2 Modelling and Model Building

Mark Mulligan[1] and John Wainwright[2]

[1]Department of Geography, King's College London, UK
[2]Department of Geography, Durham University, UK

Modelling is like sin. Once you begin with one form of it you are pushed to others. In fact, as with sin, once you begin with one form you ought to consider other forms . . . But unlike sin – or at any rate unlike sin as a moral purist conceives of it – modelling is the best reaction to the situation in which we find ourselves. Given the meagreness of our intelligence in comparison with the complexity and subtlety of nature, if we want to say things which are true, as well as things which are useful and things which are testable, then we had better relate our bids for truth, application and testability in some fairly sophisticated ways. This is what modelling does.

(Morton and Suárez, 'Kinds of models', 2001)

2.1 The role of modelling in environmental research

2.1.1 The nature of research

Research is a means of improvement through understanding. This improvement may be personal but it may also be tied to development. We may hope to improve human health and wellbeing through research into diseases such as cancer and heart disease. We may wish to improve the design of bridges or aircraft through research in materials science, which provides lighter, stronger, longer lasting or cheaper (in terms of building and maintenance) bridge structures. We may wish to produce more or better crops with less adverse impact on the environment through research in biotechnology. In all of these cases research provides, in the first instance, better understanding of how things are and how they work, which can then contribute to the improvement or optimization of these systems through the development of new techniques, processes, materials and protocols.

Research is traditionally carried out through the accumulation of observations of systems and system behaviour under 'natural' circumstances and during experimental manipulation. These observations provide the evidence upon which hypotheses can be generated about the structure and operation (function) of the systems. These hypotheses can be tested against new observations and, where they prove to be reliable descriptors of the system or system behaviour, then they may eventually gain recognition as proven theory or general law as far as that is possible.

The conditions, which are required to facilitate research, include:

(a) a means of observation and comparative observation (measurement);
(b) a means of controlling or forcing aspects of the system (experimentation);
(c) an understanding of previous research and the state of knowledge (context); and
(d) a means of cross-referencing and connecting threads of (a), (b) and (c) (imagination).

Environmental Modelling: Finding Simplicity in Complexity, Second Edition. Edited by John Wainwright and Mark Mulligan.

2.1.2 A Model for environmental research

What do we mean by the term *model*? A model is an abstraction of reality. This abstraction represents a complex reality in the simplest way that is adequate for the purpose of modelling. The best model is always that which achieves the greatest realism with the least parameter complexity (parsimony) and the least model complexity. Realism can be measured objectively as agreement between model outputs and real-world observations, or less objectively as the process insight or new understanding gained from the model.

Parsimony (using no more complex a model or representation of reality than is absolutely necessary) has been a guiding principle in scientific investigations since Aristotle who claimed:

> It is the mark of an instructed mind to rest satisfied with the degree of precision which the nature of the subject permits and not to seek an exactness where only an approximation of the truth is possible

though it was particularly strong in Mediaeval times and was enunciated then by William of Ockham, in his famous 'razor' (Lark, 2001). Newton stated it as the first of his principles for fruitful scientific research in *Principia* as:

> We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.

Parsimony is a prerequisite for effective scientific explanation, not an indication that nature *necessarily* operates on the basis of parsimonious principles. It is an important principle in fields as far apart as taxonomy and biochemistry and is fundamental to likelihood and Bayesian approaches of statistical inference. In a modelling context, a parsimonious model is usually the one with the greatest explanation or predictive power and the least parameters or process complexity. It is a particularly important principle in modelling because our ability to model complexity is much greater than our ability to provide the data to parameterize, calibrate and validate those same models. Scientific explanations must be both relevant *and* testable. Unevaluated models are no better than untested hypotheses. If the application of the principle of parsimony facilitates model evaluation then it also facilitates utility of models.

2.1.3 The nature of modelling

Modelling is not an alternative to observation but, under certain circumstances, can be a powerful tool in understanding observations and in developing and testing theory. Direct observation (as opposed to remote observation or estimation through spatial or temporal statistical inference) will always be closer to truth and must remain the most important component of scientific investigation. Klemeš (1997: 48) describes the forces at work in putting the modelling 'cart' before the observational 'horse' as is sometimes apparent in modelling studies:

> It is easier and more fun to play with a computer than to face the rigors of fieldwork especially hydrologic fieldwork, which is usually most intensive during the most adverse conditions. It is faster to get a result by modeling than through acquisition and analysis of more data, which suits managers and politicians as well as staff scientists and professors to whom it means more publications per unit time and thus an easier passage of the hurdles of annual evaluations and other paper-counting rituals. And it is more glamorous to polish mathematical equations (even bad ones) in the office than muddied boots (even good ones) in the field.
>
> Klemeš (1997: 48)

