 5, and 6 but worse for the others).

In the fields of management science and operational research, modelling projects are often carried out for clients with little modelling experience. There are a number of additional reasons why clients may prefer a simple model, and so are more likely to implement the results. Ward [7] set out a number of advantages, from a client's point of view, of using a simple model including the quicker generation of results and the production of results that are easier to understand and less specific, thus allowing the clients own preferences to be incorporated. A simple model is also more flexible and so can be more easily adapted if the project objectives change [9].

The precise nature of the relationships between the level of detail or complexity and the aspects of model performance are poorly understood. There are a few studies which have compared alternative models from different points of view. However, the objectives of these studies have not specifically been to relate model performance to either the level of detail or complexity of the models and as a result they have tended not to use quantitative measures. We briefly review these studies as they do provide some indication of the possible relationships.

The studies of Stockle [17] and Rexstad and Innis [14] took existing ecological models and attempted to simplify them. Stockle simplified a model of the amount of radiation intercepted by plant canopies. The most detailed model had nine leaf inclination classes, nine azimuth angle classes and 20 layers of leaves. The number of classes of each of these three elements could be reduced to simplify the model and Stockle found that by doing this the computation time of the model could be reduced by a factor of 12 with a negligible change in results, and by 63 with only a small change in results. This suggests that many models may be more complicated than they need to be. Our own experience of discrete event simulation models also indicates this. Innis and Rexstad [55] produced a list of simplification techniques and they subsequently applied some of these techniques to three models [14], but this paper focused on the applicability of their techniques to the models rather than on a detailed comparison of the original and simplified models. Consequently, quantitative measures of model performance were not reported apart from fitness measures for one of the models. It is, therefore, difficult to assess the extent to which the models had been simplified or the effect of the simplifications on their performance.

Costanza and Sklar [56], by contrast, did carry out a quantitative comparison of different models. They compared 87 freshwater wetland ecosystem models using measures termed articulation and descriptive accuracy. Diminishing returns indices were calculated for the number of components, time steps, and spatial units in the model (with each index having a different scaling factor). The average of the three indices was calculated for the data and for the model and the minimum of these two numbers used for the articulation measure. This was, therefore, a measure of the scope and complexity of the problem (i.e., of the experimental frame). The descriptive accuracy index was a validity measure of the fit of the model data against the actual historical data. They were able to calculate both of these measures for 26 of the models and they also calculated a combined articulation and accuracy measure called effectiveness. They found that the models with the highest descriptive accuracy had low articulation (although the majority of the 26 models had low articulation), i.e., the models with the greatest validity tended to be those addressing the simpler problems. It is difficult to draw absolute conclusions from this result as the amount of data was relatively small but Costanza and Sklar hypothesised that, ultimately, greater articulation necessitates less accuracy, and that there might be a level of articulation which maximises effectiveness (which they considered to be the best model). This assumes that greater articulation is desirable, in the sense that a model with greater articulation provides more information about the system. Often, however, the modelling objectives are quite specific and only greater information relevant to the problem (i.e., within the experimental frame) is a benefit.

Webster *et al.* [39] viewed the selection of the level of detail as part of the validation process, and so the only measure reported was the goodness of fit against actual data for the alternative timber harvesting models that they compared. They considered the appropriate level of detail to be the simplest model of adequate validity which is consistent with the expected system relationships (ignoring the accuracy of results which can only be assessed subsequently). Their discussion of adequacy suggests including performance measures other than just validity in assessing adequacy, although their approach implies that models are either valid or not valid, whereas in reality there are degrees of validity. They used three alternative methods to generate sample data for three input variables in a simulation model (giving 27 alternatives in all): mean value, regression, and a histogram of actual data. For one of the variables they found that the histogram method (which they considered the most complex level) gave output of lower validity than the simpler methods. Four of the models gave adequate validity and so they chose the simplest of these as the final model.

Halfon's studies [43,44] compared the structure of alternative models of a toxic substance in a lake at six levels of detail. This was conducted for a model with six state variables and for a model with ten state variables, and repeated in each case with and without internal recycling (giving four sets of results). He compared the structures of the models, mainly using Bosserman's [45] $\bar{c}$ measure (described earlier) which was applied to the graphs of interactions between state variables. The level of detail of the models was increased by adding the physical processes in stages in a logical order. He found that, in each case, adding the last few levels of detail only caused a small increase in the number of connections. He argued that it was not worth including these processes as they were unlikely to affect model behaviour significantly and the additional parameters would add to the amount of uncertainty in the model. It is reasonable to expect diminishing returns as complexity is added. However, the actual performance of the models was not assessed to confirm this. Halfon [44] also suggested displaying the comparisons of alternative model structures as a Hasse diagram.

