## COMPUTER SIMULATION IN PHYSICAL GEOGRAPHY

M. J. Kirkby, P. S. Naden, T. P. Burt *and* D. P. Butcher

Simulation modelling has grown in importance in geographical teaching and project work as microcomputers have become more readily available. By developing a wide range of usable programs in an environmental context, *Computer Simulation in Physical Geography* provides a means both of teaching physical geography using microcomputers and of teaching BASIC programming with geographical examples. Examples are drawn from hydrology, geomorphology, biogeography and climatology, and are used to illustrate different approaches to modelling and how to build rationally based computer simulations from geographical assumptions.

The book opens with an introduction to the subject, followed by a series of chapters, each of which is devoted to a different type of model, including black box models, process models, mass and energy balance models and stochastic models. Program listings are included for relevant examples together with references to the geographical basis of the simulations. The second half of the book contains methods of model or program formulation and various means of verifying and calibrating models against field data. The choice of an appropriate model for a given situation is also considered.

The programs are written in a standard subset of BASIC, and disks of the programs are available for both IBM and BBC microcomputers. Appendices give advice for converting the programs to other machines.

ISBN 0-471-90604-2

9 780471 906049

**JOHN WILEY & SONS**

*Chichester · New York · Brisbane · Toronto · Singapore*

# Computer Simulation in Physical Geography

# Computer Simulation in Physical Geography

**M. J. Kirkby**

**P. S. Naden**

**T. P. Burt**

**D. P. Butcher**

JOHN WILEY & SONS

Chichester · New York · Brisbane · Toronto · Singapore

***Library of Congress Cataloging-in-Publication Data:***

Computer simulation in physical geography.

1. Physical geography—Mathematical models.
2. Physical geography—Data processing. I. Kirkby, M. J.
GB21.5.M33C65 1987 910'.02'0724 87–8125

ISBN 0 471 90604 2

***British Library Cataloguing in Publication Data:***

Computer simulation in physical geography.
1. Physical geography—Simulation methods 2. Physical geography—Data processing
I. Kirkby, M. J.
910.02'0724 GB21.5.S55

ISBN 0 471 90604 2

Photo-typeset by Photo-Graphics, Honiton, Devon, Great Britain

Printed and bound in Great Britain

# *Preface*

During the 1980s computers have changed from being typically large mainframe machines for skilled users at large centres, to being most commonly seen in the form of desktop microcomputers which are accessible to a much larger number of people who have come to be aware of them in the home, at school or in the office. In this book we have tried to lead some of these new users with interests in physical geography towards the use and programming of simulation models. These models may be used in sixth form and university courses both to learn about aspects of the real world in an experimental situation, and as a relevant context in which to learn programming. The material in this book is developed from a workshop course given in September 1985 as part of the First International Geomorphology Conference, based in Manchester.

In any book which includes computer program listings, there is a difficulty in achieving any degree of transferability between machines. This is particularly true for the many dialects of the programming language BASIC, which is overwhelmingly the most common in use on microcomputers and has, therefore, been adopted for use here. Some degree of transferability has, however, been achieved by writing the programs in the main part of the text using only a standard subset of commands which are almost, though not completely, common to all common microcomputers. The result of this is to make the programs rather less succinct and elegant than they might have been if they had taken full advantage of the particular features of any single machine. Where machines are known to differ from the standards used, some comments have been given for guidance in the relevant appendix. The restrictions imposed by these standard commands are most severe with regard to input and output, both for text and graphics. The solution adopted has been to use standard line numbering to refer to a series of input and output subroutines for formatting text and drawing simple graphs and histograms on scaled axes. These subroutines are specific to individual types of microcomputer, so that they need to be rewritten for each machine. Appendices specify the function of each subroutine, and include listings for Acorn 'BBC' in BBC Basic, and for IBM PCs in BASICA. Users of other systems may be able to use these listings as a basis for adaptation. The listings in the text, together with the appropriate subroutines for these machines, are also available on disk, and may be ordered from the publisher in an appropriate format.

The book is divided into two parts. In Part 1, after an introductory chapter on our view of models in physical geography, and introducing the program and subroutine structure used, Chapters 2 to 5 present a series of simulation models which are grouped into model types. Chapter 2 is concerned with 'Black Box' or input–output models; Chapter 3 examines process models; Chapter 4 deals with mass balance models and Chapter 5 with stochastic models. It is recognized that these types of model are not mutually exclusive, and that many models contain elements of one or more types. Within each chapter individual models are placed primarily in their geographical context, to emphasize the range of problems to which related simulation models may be applied, and to allow the models to be used with a minimum of prior computer expertise.