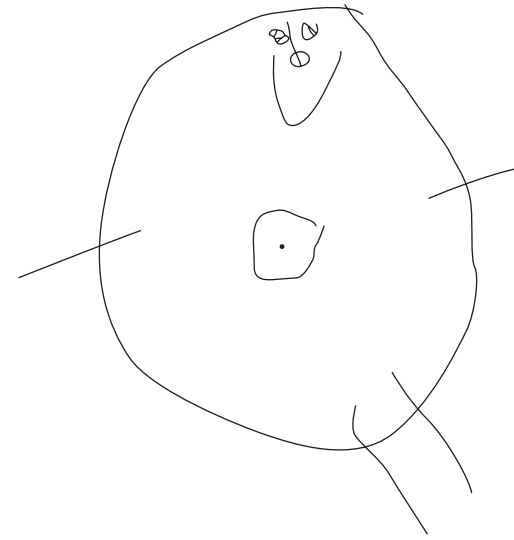
A model is an abstraction of a real system; it is a simplification in which only those components that are seen to be significant to the problem at hand are represented in the model. In this representation, a model takes influence from aspects of the real system and aspects from the modeller's perception of the system and its importance to the problem at hand. Modelling supports the conceptualization and exploration of the behaviour of objects or processes and their interaction. Modelling is a means of better understanding and generating hypotheses. Modelling also supports the development of (numerical) experiments in which hypotheses can be tested and outcomes predicted. In science understanding is the goal and models serve as one tool in the toolkit used towards that end (Baker, 1998).

Cross and Moscardini (1985: 22) describe modelling as 'an art with a rational basis which requires the use of common sense at least as much as mathematical expertise.' Modelling is described as an art because it involves experience and intuition as well as the development of a set of (mathematical) skills (although many mathematicians would argue that mathematics also requires intuition and experience to be carried out well). Cross and Moscardini (1985) argue that it is intuition and the resulting insight that distinguish good modellers from mediocre ones. Intuition cannot be taught and comes from the experience of designing, building and using models. One learns modelling by doing modelling. The reader should look at the environmental issues presented

in this book and abstract from them the key elements that might be required to build a useful simulation model. Abstraction is a difficult skill to acquire in adults (we tend to overcomplicate) though young children have the skill well honed as they operate their own mental models of how the world works before parents and teachers provide them with alternative models. A good exercise in judging your own abstraction skills may be carried out with a simple piece of paper. Think of all the faces that you know: the short round ones, the long thin ones, the European, African, Asian and South American ones; the ones with beards and those without. How might we abstract from this sea of faces a simple model for the human face? Try that on your piece of paper. Give yourself two minutes.

Our guess is that you made it too complex. The bare minimum we need is a circle, dots for eyes and a upwards facing curve for a mouth. The yellow smiley face is a good example and is one of the most common images in modern life. If you are not sure what we mean, do a Web search for 'yellow smiley face'. We do not need hair, ears, eyebrows, eyelashes or anything else to recognize this as a face. Indeed some real faces do not have those features (or at least they cannot be seen) so adding them to your model as a necessary condition for recognition as a face, reduces the generality of your model. Children are very good at abstraction as the four year old's image of a person in Figure 2.1 indicates: a single shape for the body, stick arms and legs, button eyes and nose and smiley mouth. Nothing else is needed as this is very clearly an abstraction of the human body. An element of bias is added as for this child the belly button is also an important component of the human form, hence it is in the model!

Arm yourself with a spreadsheet and turn your abstraction into numbers and simple equations. Play, examine, delete, add, think and play some more with the numbers and the equations. What can you learn about the system? What still confuses? Experience of this kind will help develop intuition and insight where it is lacking. We present you with a series of modelling problems on the web site that complements this book and going over them repeatedly will help further. The key to successful modelling is to be able to abstract carefully so that your model is on the one hand simple but on the other hand realistic enough to offer a solution to the problem at hand. Considering a cow as spherical may be appropriate for understanding some elements of how a cow works (Harte, 1985), but will not be all that helpful in understanding its locomotion!



Olive Mulligan, aged 4

Figure 2.1 Children are often very good at abstraction because they tend not to see things in the complicated ways that adults do (or to have complex preconceptions about them). This is a four year old's abstraction of a human – clearly recognizable, if not detailed (Courtesy of Olive Mulligan [aged 4]).

You are not new to modelling – everyone does it! All scientists use some form of conceptual or mental model of the data they work with. Even data are, in fact, models; they are simplified representations of (unobservable) processes, time and space, compared with the reality, all sensors form a model of reality. For example, a temperature sensor measures change in the level of a column of mercury as this level is linearly related to a change in temperature. The changing level of mercury *is* an empirical model for a temperature change. (Consider how different a digital thermometer actually is from an analogue one using mercury.) Your whole perception of reality is a model, not the reality itself. You are armed with a series of sensors for light in the visible spectrum (eyes) and certain wavelengths of sound (ears), which are only fractions of what can be sensed. Other animals have different perceptions of the same environmental characteristics because they have different sensors, but also a different mental model and context for decoding those signals. There is thus little difference between modelling and other scientific endeavours (and indeed life itself).