The lack of studies which have specifically sought to examine the effects of level of detail or complexity on model performance means that even if the expected relationships described at the beginning of this section are generally true, the nature of the relationships are unclear (linear, increasing returns, decreasing returns, etc.), and the circumstances in which the relationships break down are not understood. The particular elements of model complexity which have the greatest effect on each performance element have also not been identified.

Consider, for example, the accuracy of model results. Generally a more complex model is expected to be more accurate and as the model becomes more complex the increase in accuracy of adding further complexity is likely to reduce (assuming that additional detail is added in order of relevance), i.e., decreasing returns. Certainly, if there is a mapping between the models so that the more complex model can be reduced to the simpler model by a suitable choice of parameters, then the most accurate complex model must be at least as accurate as the most accurate simple model. However, the choice is often between models of different types or between models for which only an approximate relationship exists. In this case, it is possible for the simpler model to be more accurate, although a comparison of the level of detail of the models is more difficult. For example, empirical models are sometimes more accurate than quasi-physically based models which would generally be considered to be more complex [57]. For some modelling (such as physically based distributed parameter models), the input parameters cannot be directly measured, but must be inferred by calibrating the model against historical data [58]. This is called the inverse problem of parameter identification and its nature means that there may be a wide range of parameter values that give a good fit. In this case, the results of a model should be a range of predictions rather than a single prediction [59], and a more complex model may give a wider range. The range will depend on the number of parameters in the model and the extent to which they are allowed to vary (i.e., the size of the parameter space).

There may also be occasions when a simpler model takes longer to build. Garfinkel [60] pointed out that in modelling a large system which is considered to consist of a large number of subsystems there is a much greater choice of simple models (which just model a few subsystems thought to be important) than complex models, and it may take longer to choose between the alternative simple models than it would have taken to build a model of the whole system.

Also, a simple model, by incorporating only some of the system elements, may allow the identification of system relationships which are obscured in a more complex model and so may give a greater understanding of the system. On the other hand, a complex model may extend understanding by allowing the investigation of the effect on the system of many more factors. We found that, in the case of the poppy population genetics model, the process of identifying, building and comparing models at different levels of detail greatly increased our understanding of the system and led to the development of a new mathematical theory. Such a process could be used to link simple strategic models that are difficult to verify with more detailed tactical models which can be tested against available data [61]. If the main purpose of the study is gaining an understanding of the system then the benefits of building models at several levels of detail may be well worth the additional effort involved.

The relationship between the level of detail or complexity of a model and model performance is, therefore, more complicated than many of the comments in the literature would suggest, and the lack of studies in this area means that the relationship is poorly understood. What is required is a number of studies that measure the elements of the complexity and performance of alternative models. This would provide data from which to develop empirical relationships and may lead to a theoretical basis for the relationships. This is similar to the aim in computer science of developing software metrics to control and predict the performance of software projects [62]. Attempts have been made to predict the resources required for and the likely number of errors in a piece of software from particular software attributes (such as "complexity"). Fairly strong relationships have been found within particular environments (for example, by Boehm [63]) although none of these appear to be generally applicable. The development of metrics is still at a reasonably early stage and they do not yet appear to be widely used in software projects. The production of a piece of software is often one step in the modelling process (the model construction step), and so the identification of relationships between attributes and performance is likely to be harder for models than for software. It is very unlikely that universal relationships exist since the great diversity of model types is likely to contain an exception to any rule. However the assumed relationships identified at the beginning of this section indicate the probable existence of general relationships at least for particular types of models or for particular circumstances. A common piece of advice in choosing the level of detail is to use past experience and so, at the very least, the quantitative comparison of alternative models would provide a source of modelling experience from which to draw.

6. SIMPLIFICATION AND OTHER RELATED AREAS

The selection of the best model requires not just an appreciation of the likely performance of each model but also a knowledge of the possible alternative models. There are very many models that could be built in most cases and so the best model may not even be identified as a possible model. Structuring the models by level of detail can help in the search for better models, and one way of identifying new models is to take an existing model and then attempt to simplify it [64]. Zeigler [12] set out four categories of simplification methods: dropping unimportant parts of the model, replacing part of the model by a random variable, coarsening the range of values taken by a variable, and grouping parts of the model together. The simplified model using these methods will be of the same type as the original; it is also possible to replace part of a model with a model of a different type such as analysing the inputs and outputs of the particular part and replacing it with a regression equation, analytical equation or neural network (if the original part was very complex). Sevinc [65] developed a semiautomatic simplification program for discrete event simulation models based on Zeigler's [12] DEVS model formalism and simplification ideas. Innis and Rexstad [55] listed and described 17 simplification techniques. These are specific techniques, some of which fall under Zeigler's [12] categories, as well as techniques for replacing part of the model with a different type. Innis and Rexstad also included techniques for identifying which parts of the model might be suitable for simplification, techniques for reducing the number of model runs or run times and techniques for improving the readability of the model code. They stated that their list was not exhaustive, and it would appear that a general simplification methodology does not exist.

