CHAPTER 7

Environmental Simulation Modelling

There are a number of quite different types of models, as well as several approaches to model building. It will be useful to bring out some important distinctions and to discuss which of the techniques are likely to be the most appropriate in modelling the complex issues of the environment. Not all readers will need to understand the technicalities, but it is desirable that they gain at least some appreciation of the major issues of methodology, such as those discussed in Section 7.3 below. Some of the main points are illustrated by specific case studies in Chapter 9.

7.1 THE NATURE OF SIMULATION MODELLING

Theoretical modelling of the real world is as old as science itself. With the advent of the computer, however, the technique has been transformed. Where, previously, models had to be soluble by analytical means, it suddenly became possible to solve very large systems of equations of a much more general form. In some cases, the necessary level of mathematical expertise was greatly reduced, since the process of solution by simulating the system on a computer required the investigator to have only sufficient mathematics to formulate and programme the equations, and not to solve them analytically.

Simulation may, in the present context, be defined as the use of the computer to trace the lengthy chains of indirect repercussions in the cause-and-effect relations describing some complex system, and thus to imitate the dynamic behaviour of the system. The ability of the computer automatically to keep track of the interactions between numerous variables, each of which can affect, and simultaneously be affected by, the others, is a strong point of simulation-modelling technique.

The terms 'modelling' and 'simulation' are often used interchangeably, although a recent trend is to reserve the first for the process of developing a model and the second for its manipulation and use. In the present report, we have used the two terms in combination to cover all phases of model development and application, as well as to distinguish this type of modelling from purely mathematical modelling.

Until about 1960, the biggest influence in the development of this newer style of simulation modelling came from the analogue computer, so called because its operation is directly analogous to that of the real system. In more recent times the digital computer has become fast enough to be used for numerical solutions in a way that is conceptually a close parallel to the analogue mode. Hybrid computers, which aim at combining some of the best features of the analogue and digital types, have also been developed, but it would not be appropriate to go into the details here. In any case, the digital computer has, because of its many other uses, become so readily available that it is likely to be the type most used for environmental studies in the foreseeable future.

Not only has the computer provided an enormous stimulus to modelling but the reverse is also true: the desire to model ever more complex systems has been a major incentive in the development of ever more powerful computers, as well as of more powerful and flexible techniques for computer programming and numerical analysis. This mutual reinforcement of computing facilities and modelling techniques is likely to continue in the future.

Besides the *continuous* simulation techniques discussed so far, two other techniques of fundamental importance in simulation modelling have been made possible by the modern digital computer – *discrete-event* simulation and *stochastic* simulation.

For situations where the system being studied contains numbers of separate items, each having its own characteristics and period of existence within the system, the technique of discrete-event simulation is often appropriate. In this, changes in the state of the system are conceptualized as taking place in discrete jumps corresponding to the arrivals, departures, or other critical changes in status of the individual items. This approach could, for instance, appropriately be used in modelling demographic processes in a small population where each birth and death was to be noted separately. This contrasts with the first type of simulation technique where the changes are *thought of* as taking place in a continuous manner (although in a digital computer they may be programmed in terms of discrete equations).

Where the number of separate items is sufficiently large, and only their average or group behaviour is of interest, a continuous simulation model may still be appropriate. But where it is essential to follow the fortunes of each item and preserve its individuality the discrete-event form is usually necessary. So far, continuous simulation models have been the choice in the majority of environmental applications, but more use of discrete-event simulation can be expected in the future.

Discrete-event simulation owes much less to any prior developments in mathematical modelling than does continuous simulation, and is very largely an outgrowth of developments in computer programming, with some ideas also coming from critical-path network techniques such as the Program Evaluation Review Technique (PERT) (Shore, 1973). Particular programming techniques that are exploited in discrete-event simulation include list processing, which allows easy handling of lists that must be frequently amended, and dynamic storage allocation, which facilitates the storage of data in cases where it is not possible to predict in advance how much computer storage will be taken up by each category of data, and the space needed may change during the course of the simulation.

Where the response of the system to changed inputs is subject to random fluctuation (noise), or where two variables are only moderately correlated (the exact relationship being unknown), various types of Monte Carlo or stochastic simulation are possible. These utilize the ability of the computer to generate pseudo-random numbers. In other words, it is possible for the computer to carry out the equivalent of tossing a coin or drawing numbers from a hat.