2.1.4 Researching environmental systems

According to some, we have crossed a geological boundary from the Holocene to the Anthropocene (Crutzen,

2002; Steffen *et al*., 2007; Zalasiewicz *et al*., 2010; Brown, 2011). The Holocene was an epoch of unprecedented stability that enabled complex societies, cultures, agricultures and infrastructures to be developed eventually supporting some seven billion people (Ruddiman, 2007). In the Anthropocene, humans are a major geological force generating planetary scale change in climate, land, water and ecosystems. Our increasing individual impacts on the environment coupled with our sheer numbers and their growth promises to put an end to this era of stability in favour of an epoch of unprecedented instability. In order to maintain and sustain water, food, shelter, livelihoods and culture we will need to manage our impact on nature much more effectively than ever before. We can only manage what we understand, so researching environmental systems is more important than ever.

Modelling has grown significantly as a research activity since the 1950s, reflecting conceptual developments in the modelling techniques themselves, technological developments in computation, scientific developments indicating increased need to study systems (especially environmental ones) in an integrated manner and an increased demand for extrapolation (especially prediction) in space and time.

Modelling has become one of the most powerful tools in the workshop of environmental scientists who are charged with better understanding the interactions between the environment, ecosystems and the populations of humans and other animals. This understanding is increasingly important in environmental stewardship (monitoring and management) and the development of increasingly sustainable means of human dependency on environmental systems and the services that they provide.

Environmental systems are, of course, the same systems as those studied by physicists, chemists and biologists but the level of abstraction of the environmental scientist is very different from that of many of these scientists. Whereas a physicist might study the behaviour of gases, liquids or solids under controlled conditions of temperature or pressure and a chemist might study the interaction of molecules in aqueous solution, a biologist must integrate what we know from these sciences to understand how a cell – or a plant – or an animal, lives and functions. The environmental scientist or geographer or ecologist approaches their science at a much greater level of abstraction in which physical and chemical 'laws' provide the rule base for understanding the interaction between living organisms and their nonliving environments, the characteristics of each and the processes through which each functions.

Integrated environmental systems are different in many ways from the isolated objects of study in physics and chemistry although the integrated study of the environment cannot take place without the building blocks provided by research in physics and chemistry. The systems studied by environmental scientists are characteristically:

Large scale, long term. Though the environmental scientist may only study a small time- and space-scale slice of the system, this slice invariably fits within the context of a system that has evolved over hundreds, thousands or millions of years and which will continue to evolve into the future. It is also a slice that takes in material and energy from a hierarchy of neighbours from the local, through regional, to global scale. It is this context, which provides much of the complexity of environmental systems compared with the much more reductionist systems of the traditional 'hard' sciences. To the environmental scientist models are a means of integrating across time and through space in order to understand how these contexts determine the nature and functioning of the system under study.

Multicomponent. Environmental scientists rarely have the good fortune of studying a single component of their system in isolation. Most questions asked of environmental scientists require understanding of interactions between multiple living (biotic) and nonliving (abiotic) systems and their interaction. Complexity increases greatly as number of components increases, where their interactions are also taken into account. Since the human mind has some considerable difficulty in dealing with chains of causality with more than a few links, to an environmental scientist models are an important means of breaking systems into intellectually manageable components and combining them and making explicit the interactions between them.

Non-laboratory controllable. The luxury of controlled conditions under which to test the impact of individual forcing factors on the behaviour of the study system is very rarely available to environmental scientists. Very few environmental systems can be rebuilt in the laboratory (laboratory-based physical modelling) with an appropriate level of sophistication to represent them adequately. Taking the laboratory to the field (field-based physical modelling) is an alternative as has been shown by the Free Atmosphere CO_2 Enrichment (FACE) experiments (Hall, 2001), BIOSPHERE 2 (Cohn, 2002) and a range of other environmental manipulation experiments. Field-based

physical models are very limited in the degree of control available to the scientist because of the enormous expense associated with them. They are also very limited in the scale at which they can be applied, again because of expense and engineering limitations. So, the fact remains that, at the scale at which environmental scientists work, their systems remain effectively noncontrollable with only small components capable of undergoing controlled experiments. However, some do argue that the environment itself is one large laboratory, which is sustaining global-scale experiments through, for example, greenhouse-gas emissions (Govindasamy *et al.*, 2003). These are not the kind of experiments that enable us to predict (as they are real time) nor which help us, in the short term at least, to better interact with or manage the environment (notwithstanding the moral implications of this attitude!). Models provide an inexpensive laboratory in which mathematical descriptions of systems and processes can be forced in a controlled way.