Zeigler's [12] DEVS model formalism provides a framework within which alternative discrete event simulation models can be compared. Addanki *et al.* [66] proposed representing the alternative models as nodes on a graph with the edges representing the changes in assumptions from one model to another. Moving around the graph is an alternative way of searching the space of models to which Addanki *et al.* applied artificial intelligence techniques. An approach applied to engineering models has been to generate a database of model fragments and then to automate the process of selecting and combining the fragments to produce the model [67–69]. Developments have also taken place recently in variable resolution modelling which allows the level of detail of the model to be changed easily even while the model is running (e.g., [70]) and this may be a suitable environment within which to investigate the effect of level of detail.

7. CONCLUSIONS

The lack of research into the process of choosing the best model is surprising given the importance of modelling in science. There are very few studies that have made any quantitative assessment of the effect of different model attributes on the modelling process. This probably stems from the difficulty in measuring either suitable attributes or model performance, and also the effort required to build several alternative models. Different models are most often compared by their level of detail or complexity although such a comparison is usually only qualitative. This has resulted in only vague guidelines to aid the choice of the level of detail and so a greater understanding of its effect would be very beneficial. The initial requirement is for a considerable number of studies that compare, preferably quantitatively, some aspect of model performance with appropriate model attributes, such as level of detail or complexity. Such an approach has other benefits as we have found that it can greatly increase the understanding of the system being studied. The alternative models will, at least, partially verify each other [64]. Ultimately this approach should lead to the development of general principles and hopefully to a methodology for choosing the best model. A corresponding methodology for simplification is also necessary.

REFERENCES

1. J. Fripp, How effective are models?, *Omega* **13** (1), 19–28 (1985).
2. J.N.R. Jeffers, From free-hand curves to chaos: Computer modelling in ecology, In *Computer Modelling in the Environmental Sciences*, (Edited by D.G. Farmer and M.J. Rycroft), pp. 299–308, The Institute of Mathematics and its Applications Conference Series No. 28, Clarendon Press, Oxford, (1991).
3. A.M. Tobias, Verification, validation and experimentation with visual interactive simulation models, In *Operational Research Tutorial Papers*, The Operational Research Society, (1991).
4. P.-J. Courtois, On time and space decomposition of complex structures, *Communications of the ACM* **28** (6), 590–603 (1985).
5. A.M. Law, Simulation model's level of detail determines effectiveness, *Industrial Engineering* **23** (10), 16–18 (1991).
6. C.B. Tilanus, Failures and successes of quantitative methods in management, *European Journal of Operational Research* **19**, 170–175 (1985).
7. S.C. Ward, Arguments for constructively simple models, *Journal of Operational Research Society* **40** (2), 141–153 (1989).
8. J.D. Salt, Keynote address: Simulation should be easy and fun!, In *Proceedings of the 1993 Winter Simulation Conference*, (Edited by G.W. Evans *et al.*), pp. 1–5, IEEE, New York, (1993).
9. A.M. Law, J.S. Carson, K.J. Musselman, J.G. Fox, S.K. Halladin and O.M. Ulgen, A forum on crucial issues in the simulation of manufacturing systems, In *Proceedings of the 1993 Winter Simulation Conference*, (Edited by G.W. Evans *et al.*), pp. 916–922, IEEE, New York, (1993).
10. J.C.R. Hunt, Presidential address: Contributions of mathematics to the solution of industrial and environmental problems, *IMA Bulletin* **30**, 35–45 (1994).
11. A.J. Jakeman and G.M. Hornberger, How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research* **29**, 2637–2649 (1993).