The three major types of simulation that have been described, continuous, discrete, and stochastic, can of course be used in combination; an example would

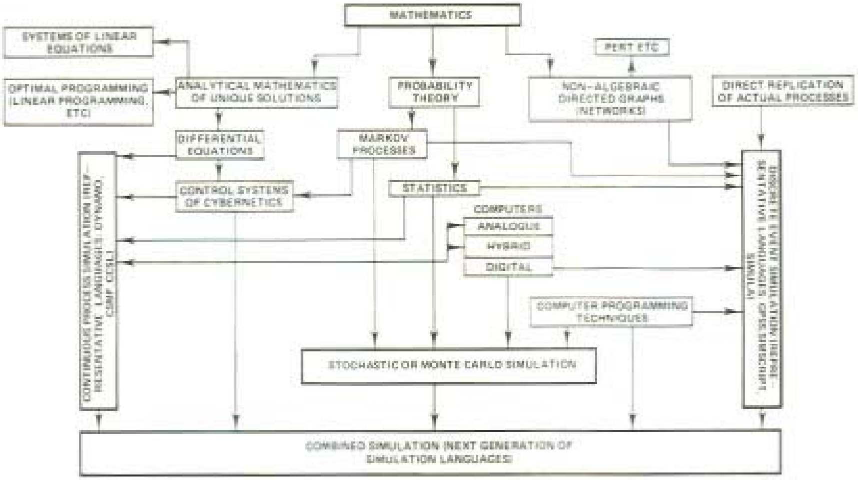

Figure 7.1 Relationships between the major simulation techniques

be the modelling of atmospheric pollution as a continuous process, but depending on the sudden commencement or cessation of discrete sources, and subject to the stochastic nature of atmospheric turbulence. Figure 7.1 illustrates the relationships between these types of simulation approach and shows the three distinct lines of development that have led to them. Continuous simulation evolved from analytical and analogue methods to the point where systems of considerable complexity are modelled digitally. Elaborate graphical and pictorial displays of the dynamic behaviour of such systems are now possible. The discrete-event methodology may be considered as combining ideas from directed graph (or network) theory and the mathematical study of Markov processes. As shown in Figure 7.1, discrete-event simulation originated independently of the other two methodologies and their underlying mathematics, and derives historically from attempts to imitate directly various real-life processes.

A number of specialized computer languages (shown in Figure 7.1) have been developed to facilitate the programming of continuous or discrete-event models. Although these are not essential (many models are still programmed using general-purpose languages such as FORTRAN), they make it possible to develop programmes easy for others to follow. This latter point is quite important; the difficulties in unravelling someone else's computer programme are such that modellers sometimes find it easier to start again from scratch. Unfortunately, no one simulation language is yet generally used and accepted. The widespread acceptance and availability of a suitable simulation language would assist modellers everywhere in exchanging ideas and building on each other's work.

The general principles of model design do not result in unique model structures and, needless to say, do not eliminate the need for the types of substantive knowledge that are provided by individual disciplines. Moreover, each of the three types of simulation model implies a different view of the world and how it is structured. Accordingly, each employs somewhat different principles of model design and requires different types of data.

It is possible therefore that equally valid models may differ substantially in structure, data, and simulation technique, not to mention differences in programming language, and yet produce similar representation of the behaviour patterns of the system under study. This makes the task of evaluating simulation models relatively difficult, particularly since valid models may well display patterns of behaviour which differ from any that have been historically observed. This is natural if one keeps in mind that the purpose of simulation experiment is not only forecasting, but the display of possible behaviour modes under different conditions – including conditions that differ from the historical ones. As an example, it is possible to simulate world climate on the assumption that Australia is covered with forest, or that the earth rotates in the opposite direction.

One further distinction regarding types of model needs to be made. The term 'model' is often applied rather loosely to mathematical procedures which are intended not only to reflect the real world, but also to identify the best way of attaining some specified goal – optimization studies using methods such as linear programming. In these, however, much of what is called the model is really a specification of the process of optimization – in particular the constraints under which it is to be performed – the model proper often being of rather a simple form. The optimizing models certainly have their uses but should not be confused with simulation models.