Multiscale, multidisciplinary. Environmental systems are multiscale with environmental scientists needing to understand or experiment at scales from the atom through the molecule to the cell, organism or object, population or objects, community or landscape through to the ecosystem and beyond. This presence of multiple scales means that environmental scientists are rarely just environmental scientists; they may be physicists, chemists, physical chemists, engineers, biologists, botanists, zoologists, anthropologists, population geographers, physical geographers, ecologists, social geographers, political scientists, lawyers, environmental economists or indeed environmental scientists in their training but who later apply themselves to environmental science. Environmental science is thus an interdisciplinary science that cuts across the traditional boundaries of academic research. Tackling contemporary environmental problems often involves large multidisciplinary (and often multinational) teams working together on different aspects of the system. Modelling provides an integrative framework in which these disparate disciplines can work on individual aspects of the research problem and supply a module for integration within the modelling framework. Disciplinary and national boundaries, research 'cultures' and research 'languages' are thus less of a barrier.

Multivariate, nonlinear and complex. It goes without saying that complex and integrated systems such as those handled by environmental scientists are multivariate and, as a result, the relationships between individual variables are often nonlinear and complex. Models provide a means of deconstructing the complexity of environmental systems and, through experimentation, of understanding the univariate contribution to multivariate complexity.

In addition to these properties of environmental systems the rationale behind much research in environmental systems is often a practical or applied one such that research in environmental science also has to incorporate the following needs.

The need to look into the future. Environmental research often involves extrapolation into the future in order to understand the impacts of some current state or process. Prediction is difficult, not least because predictions of the future can only be tested in the future (at which point they are no longer predictions). Models are very often used as a tool for integration of understanding over time and thus are well suited for prediction and retrodiction. As with any means of predicting the future, the prediction in only as good as the information and understanding upon which it is based. This limitation may be sufficient where one is working within process domains that have already been experienced during the period in which the understanding was developed, but when future conditions cross a process domain, the reality may be quite different to the expectation. Thus we often talk about *projecting* into the future rather than *predicting* into the future, in recognition of the fact that we are fundamentally limited to projecting our present understanding into the future as one *possible* outcome rather than providing a reliable forecast of future processes and their outcomes.

The need to understand the impact of events that have not happened (yet). Environmental research very often concerns developing scenarios for change and understanding the impacts of these scenarios on systems upon which humans depend. These changes may be developmental, such as the building of houses, industrial units, bridges, ports or golf courses and thus requiring environmental impact assessments (EIAs). Alternatively they may be more abstract events such as climate change or land-use and cover change (LUCC). In either case, where models have been developed on the basis of process understanding or a knowledge of the response of similar systems to similar or analogous change, they are often used as a means of understanding the impact of expected events.

The need to understand the impacts of human behaviour. With global human populations continuing to increase and *per capita* resource use high and increasing in the developed world and low but increasing in much of the developing world, the need to achieve renewable and nonrenewable resource use that can be sustained into the distant future becomes more pressing. Better understanding the impacts of human resource use (fishing, foresting, hunting, agriculture, mining) on the environment and its ability to sustain these resources is thus an increasing thrust of environmental research. Models, for many of the reasons outlined above, are often employed to investigate the enhancement and degradation of resources through human impact.

The need to understand the impacts on human behaviour. With the human population so high and concentrated and with *per capita* resource needs so high and sites of production so disparate from sites of consumption, human society is increasingly sensitive to environmental change. Where environmental change affects resource supply, resource demand or the ease and cost of resource transportation, the impact on human populations is likely to be high. Therefore understanding the nature of variation and change in environmental systems and the feedbacks of human impacts on the environment to human populations are both increasingly important. Environmental science increasingly needs to be a supplier of reliable forecasts and understanding to the world of human health and welfare, food and water security, development, politics, peacekeeping and warmongering.

2.2 Approaches to model building: chickens, eggs, models and parameters?

Should a model be designed around available measurements or should data collection be carried out only once the model structure has been fully developed? Many hardened modellers would specify the latter choice as the most appropriate. The parameters that are required to carry out specific model applications are clearly best defined by the model structure that best represents the processes at work. Indeed, modelling can be used in this way to design the larger research programme. Only by taking the measurements that can demonstrate that the operation of the model conforms to the 'real world' is it possible to decide whether we have truly understood the processes and their interactions.

However, actual model applications may not be so simple. We may be interested in trying to reconstruct past environments, or the conditions that led to catastrophic slope collapse or major flooding. In such cases, it is not possible to measure all of the parameters of a model that has a reasonable process basis, as the conditions we are interested in no longer exist. In such cases, we may have to make reasonable guesses (or estimates, if you prefer) based on indirect evidence. The modelling procedure may be carried out iteratively to investigate which of a number of reconstructions may be most feasible.