12. B.P. Zeigler, *Theory of Modelling and Simulation*, John Wiley, New York, (1976).
13. S.I. Gass, What is a computer-based mathematical model?, *Mathl. Modelling* **4** (5), 467–472 (1983).
14. E. Rexstad, and G.S. Innis, Model simplification—Three applications, *Ecological Modelling* **27** (1/2), 1–13 (1985).
15. R.E. Shannon, *Systems Simulation: The Art and Science*, Prentice Hall, Englewood Cliffs, NJ, (1975).
16. J. Banks and J.S. Carson, *Discrete-Event System Simulation*, Prentice Hall, Englewood Cliffs, NJ, (1984).
17. C.O. Stockle, Canopy photosynthesis and transpiration estimates using radiation interception models with different levels of detail, *Ecological Modelling* **60** (1), 31–44 (1992).
18. A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, (1991).
19. M. Bunge, *The Myth of Simplicity: Problems of Scientific Philosophy*, Prentice Hall, Englewood Cliffs, NJ, (1963).
20. J.L. Casti, *Connectivity, Complexity, and Catastrophe in Large-Scale Systems*, John Wiley and Sons, New York, (1979).
21. R.L. Henneman and W.B. Rouse, On measuring the complexity of monitoring and controlling large-scale systems, *IEEE Transactions on Systems, Man and Cybernetics* **SMC-16**, 193–207 (1986).
22. M. Gell-Mann, What is complexity?, *Complexity* **1** (1), 16–19 (1995).
23. W. Weaver, Science and complexity, *American Scientist* **36** (Autumn), 536–544 (1948).
24. G.J. Chaitin, Randomness and mathematical proof, *Scientific American* **232** (May), 47–52 (1975).
25. H. Morowitz, The emergence of complexity, *Complexity* **1** (1), 4–5 (1995).
26. M. Flatau, Review Article: When order is no longer order—Organising and the new science of complexity, *Organization* **2** (3/4), 566–575 (1995).
27. R.L. Flood and E.R. Carson, *Dealing with Complexity: An Introduction to the Theory and Application of Systems Science*, 2^{nd} edition, Plenum Press, New York, (1993).
28. M.W. Golay, P.H. Seong and V.P. Manno, A measure of the difficulty of system diagnosis and its relationship to complexity, *International Journal of General Systems* **16** (1), 1–23 (1989).
29. H.A. Simon, The architecture of complexity, *General Systems Yearbook* **10**, 63–76 (1964).
30. R. Rosen, Complexity as a system property, *International Journal of General Systems* **3**, 227–232 (1977).
31. L. George, Tests for system complexity, *International Journal of General Systems* **3**, 253–258 (1977).
32. B.P. Zeigler, Multifaceted modeling methodology: Grappling with the irreducible complexity of systems, *Behavioral Science* **29**, 169–178 (1984).
33. L. Schruben and E. Yücesan, Complexity of simulation models: A graph theoretic approach, In *Proceedings of the 1993 Winter Simulation Conference*, (Edited by G.W. Evans *et al.*), pp. 641–649, IEEE, New York, (1993).
34. G. Blöschl and R. Kirnbauer, Point snowmelt models with different degrees of complexity—Internal processes, *Journal of Hydrology* **129**, 127–147 (1991).
35. J.A. le Roux and M.S. Potgieter, The simulation of Forbush decreases with time-dependent cosmic-ray modulation models of varying complexity, *Astronomy and Astrophysics* **243**, 531–545 (1991).
36. B.O. Palsson and I. Lee, Model complexity has a significant effect on the numerical value and interpretation of metabolic sensitivity coefficients, *Journal of Theoretical Biology* **161**, 299–315 (1993).
37. W.K. Durfee, Control of standing and gait using electrical stimulation: Influence of muscle model complexity on control strategy, *Progress in Brain Research* **97**, 369–381 (1993).
38. D.E. Smith and J.M. Starkey, Effects of model complexity on the performance of automated vehicle steering controllers: Model development, validation and comparison, *Vehicle System Dynamics* **24**, 163–181 (1995).
39. D.B. Webster, M.L. Padgett, G.S. Hines and D.L. Sirois, Determining the level of detail in a simulation model—A case study, *Computers and Industrial Engineering* **8** (3/4), 215–225 (1984).
40. T.J. McCabe, A complexity measure, *IEEE Transactions on Software Engineering* **2** (4), 308–320 (1976).
41. L. Schruben, Simulation modelling with event graphs, *Communications of the ACM* **26** (11), 957–963 (1983).
42. M. Pidd, *Computer Simulation in Management Science*, John Wiley and Sons, Chichester, (1988).
43. E. Halfon, Is there a best model structure? I. Modelling the fate of a toxic substance in a lake, *Ecological Modelling* **20**, 135–152 (1983).
44. E. Halfon, Is there a best model structure? II. Comparing the model structures of different fate models, *Ecological Modelling* **20**, 153–163 (1983).
45. R.W. Bosserman, Structural comparison for four lake ecosystem models, In *A General Survey of Systems Methodology. Proceedings of the Twenty-sixth Annual Meeting of the Society for General Systems Research*, (Edited by L. Troncale), pp. 559–568, Washington, DC, (1982).
46. J.D. Farmer, A rosetta stone for connectionism, *Physica D* **42**, 153–187 (1990).
47. R.J. Brooks, A.M. Tobias and M.J. Lawrence, The population genetics of the self-incompatibility polymorphism in *Papaver Rhoeas*. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of *S*-allele frequencies in populations at equilibrium, *Heredity* **76**, 367–376 (1996).
48. C.G. Langton, Computation at the edge of chaos: Phase transitions and emergent computation, *Physica D* **42**, 12–37 (1990).
49. S. Henry and D. Kafura, Software quality metrics based on information flow, *IEEE Transactions on Software Engineering* **7** (5), 510–518 (1981).
50. N.E. Fenton, *Software Metrics: A Rigorous Approach*, Chapman and Hall, London, (1991).