In Part 2 the focus of the book shifts from the geographical context to the computing rationale. Chapter 6 gives advice on the construction of a computer simulation model from the stage of problem definition to the completion of a finished and working program. Chapter 7 is concerned with problems of choosing appropriate parameters, both in terms of optimization of a forecasting model to match an observed outcome, and in terms of sensitivity of the outcome to parameter changes. Chapter 8 addresses the issue of defining one's problem, illustrating the range of alternative model types which may be appropriate to a particular geographical context and guiding the reader towards the difficult initial step of seeing a problem in a way which is amenable to simulation. The book is completed with a series of appendices describing the standard set of BASIC commands used, the subroutine specifications, and subroutine listings for BBC and IBM microcomputers.

In the world of computing, technological advance has a habit of overtaking publication but it is hoped that the principles set out here will be of use for several years. Even if the BBC microcomputers on which the programs were developed have become obsolete, the geographical context will remain relevant to a growing number of users in both schools and institutions of higher education, as more and more people discover that computer simulation is a cost-effective tool for extending and understanding our field observations and, hence, learning about our natural environment.

March 1987

Mike Kirkby
Pam Naden
Tim Burt
Dave Butcher

# Contents

# Part I

# CHAPTER 1

# *Introduction: Getting Started*

## 1.1 WHAT IS A MODEL IN PHYSICAL GEOGRAPHY?

A model simulates the effect of an actual or hypothetical set of processes, and forecasts one or more possible outcomes. At one extreme the simulation may be as simple as substituting values into an equation or matching a pattern. At the other extreme the simulation may closely follow the detailed processes which operate in the real world. Most current models are deterministic, so that for a given set of inputs there is a unique forecast. Others are stochastic: that is they contain at least some random or chance element in the process operation or the inputs to the model, so that more than one, and usually a very large number of outcomes are possible.

Models can never fully represent the real world, but can only be analogies or analogues which have some features and behaviours in common with it. They differ very widely in their degree of similarity to the real world prototype, and in a number of ways. For example they differ in how much they physically 'look like' the prototype and in how well they forecast its behaviour.

Physical resemblance does not necessarily guarantee that a model is effective. To say that the moon looks like green cheese tells us very little about its true properties! On the other hand accurate scale models are sometimes used to model landform features and forecast their response, and this approach has been successfully used to forecast, for example, coastal erosion and sedimentation, river channel changes and drainage basin erosion. Physical similarity does not necessarily require an obvious physical resemblance, but must depend on the operation of similar physical laws in the model and real world. For example the (laminar) flow of water in a saturated aquifer is proportional to the hydraulic gradient and the hydraulic conductivity of the aquifer material (Darcy's law). Similarly flow of electricity (current) is proportional to electrical pressure (voltage) gradient and electrical conductivity (1/resistance), through Ohm's law. Thus flow through an aquifer may be modelled by measuring current in a conducting material of the same physical shape as the aquifer. This kind of analogue model is one step along a road towards an abstract representation of each process in the real world by a mathematical or logical expression. Thus the rate of flow in the aquifer is represented or modelled by the mathematical expression:

$$Q = KG \tag{1.1}$$

where $Q$ is the flow through each square metre of aquifer
$K$ is the hydraulic conductivity of the aquifer
$G$ is the hydraulic pressure gradient.

In this form the flow can be calculated at each point, either with a pencil and paper, or using a calculator or computer. If a computer model calculates flow from this expression, the model may still be said to resemble the real world, even though the computer does not physically 'look like' an aquifer.

It is clear that computer models of environmental forms or processes can only look like the real world at this abstract, mathematical level. This kind of similarity does nothing to confirm or refute the effectiveness of a model to produce realistic forecasts. It is easy, and plainly wrong, to assume that because the computer model relies on abstract mathematical expressions it must be right in its forecasts. It is equally easy, and equally wrong, to assume that because the computer model has no physical resemblance to a landscape then its forecasts must be incorrect. The truth generally lies between these two extremes for computer models, as for all kinds of models, hypotheses or theories. No model is better than the assumptions and data it relies on. In circumstances where the assumptions are valid and/or forecasts are being made within the range of data used to establish the models, then it will probably give reasonable results. Where one is forecasting outside this area of tested reliability, no model should be relied on without further testing. In other words the well-worn computer dictum of 'Garbage in – Garbage out' applies to models as to other computer operations, and the inscrutability of the computer should never be confused with reliability.