Our optimal model structure may also produce parameters that it is not possible to measure in the field setting, especially at the scales in which they are represented in the model. The limitations may be due to cost, or the lack of appropriate techniques. It may be necessary to derive transfer functions from (surrogate) parameters that are simpler to measure. For example, in the case of infiltration into hillslopes, the most realistic results are generally obtained using rainfall simulation, as this approach best represents the process we are trying to parameterize (although simulated rain is never exactly the same as real rain – see Wainwright *et al.*, 2000, for implications). However, rainfall simulation is relatively difficult and expensive to carry out, and generally requires large volumes of water. It may not be feasible to obtain or transport such quantities, particularly in remote locations – and most catchments contain some remote locations. Thus, it may be better to parameterize using an alternative measurement such as cylinder infiltration, or pedo-transfer functions that only require information about soil texture. Such measurements may not give exactly the same values as would occur under real rainfall, so it may be necessary to use some form of calibration or tuning for such parameters to ensure agreement between model output and observations. In extreme cases, it may be necessary to attempt to calibrate the model parameter relative to a known output if information is not available. We will return to the problems with this approach later.

Parameterization is also costly. Work in the field requires considerable investment of time and generally also money. Indeed, some sceptics suggest that the research focus on modelling is driven by the need to keep costs down and PhDs finished within three years (Klemeš, 1997). Equipment may also be expensive and if it is providing a continuous monitored record, will need periodic attention to download data and carry out repairs. Therefore, it will generally never be possible to obtain as many measurements as might be desirable in any particular application. As a general rule of thumb,

we should invest in parameter measurement according to how big an effect the parameter has on the model output of interest. The magnitude of the effect of parameters on model output is known as the sensitivity of a model to its parameters. This important stage of analysis will be dealt with in more detail below.

2.2.1 Defining the sampling strategy

Like models, measurements are also abstractions of reality, the results of a measurement campaign will depend as much upon the timing, technique, spatial distribution, scale and density of sampling as on the reality of the data being measured. As in modelling, it is imperative that careful thought is given to the conceptualization and design of a sampling strategy appropriate to the parameter being measured and the objective of the measurement. This is particularly true when the sampled data are to be used to parameterize or to validate models. If a model underperforms in terms of predictive or explanatory power this can be the result of in appropriate sampling for parameterization or validation as much as model performance itself. It is often assumed implicitly that data represents reality better than model does (or indeed that data is reality). Both are models and it is important to be critical of both.

We can think of the sampling strategy in terms of (i) the variables and parameters to be measured for parameterization, calibration and validation, (ii) the direct or indirect techniques to be used in measurement and their inherent scale of representation, (iii) the spatial sampling scheme (distributed, semi distributed, lumped) and its density, (iv) the temporal sampling scheme (duration and temporal resolution). Choosing which variables will be measured for parameterization and the intensity of measurement will depend very much of the sensitivity of the significant model outputs to those (see below). Highly sensitive parameters should be high on the agenda of monitoring programmes but as model sensitivity to a parameter is usually also dependent on the value of other parameters, this is not always as straightforward as it might at first appear. Where variables are insensitive either they should not be in the model in the first place or their measurement can be simplified to reflect this. Calibration parameters should be, as much as possible, those without physical meaning so as not to compromise the physical basis of the model and their measurement will be necessary for the application of models to new environments or epochs. Validation parameters and variables should be those that are the critical model outputs in terms of the purpose of the model. A robust validation of the key model output would tend to indicate that the model has performed well in a predictive sense. This outcome does not mean that the results have been obtained for the correct reasons, in other words good prediction is no guarantee of good explanation. In this way, if one were to validate the output of a catchment hydrological model using measured discharge data and obtain good agreement between model and data, this success can come about as the result of many different configurations of the driving variables for discharge. It is thus important in validation to validate the output required but also some internal variable that would indicate whether that output has been arrived at for the correct reasons, in this case the spatial distribution of soil moisture around the catchment.

The techniques used for measurement will depend upon a number of logistic constraints such as availability, cost, dependence on power supplies, training required for use and safety but must also depend upon the spatial and temporal structure of the model for which these techniques will provide data since it is important that the model and the data are representing the same thing. A good example is soil moisture. Soil is a three-phase medium consisting of the soil matrix, rock fragments greater than 2 mm in diameter and of a porosity. Soil moisture occupies the porosity, which is usually around half of the soil volume. In many soils, rock-fragment content can be in excess of 30% (van Wesemael *et al.*, 2000) and whilst rock fragments sometimes have a small porosity, it is usually quite insignificant for the purposes of moisture retention. Volumetric measurement of soil moisture usually provides an output of m^3 water per m^3 soil fine fraction which does not usually contain rock fragments. The latter tend to be avoided in the installation of, and not accounted for in calibration of, electronic sensors of soil moisture and tend to be avoided or sieved out of gravimetric samples. Soil-moisture measurements are usually an aggregate of small sample measurements of the fine soil fraction. However soil tends to be represented as large blocks with dimensions of tens to hundreds of metres in hydrological models. The move to this larger-scale representation must therefore incorporate a significant loss of available porespace because of the presence of rock fragments and thus the nature of soil moisture at this scale is quite different to that at the point scale of measurement. The need to balance data and model attributes is particularly clear where indirect measurements, in particular remote sensing, are used for model parameterization.