Of course, simulation models can be used in optimization studies. However, with many environmental issues the most urgent requirement will be to obtain a model that provides good insight into the general working of, and interactions between, the different parts of the environmental system. By the time such a model has been developed and such insight gained, the main actions that need to be taken will often have become obvious, so that formal optimization studies will be either unnecessary or concerned merely with minor improvements in some basic strategy. It is for these reasons that the main emphasis of the present report is placed on simulation rather than optimizing models.

Some types of computer equipment and organization lend themselves particularly well to interactions between computer and user. For instance, it may be possible to arrange model input and output so that a user can sit at a computer terminal, insert instructions (perhaps as a response to enquiries generated by the computer programme and displayed on a screen), and see the results produced by the model in a graphical form – for instance, as a time plot of critical variables. He could then insert alternative instructions and compare the results. This type of man-machine interaction at the output end of the model can make its use far more effective and greatly encourage its application in the development of strategies for environmental management.

At the input end, too, interactive facilities can be useful if, for instance, a model needs to make use of a variety of data. These data can then be inserted at the time the model programme is executed; or, alternatively, they can be extracted from an existing data bank if the interactive programmes have been organized to permit this.

If simulation systems operating interactively are to provide a basis for making sensible management and planning decisions, their structure should also define the interdisciplinary research programmes which provide their data base. This suggests that studies both of the external environment and of processes taking place within society might constitute a single complex, internally linked through a common information base, a common language, and a common set of principles governing its utilization. Such a degree of integration, however, lies well in the future.

For the future, too, some modellers envisage the development of libraries of generalized models available on an interactive basis within a computer network. Models of decomposition and of primary and secondary production, for instance, could be integrated with models of other biological and physical processes to produce a complete description of how the sun's energy drives the entire environmental system as we know it; these environmental models could be linked with models of human economic and social activities, to enable one to study both the stresses and perturbations of the environment due to man, and the effects of environmental change on human activity. Until recently, economic models have emphasized the internal dynamics of organized economic systems by studying market processes and money flow rather than interactions with nature and the factors that govern the real flows in the system. There have recently been attempts at using simulation in order to consider the economy in this wider sense, and it may be that such attempts will improve economic modelling in the future.

The approach to environmental simulation modelling outlined in the previous paragraphs should not be considered as generally applicable – or even as representing consensus. All would probably agree with the need to integrate models of the environment with socio-economic models of human affairs affecting it and affected by it. But opinions differ on the value for specific problems of environmental models which are highly general. In such problems, the objectives of the study will tend to indicate the specific nature of the model which seems most appropriate under the circumstances: indeed it could be argued that one of the principal tasks of the systems analyst is to choose a model which best suits the nature of the problem at hand. Certainly, the history of simulation modelling so far suggests that the search for universally applicable models may be a vain one which, if pursued, can lead to inefficiency and increase the probability of failure.

7.2 CLASSIFICATION OF ENVIRONMENTAL SIMULATION MODELS

Apart from the possible varieties of approach in the technical construction of models mentioned in the previous section, models may also vary greatly in subject matter, in scale, in the purpose for which they are to be used, and in many other ways. There is thus a need for some form of classification in order to improve communication among modellers, between the modeller and the rest of the scientific community, and between the modeller and the decision-maker.

The requirements for a classification of environmental simulation models vary from one situation to another. Appendix 2 gives a detailed list of attributes that specify an environmental simulation model in different respects. From these attributes a subset can be defined in most cases to meet the classification needs of a special situation.

7.3 APPROACHES TO MODEL BUILDING: THE PROBLEM OF COMPLEXITY

Having decided on the type of model which appears best to suit the nature of the study objectives, it is necessary to consider how an appropriate model *structure* may be defined or *identified*, and then how the coefficients or parameters that characterize the model within this structure may be *estimated*. Finally, the modeller must attempt to validate his model – that is, in some way to test the reliability of the model as a guide to the behaviour of the real system. The word 'validation' has been used to designate a variety of procedures (see below, Section 8.1.5), but in the strict sense it implies the statistical comparison of model output with independent observation of the actual system, leading in principle to the definition of confidence limits.