If a model is to be of any use, either practical or theoretical, it should produce testable outcomes. Computer models are necessarily numerical and logical, so that they are one form of quantitative model, producing a definite forecast in each run of the model. In many cases this forecast can be compared numerically with real world measurements (which should be independent of the data used to set up the model) to make a very exact test of performance. It is, however, also possible to make qualitative deductions from model output which can be compared with the real world, at a less precise although not always at a 'worse' level. In fact, where models have a stochastic element, either in the processes they represent or through uncertainty in the constants put into them, the numerical outcome must become less precise, so that there is a continuous range of models from strictly deterministic quantitative models with a unique outcome; through a range of stochastic models in which outcomes are more or less probable; to entirely qualitative models including many of those which have been traditionally successful in physical geography like W. M. Davis' geographical cycle of erosion.

Models which are suitable for computer programming must be expressible in the form of a strictly logical and/or numerical procedure. Most of the examples in this book are mainly deterministic though some have a significant random component. A logical procedure is necessary if a model is to be implemented on a computer, but not all logical models make useful programs. In some cases the trouble taken in writing a program does not justify the time saved in following the procedure, however logical. This is particularly true if the model will only be used a few times.

## 1.2 TYPES OF COMPUTER MODEL

The various types of computer model are discussed in Chapters 2 to 5. Although they build up in some senses from simple to complex, the categories overlap considerably, and many working models contain elements of several types. The categories used here are 'black box models' (Chapter 2), 'process models' (Chapter 3), 'mass balance models' (Chapter 4) and 'stochastic models' (Chapter 5), although there is no agreed categorization of model types along these or other lines. In the simplest models, the workings of the model are treated as invisible to the user, who feeds in input and takes out forecasts. Most computer models can be treated in this way by their users, even though we hope that you will not be content with this level of knowledge! Such models are called 'black box' or 'input/output' models. These terms are usually applied to models in which the internal workings of the model are not intended to directly represent the processes operating in the real world, even at an abstract mathematical level. Perhaps the most widespread example of this type of model is where input and output are related to one another by a statistical curve fitting or other regression procedure. Output may then be estimated from input data, but little is learned about the operation of the linking process. This type of model may be very effective, and in cases where processes are poorly understood is likely to be the best available in terms of forecasting accuracy. It tends, however, to behave worse than process and/or mass balance models when input lies outside the range of the original test data.

Both process models and mass balance models attempt to shine some light into the black box. If all processes and relationships are fully represented in the model, then we may be said to have a 'white box' model. Such a model is usually exceedingly complicated, even where we have the knowledge to construct it, so that most achievable models may be considered as boxes in various rather dark shades of grey. There is always a conflict between greater reality, associated with greater complexity, and greater simplicity at the expense of detail in representing reality. The 'best' model can only be judged by how well it satisfies its original purpose.

Process models describe the mechanisms of particular operations which occur in the real world. For example a black box model of soil erosion may estimate erosion rates in terms of empirical equations, from storm rainfall, slope length and slope gradient. In a process model the erosion may be separated into rainsplash, and sheetwash in overland flow. Rainsplash may be forecast in terms of rainfall intensity and soil properties; and sheetwash through estimates of overland flow and its transporting capacity. In other words the model is built up from a flow diagram which represents the physical storages and/or flows of energy or material in the real world. The process model can never be perfect or complete, but it is an attempt to make the conceptual behaviour of the model resemble the real world more closely than in a black box model. In many if not most models there are many processes operating at once, and interacting with one another. The overall model can therefore be assembled from a number of sub-models, each representing a single process or group of processes. In an effective model the various sub-models need to be dealing in flows of the same commodity, usually mass or energy; and to be dealing in it at similar levels of space and time resolution. It is, for example, extremely difficult to reconcile a model of atmospheric circulation which deals in square $100 \times 100$ kilometre 'cells' with a hydrological model which works with irregularly shaped small drainage basins of about one square kilometre.