Over recent decades there has been a move away from lumped models in which spatial heterogeneity is not

represented and towards distributed models in which it is. Advances in computing power and GIS technologies have enabled the development of complex spatial models based on the discretization of landscapes into vector polygons, triangular irregular networks, objects of complex form or simple raster grids. Despite recent advances in remote sensing there are still very many parameters that cannot be measured using electromagnetic radiation and thus remote sensing. The sophistication of spatial models has rapidly outgrown our ability to parameterize them spatially and they thus remain conceptually lumped (Beven, 1992). The appropriate scale of distribution and the optimum configuration of measurements for model parameterization or calibration is the subject of much debate. For example, Musters and Bouten (2000) used their model of root-water uptake to determine the optimal sampling strategy for the soil-moisture probes used to parameterize it. Fieldwork is an expensive, labour-intensive, time-consuming and sometimes uncomfortable or even hazardous activity. Traditional random or structured sampling procedures usually require that a very large number of samples be collected in order to fulfil the assumptions of statistical inference. In order to reduce the sampling effort, prior knowledge about the system under study may be used to guide convenience or non-random sampling which is still statistically viable, with the appropriate method depending on the type of prior knowledge available (Mode *et al*., 2002). Ranked set sampling (Mode *et al*., 1999) reduces the cost of sampling by using 'rough but cheap' quantitative or qualitative information to guide the sampling process for the real, more expensive sampling process. Chao and Thompson (2001) and others indicate the value of optimal adaptive sampling strategies in which the spatial or temporal sampling evolves over time according to the values of sites or times already sampled. A number of authors indicate how optimal sampling can be achieved by algorithmic approaches that maximize entropy in the results obtained (e.g. Bueso *et al*., 1998; Schaetzen *et al*., 2000). The luxury of optimizing your sampling scheme in this way is, however, not always available to the modeller, especially within the context of policy models that are applied using existing datasets generated by government agencies, for example where 'you get what you are given' and which may not be collected with uniform or standard protocols (e.g. as outlined for soils data in Spain by Barahona and Iriarte, 2001) or where the protocol may evolve over time affecting the legitimacy of time-series analysis. Usually the spatial sampling scheme chosen is a compromise between that which best represents the system under investigation and the computational resources and data available. This compromise is most clearly seen in the extensive discussions on the problem of grid size and subgrid variability in general circulation models (GCMs). May and Roeckner (2001), amongst others, indicate the importance of grid resolution in affecting the results of GCMs. Smaller grid sizes produce more realistic results, especially, in highly mountainous areas, but smaller grids also have substantially higher computational and data costs.

Wainwright *et al*. (1999a) indicated the importance of the temporal detail of climate data for accurate hydrological modelling. The calculation of evapotranspiration using the Penman–Monteith formula for hourly data and then the same data aggregated to a single value for each day and then separately for each day and night indicates that the day-night aggregation produces much closer results to the original hourly data than does the daily aggregation because of the domain change in net radiation values from daylight hours when they are positive to night time hours when they are negative. The error induced by aggregation to daily timestep is of the order of 100% and varies with the month of the year too. This indicates that one must pay attention to the natural scales and boundaries of the processes being modelled when devising the time (or space) scale for sampling. Similarly, Mulligan (1998) demonstrated the importance of high temporal resolution rainfall intensity data for understanding the partitioning between infiltration and overland flow. Where soil infiltration rates fall within the range of measured instantaneous rainfall intensities (as they often do), it is important to understand the distribution function of instantaneous intensities. The greater the timescale over which these intensities are aggregated, the lower the measured intensity would be. Such aggregation can have major effects on the predicted levels of Hortonian or infiltration excess overland flow production – which is, after all, a threshold process (see Wainwright and Parsons, 2002, for spatial implications). Hansen *et al*. (1996) suggested the importance of data quality in determining streamflow prediction for the lumped IHACRES rainfall-runoff model to conclude that rain-gauge density and the sampling interval of rainfall are the most critical across a range of catchments. Understanding these sensitivities is critical to designing an appropriate sampling scheme.