However validation is attempted, the modeller is never completely sure that his model is correct and should indicate to the decision-maker the inherent uncertainty of any information derived from the model. In addition, it must be remembered that the model has only been validated for a limited range of data. Thus, if the model is to be used for forecasting the future, it may prove inadequate if, for some reason, the real system changes its characteristics in the future and behaves in a radically different manner. The system is then said to be non-stationary. For example, if a model for water pollution in a river is validated prior to major urban and industrial development, it may not provide a good indicator of pollution behaviour after this development has taken place. Similarly, if an environmental model is validated under certain climatic conditions, it may not be fully representative if these conditions should change.

It is indeed possible to build good models of such non-stationary systems, which will cover changing conditions satisfactorily; but the fact that a model has been validated previously is no guarantee in itself that it will still be valid if conditions change. Validation should therefore be regarded as a continuing need for a model in constant use, and the model should be updated whenever any built-in assumptions cease to apply.

Model identification, estimation, and validation, while essential steps in model building, can prove difficult to accomplish in practice. Environmental systems tend to be badly defined, in the sense that the mechanisms which govern the changes in the variables defining the system, and their relationships to one another, are often not well understood. Theoretical analysis of the system coupled with observations of its behaviour can help to define the inherent mechanisms better. However, ambiguity often still remains: a situation may occur in which a number of possible explanations remain feasible, but little, if any, evidence exists as to which of these explanations is the most plausible.

If it is possible to perform planned experiments on the system, then this ambiguity can often be removed by attempting to excite that mode of system behaviour which will best help to define the mechanisms involved; and since in such cases the experiment is planned, it may well be possible to obtain observations of the behaviour of the system that are relatively free of uncertainty. When dealing with larger environmental systems, however, the possibility of performing such planned experiments is usually remote, and data obtained during the normal operation of the system, even if available, are likely to be scarce. In the case of urban air pollution models, for example, investigators have exploited the occurrence of industrial strikes (and consequent reduction of point-source emissions) to help estimate and validate their models. In one particular instance, the operators of a power station in Chicago were even permitted to increase their emissions briefly for the purpose of model validation. But such opportunities are rare.

Perhaps the most popular approach to modelling in these circumstances consists of analysing the system in a fairly detailed manner, so that the resulting equations are characterized by parameters with some physical (or biological or social) significance to the model builder, as in the Hydrocomp Simulation Programme (Linsley, 1976) or in many ecological models (e.g. Goodall, 1975; Miller et al., 1975). These physical parameters (such as diffusion rate, sedimentation, and soil infiltration coefficients) are evaluated, possibly by experiment in a laboratory, and then inserted into the model at their appropriate positions. Realizing some of the difficulties involved in such a procedure, the modeller usually *calibrates* his model by comparing its output with whatever observed data are available, and 'tunes' his parameter within physically meaningful bounds until some reasonable fit to the data is obtained. This tuning of the model parameters, which may be accomplished either manually or by systematic optimization, can be quite a difficult task, since models of this type are usually characterized by a large number of unknown parameters. The number of parameters that can be tuned will, of course, be limited by the number of data points available. Each model parameter which is estimated from a set of data reduces the value of those data for validation purposes. Technically speaking, the size of the data set determines the total number of degrees of freedom, some of which are needed for parameter estimation; the larger the number of degrees of freedom left for testing model output, the more satisfactorily can the model be validated.

One method of alleviating these difficulties in an *ad hoc* way is for the model builder to use his physical intuition to adjust some of the parameters, with the remainder fixed at their assumed values. An alternative and more systematic approach is to test the sensitivity of the system to changes in the assumed values of parameters (sensitivity analysis), and then to reduce the *effective* number of unknown parameters by only considering as unknown those parameters which most affect the model response and thus need to be known with greatest accuracy. Another important way of reducing the number of parameters to be estimated simultaneously – that of decomposing the large model into a number of smaller parts for validation – will be discussed when considering alternatives to the large-model approach.

This detailed, mechanistically-based approach generally leads to a complex model which has a rich repertoire of behaviour, and we might refer to it as a '*comprehensive*' model. Such a model has several advantages:

(a) It enables good use to be made of any prior information about the internal structure and mechanisms of the system being modelled. In some cases this prior knowledge could be extensive and include some of the 'hardest' data available about the system.