A very important set of constraints on most models is that, barring nuclear explosions and radioactive decay, mass and energy are neither created nor destroyed. This conservation applies not only to total mass, but to the masses of individual chemical elements, like, say, iron or carbon. It also applies to compound materials, provided allowance is made for chemical changes and change of state. Perhaps the most obvious example of this principle, and one of the most important for physical geography, is the hydrological cycle. It requires that the mass of water is conserved if due allowance is made for chemical changes like release from volcanoes or incorporation into sediments; and for changes of state to and from ice and water vapour. In a similar way we can generally appeal to the conservation of total rock and soil materials. In following the course of weathering or nutrient cycling we may be equally interested in budgeting say silicon or carbon. In all of these mass budgets there is little loss of material from the system of interest, so that the relevant budgets form very effective overall controls of many of the systems we will be looking at. Although energy is also conserved, energy balances prove somewhat less useful than mass balances in most of the cases we will look at. The reason for this preference for mass balance is that there are large losses of energy in most mechanical systems, and the losses are not well understood, so that many of the largest terms in the energy budget are uncertain. There is, however, one important exception where an effective energy budget is crucial: in modelling the temperature of the earth's surface and evapo-transpiration from it. For this essentially thermodynamic system the energy losses are relatively small.

Wherever mass or energy balances are appropriate, the model is constrained by some form of the 'storage equation':

$$\text{Input} - \text{Output} = \text{Net increase in storage} \qquad (1.2)$$

This equation applies not only to the system as a whole, but to each spatial compartment of it. It is applied most commonly to mass of water or earth materials, and less commonly to mass of individual elements, to total or radiation energy or to biological populations. The importance of the mass balance approach cannot be overstressed. It provides a common strand to many models of interest, and gives many of them a distinct family resemblance, as will be seen in Chapter 4. In opening up the black box model, mass or energy balance may be thought of as providing the framework for a physically based model, within which the individual process models are supported. Storage equations also have an important role in providing the formal link between rates of change of flows across space and rates of change of state over time at a point. In equation (1.2), the flows provide the input and output terms on the left-hand side, while the change of state is the increase in storage on the right-hand side. A final and very practical virtue of mass or energy storage equations is that they have a much better physical foundation than many of the processes which we hang around them, so that they help to keep forecasts within the range of the possible. In fact it may be argued that in some large models much of their overall forecasting power is based on the constraining effect of mass balance rather than on our reliable understanding of the processes acting.

The last type of model discussed in Part I is the stochastic model (Chapter 5). This category tends to run across all the others, although the simplest models generally lack a stochastic element. Random elements are usually included to represent processes or forms which are outside the scope of the model. The real world we are attempting to model may usually be conceived as strictly deterministic, so that processes are not random in principle. Nevertheless we may not wish to include the causes of every process within our model. These processes may then be represented by a series of random numbers drawn from a specified probability distribution. Some examples show the sort of cases in which this approach is useful. If we have a hydrological simulation model which converts rainfall into stream runoff, we can use it to forecast the runoff from a specified storm sequence. If, alternatively we want to know the size of the flood which will, on average, be exceeded every 100 years, then a simulated random rainfall sequence may be generated from our knowledge of local rainfall distributions. This sequence may then be used to forecast flood sizes and their distribution, without needing a very long real runoff record. This approach is likely to be more cost effective and more reliable than attempting to forecast rainfall from a global circulation model.

A second example in which a stochastic model may be appropriate is in producing the initial surface on which erosion occurs. This will always have some faint irregular relief, but we are not usually interested in its exact form. We might therefore generate the irregularities as suitable random numbers instead of investigating and modelling the causes of the micro-relief. In each of these two examples, the use of random numbers does not imply that rainfall or relief could take *any* value. Instead the random numbers would be drawn from a very definite probability distribution. A random number might, for example, be drawn from a normal distribution with an average of 100 and a standard deviation of 1, so that most values lie between 98 and 102. Randomness may be constrained to this or any extent, and certainly does not mean that model outcomes lack regularity; but only that outcomes are not uniquely determined.

There is another kind of stochastic variation which is not included in Chapter 5, but which is important in testing models. The outcome of any model is determined by the *parameters* which are used to describe initial states, process rate constants, etc. Most of these parameters are not known to a high degree of accuracy, because of experimental error and field variations from site to site. If each parameter value is drawn from a random distribution which represents its range of possible values, then a series of runs of the model, even of a deterministic model, will give a distribution of outcomes. This distribution may then be compared with field site measurements to determine whether the real world values lie within the range of forecast outcomes of the model. This topic is taken up again in Chapter 7.