2.2.2 What happens when the parameters don't work?

It is frequently the case that initial parameter estimates will produce model outputs that are incompatible with

the known behaviour of the system. There are usually good reasons for this outcome, so do not give up! Given that the parameter base for distributed models is generally small relative to the detail simulated, it is perhaps not surprising. Similarly, lumped models have a sparse parameter representation relative to natural variability in the system. Point measurements or spatial averages are often poor representations of the parameter interactions in this case. Evaluating errors from these sources will be dealt with later.

However, in terms of model parameterization, it may be impossible to return to the field to carry out more measurements, but we still need to obtain results for our model application. Thus, we need to adopt an iterative approach to the evaluation of the correct model parameters. This procedure is generally known as model calibration.

2.2.3 Calibration and its limitations

Kirkby *et al.* (1992) distinguish between physical parameters, which define the physical structure of the system under study, and process parameters, which define the order of magnitude of processes. Most models will contain both types of parameter. Definition of these process parameters is known as calibration or model tuning. Where they are physically based this definition can be achieved by their measurement; otherwise they are calibrated using a process of optimization (optimized) against a measure of the agreement between model results and a set of observations used for calibration. The calibration dataset must be independent from any dataset which is used later to validate the model, if the same dataset is used for both it should be no surprise that the model is a perfect predictor! Split-sample approaches, in which the available data is separated into a calibration set and a separate validation set, is usually the solution to this problem.

Calibration should pay particular attention to the sensitivity of parameters with sensitive parameters being calibrated carefully against high quality datasets to ensure that the resulting model will produce reliable outcomes. The simplest form of optimization is trial and error whereby model parameters are altered and a measure of goodness of fit between model results and the calibration dataset is noted. This process is repeated iteratively to obtain the best possible fit of observed against predicted. Of course the calibration will be specific to the model results calibrated against and will produce a model, which should forecast this result well at the expense of other model outputs not involved in the calibration procedure. The choice of calibration parameters, measures and techniques will thus depend upon the purpose to which the model will be put. Moreover a model calibration by one user with a particular understanding of its function may be quite different from that of another (Botterweg, 1995) and a model calibrated to a particular objective such as the prediction of peak runoff may then be useless in the prediction of total annual runoff. Some prior knowledge of, for example, the reasonable ranges of parameter values, will also be necessary and calibration will usually follow a preliminary sensitivity or uncertainty analysis, which is performed to test the validity of the model. The relationship between the range of values for a parameter and the model agreement is known as the calibration curve for that parameter. A parameter that shows a significant change in error with a change in its value (with all other parameters held constant) is known as a sensitive parameter. If a model has only one parameter, it is usually fairly straightforward to find the optimal value for that parameter. This procedure becomes only marginally more difficult for models with more than one parameter where the parameters are independent. In most models, parameters are highly interdependent and this will confound the definition of an optimum parameterization. In these cases other – automated – techniques are used to define the optimum parameter set. These techniques include genetic algorithms and fuzzy logic approaches as used to calibrate a rainfall runoff model to multiple objectives (peak discharge, peak time and total run off volume) by Cheng *et al.* (2002). Calibration is particularly challenging in distributed models, which tend to have a large number of parameters and stochastic algorithms or evolutionary genetic ones seem to be the most successful approaches under these circumstances (Eckhardt and Arnold, 2001) and have been applied widely where there are multiple objectives of the calibration (Madsen, 2000; 2003). In distributed models, there may also be advantages of calibrating different areas, such as subcatchments, separately and independently rather than as an integrated whole (e.g. Seibert *et al.*, 2000). Ratto *et al.* (2001) highlight the utility of the global sensitivity analysis (GSA) and generalized likelihood uncertainty estimation (GLUE) approaches (see below) in the calibration of over-parameterized models with strong parameter interaction. Global sensitivity analysis is a model-independent approach, which is based on estimating the fractional contribution of each input factor to the variance in the model output, accounting also for interaction terms. GLUE allows model runs to be classified according to the likelihood of their being a good simulator of the system recognizing that many different

combinations of parameter values can produce accurate model results (the issue of equifinality). By applying GSA to the GLUE likelihoods, the parameter sets driving model simulations with a good fit to observations are identified along with the basic features of the parameter interaction structure. Extensive model calibration tends to remove the physical basis of a model. Part of the objective in building a physically based model should be to produce a sufficiently good model structure and conceptualization to avoid the need for substantial calibration.