(b) It makes it easy for the various specialists to incorporate their knowledge of the behaviour of the particular components with which they are familiar. Where their knowledge is based on personal observations, collected over a long period under a wide variety of circumstances, this facility could be most valuable.

(c) It provides a convenient vehicle of communication between the experts, and a means of summarizing all of their relevant knowledge in a systematic form. Because of this feature, it may be that, even where the 'comprehensive' model must be rejected as the final decision-making tool, it is nevertheless worth formulating for use as a starting point in developing the various simpler and specialized models that are actually to be used.

These considerable advantages of the 'comprehensive' model are, unfortunately, matched by disadvantages:

(a) It may be difficult, if not impossible to obtain the large amount of data, covering a wide range of circumstances, which are needed to validate such a detailed model.

(b) Even where this might be possible, the effort required could be so high as to be out of proportion to the likely benefit from the final model.

(c) Where the model is only partially validated, it may not be possible to provide good statistical information on the uncertainty associated with all of the parameter estimates. This fact, coupled with the large number of parameters of the 'comprehensive' model, leaves it open to the charge that it could contain hidden and unvalidated modes of behaviour that may not be characteristic of the real system.

An alternative technique which tends to the opposite extreme of simplicty might be termed the 'dominant-modes' approach. Here, an attempt is made to model only those primary or dominant modes of behaviour that characterize the observed response of the system. This approach, which is at present much less popular than 'comprehensive' modelling (perhaps because the necessary techniques are less well known), has been used mainly by model builders with a background in control theory and systems engineering. Problems of automatic control can often be greatly simplified by recognizing that perturbation of the system about some static or slowly changing equilibrium state leads to a response whose main characteristics can be described in rather simple terms; control of these main characteristics (*dominant modes* – Shamash [1975]) of the response is the main task of the control-system designer. Since management and planning are forms of feedback control, it is clear that such observations are relevant to models that are to be used for decision-making.

The usual approach to the construction of 'dominant-mode' models entails reducing the initial model based on the mechanisms of the system to its simplest viable form and then, by an iterative process of parameter estimation and model evaluation, successively building additional features into the model until it is satisfactory in some statistical sense (Young, 1977). Because the models identified in this manner normally have rather few parameters, it is feasible to employ relatively sophisticated statistical methods of parameter identification and estimation. These methods yield information not only on the statistical nature of the parameter estimates, but also on the stochastic disturbances that inherently characterize any badly defined systems, such as those met in the environmental context. Additional information of this type can be particularly useful, for example, in providing confidence bounds on model predictions; it also allows for the possibilities of performing stochastic simulations (Monte-Carlo analysis) and considering, in a rigorous manner, the problems of decision-making in the presence of the uncertainty which arises when random fluctuations ('noise') are superimposed on the effects of interest.

In addition to requiring a reasonable data base, the major disadvantage of the 'dominant-modes' approach to model building is that the modeller gains little insight into possible hidden modes of behaviour that might become important in the future. For this reason, this kind of model has its principal application in the 'adaptive' or 'incremental' approach to decision-making which is discussed elsewhere in this report (pages 23 *et seq.*).

The 'comprehensive' and 'dominant-modes' approaches to modelling are clearly not mutually exclusive; indeed, while there are few existing examples, it may be anticipated that they will eventually be used to complement each other, with the advantages of one compensating for the disadvantages of the other. Furthermore, they are not the only approaches to environmental simulation; there are various other techniques that tend to incorporate, to a greater or lesser degree, aspects of both.

In the case of very large systems, it is often convenient to decompose the overall model into sub-models, each representing a different part of the system. In this way, the model builder reduces the complexity of his problem, and the smaller sub-systems can each be modelled using either the 'comprehensive' or the 'dominant-modes' approach. Decomposition of the large model into smaller parts is an important way of reducing the number of parameters to be estimated simultaneously. It can also make it easier to find the reasons for any discrepancies that show up during validation. This depends, however, on the availability of detailed data on the many internal variables that are inputs and outputs of the decomposed parts. If the validation of the sub-model is performed on data from an isolated sub-system (collected, for instance, in the laboratory), it does not cover possible unforeseen interactions between sub-systems. If data from the whole system are used, this objection does not apply; but it is important that the sub-system inputs cover an adequate range, which may be difficult to achieve when important feedbacks occur elsewhere in the system.