## 1.3 WRITING A COMPUTER SIMULATION MODEL

This section provides a brief introduction to model development, so that you can read and use the programs of Part I. Methods of model creation and development are discussed in Chapter 6, but are briefly summarized in the flow diagram of Figure 1.1. The first step is to define a procedure for forecasting on the basis of your understanding of the environmental problem. This procedure will usually conform to one or more of the model types discussed above, and must ultimately be expressible in logical or numerical form. The best guide to developing suitable procedures is to follow other examples as far as appropriate. However, this is not always possible and you will at times need to refer directly to scientific or mathematical principles.

Once a procedure or algorithm has been designed, the various steps of the routine must be put into a definite order of operations, which the computer must perform. Even though a number of operations may in the real world be simultaneous, the computer must be instructed to do them in a definite order. Construction of a flow diagram like Figure 1.1 is helpful to clarify this sequence. The next step is to ask how you wish to enter the parameter values and extract data on the model outcome at one or more stages. Values may be entered from the keyboard, through DATA statements in the BASIC program, by calculation from a mathematical or logical function, or from a file stored on magnetic disc, tape or other medium. In the simple programs illustrated in this book, the last of these is not usually used, but for models with large data requirements a file is a likely choice. In many cases it is helpful to the program user, even if only yourself, to prompt keyboard entry with a plausible value or range of values, which can then be revised by the user. This method is illustrated below. Output from

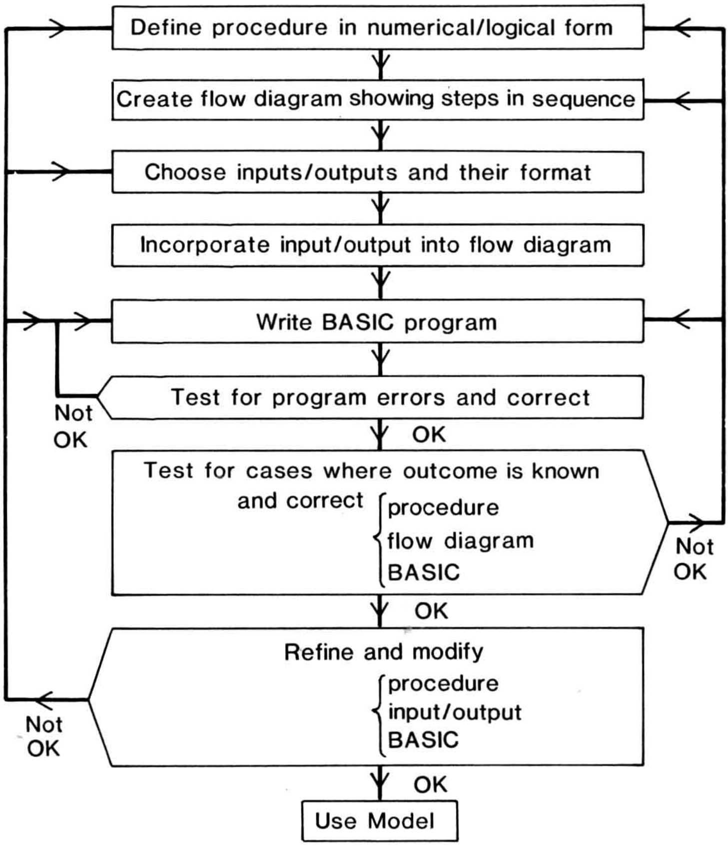


Figure 1.1 Summary procedure for designing and writing a computer model

the program is primarily through the monitor/VDU/TV screen, and may be in text form or as some type of graph (or other graphic design). Where a printer is connected, this screen output may be replicated on paper. Another possibility for large models is to store results to disc or tape as a data file. The flow diagram now needs to be modified to fit input and output into the sequence of computer operations. While developing a program and testing it for errors, it is often useful to obtain output at many more stages than will be required in the final running model.

With what is now a complete flow diagram, you are in a position to write a program in BASIC or any other computer language. If other people are to use your program, and even if you wish to return to it after the lapse of a few days or weeks, it is helpful to sprinkle the program text with remark (REM) statements to explain what the program is doing, preferably in terms of the flow diagram.