51. W.S. Meisel and D.C. Collins, Repro-modelling: An approach to efficient model utilization and interpretation, *IEEE Transactions on Systems, Man and Cybernetics* **SMC-3**, 349–358 (1973).
52. S.L.S. Jacoby and J.S. Kowalik, *Mathematical Modeling with Computers*, Prentice Hall, Englewood Cliffs, NJ, (1980).
53. R.N. Palmer and J.L. Cohan, Complexity in Columbia River systems modeling, *Journal of Water Resources Planning and Management* **112**, 453–468 (1986).
54. P.A. Fishwick, The role of process abstraction in simulation, *IEEE Transactions on Systems, Man and Cybernetics* **18** (1), 19–39 (1988).
55. G.S. Innis and E. Rexstad, Simulation model simplification techniques, *Simulation* **41** (1), 7–15 (1983).
56. R. Costanza and F.H. Sklar, Articulation, accuracy and effectiveness of mathematical models: A review of freshwater wetland applications, *Ecological Modelling* **27** (1/2), 45–68 (1985).
57. D.G. Decoursey, Developing models with more detail: Do more algorithms give more truth?, *Weed Technology* **6** (3), 709–715 (1992).
58. H. Allison, Inverse unstable problems and some of their applications, *Mathematical Scientist* **4**, 9–30 (1979).
59. R.J. Brooks, D.N. Lerner and A.M. Tobias, Determining a range of predictions of a groundwater model which arise from alternative calibrations, *Water Resources Research* **30** (11), 2993–3000 (1994).
60. D. Garfinkel, Modelling of inherently complex biological systems: Problems, strategies, and methods, *Mathematical Biosciences* **72** (2), 131–139 (1984).
61. W.W. Murdoch, E. McCauley, R.M. Nisbet, W.S.C. Gurney and A.M. De Roos, Individual-based models: Combining testability and generality, In *Individual-based Models and Approaches in Ecology: Populations, Communities and Ecosystems*, (Edited by D.L. DeAngelis and L.J. Gross), pp. 18–35, Chapman and Hall, New York, (1992).
62. T. Demarco, *Controlling Software Projects: Management, Measurement and Estimation*, Yourdon Press, New York, (1982).
63. B.W. Boehm, *Software Engineering Economics*, Prentice Hall, Englewood Cliffs, NJ, (1981).
64. B.P. Zeigler, Multilevel multiformalism modeling: An ecosystem example, In *Theoretical Systems Ecology: Advances and Case Studies*, (Edited by E. Halfon), pp. 17–54, Academic Press, New York, (1979).
65. S. Sevinc, Automation of simplification in discrete event modelling and simulation, *International Journal of General Systems* **18** (2), 125–142 (1990).
66. S. Addanki, R. Cremonini and J.S. Penberthy, Graphs of models, *Artificial Intelligence* **51**, 145–177 (1991).
67. B. Falkenheimer and K.D. Forbus, Compositional modelling: Finding the right model for the job, *Artificial Intelligence* **51**, 95–143 (1991).
68. P.P. Nayak, Automated modelling of physical systems, Ph.D. Thesis, Computer Science Department, Stanford University, Technical Report STAN-CS-92-1443, (1992).
69. T.R. Gruber, Model formulation as a problem solving task—Computer-assisted engineering modelling, *International Journal of Intelligent Systems* **8** (1), 105–127 (1993).
70. P.K. Davis and R. Hillestad, Families of models that cross levels of resolution: Issues for design, calibration and management, In *Proceedings of the 1993 Winter Simulation Conference*, (Edited by G.W. Evans, M. Mollaghasemi, E.C. Russell and W.E. Biles), pp. 1003–1012, IEEE, New York, (1993).