2.2.4 Top-down or bottom-up?

For a long time the focus in science has been on reductionism, splitting nature into manageable pieces (specialization) in order to understand how the individual parts work. This top-down approach assumes that it is best to start with the conceptualization of the whole system and the component parts in one go, and is most recently typified by the Earth-System Science approach (see discussion in Wainwright, 2009). The top-down approach has led to significant progress in many areas but its reductionist focus on the component parts means that, paradoxically, it does not allow us to understand whole systems well – and such an understanding is necessary to be able to understand and manage the environment effectively. Now, modelling and computers are enabling us, for the first time, to put these pieces together and understand their interactions and emergent properties as a (process, spatial and temporal) whole. This bottom-up capability is critical to understanding response to environmental change. (Note that Young and Leedal, in Chapter 7, reflect hydrological usage in that the terminology of 'top-down' and 'bottom-up' is used in the opposite sense from other areas of environmental modelling – see also Chapter 18).

As an example, single-domain models simulate the response of a process to a set of drivers – for example, a model may represent the fact that land-use change leads to an increase in soil erosion (see Chapters 15 and 22). Conversely, integrated models include all relevant processes and feedbacks, e.g. land-use change leads to increased soil erosion which leads to a decline in soil suitability and thus a decline in agricultural productivity and profit hence producing a further land-use change. The single-domain model, through its reductionist approach, is not able to represent this complex response. Most systems respond to change with a negative feedback so integrated models tend to produce less dramatic outcomes than single domain models.

2.3 Testing models

The terms 'verification' and 'validation' have very specific meanings in computer science. Verification is used to denote the process of checking that the computer code (program) does exactly what the algorithm is designed to do. As well as a formal confirmation, the procedure also involves the removal of coding bugs that have crept into the program during its writing (due to typing mistakes as well as misconceptualizations) as opposed to real bugs (moths etc.) that would get into relays and other early engineering structures causing the same kind of havoc as software bugs do today (which thus explains the origin of the term). Validation on the other hand refers to the testing of the model output to confirm the results that should be produced for the same inputs in reality (Fishman and Kiviat, 1968). One common method of validation is the comparison of a numerical model against the analytical solution for specific boundary conditions or against field measured data for the period and place of the model simulation.

However, Oreskes *et al.* (1994) pointed out that the difference between these specific uses and the common usage of the same terms can often lead to confusion, particularly when model results are being presented to nonmodellers. Rykiel (1996) suggests that the terms are essentially synonymous in everyday language, so the distinction is hard to see to a nonuser. Furthermore, the roots of the words may imply that a model is better than was actually intended when the author of a paper noted that the model was verified and validated. The root meaning of 'verify' comes from the Latin *verus*, meaning truth, while the Latin *validare* means to declare or give legal authority to something. Thus Oreskes *et al.* (1994) suggest that the nonmodeller may tend to feel a verified model presents the truth, and one that is validated can have legal authority, or is at least 'does not contain known or detectable flaws and is internally consistent'. They suggest that 'benchmarking' is a more appropriate term for verification and 'model evaluation' should be used in place of 'validation'. However, it could be argued that these are equally value-laden terms. In reality, most model output actually seems to generate a healthy dose of scepticism in nonmodellers (see the debate in Aber, 1997; 1998; Dale and Van Winkle, 1998; Van Winkle and Dale, 1998, for example). Lane and Richards (2001), on the other hand, suggest that validation is used as a linguistic means of hiding from such criticism.

Much more fundamentally in this debate, the nature of environmental systems and scientific practice means that

whatever a term is used for validation/model evaluation, it will always tend to overstate the case for belief in the model results. There are six reasons stated by Oreskes *et al.* (1994) for this problem. First, all environmental systems are open. Logically, it is only possible to demonstrate the truth of a closed system (although even this proposition is called into question by Gödel's theorem – see the excellent overview by Hofstadter, 1979). Secondly, there are problems due to the presence of unknown parameters and the scaling of nonadditive parameters (see below and Chapter 5). Thirdly, inferences and embedded assumptions underlie all stages of the observation and measurement of variables – dependent and independent alike. Fourthly, most scientific theories are developed by the addition of 'auxiliary hypotheses' – that is, those not central to the principal theory, but fundamental in a specific context for putting it into action. Thus, it is impossible to tell whether the principal or an auxiliary hypothesis is incorrect should deductive verification fail. Fifthly, as we have seen, more than one model formulation can provide the same output. This property is known formally as nonuniqueness or underdetermination (the Duhem–Quine thesis – Harding, 1976). Sixthly, errors in auxiliary hypotheses may cancel, causing incorrect acceptance of the model. Many modellers would now accept that full validation is a logical impossibility (e.g. Refsgaard and Storm, 1996, Senarath *et al.*, 2000). Morton and Suárez (2001) suggest that in most practical contexts the term 'model' can be thought of as synonymous with 'theory' or 'hypothesis', with the added implication that they are being confronted and evaluated with data. Often, the models represent simplifications of complex, physically based theories, analogies of other systems, summaries of data, or representations of the theories themselves. It is this set of approaches that allows the provisional nature of scientific knowledge to be tested. Conversely, it is possible for models to continue being used for a range of reasons relating to the social, economic and political contexts of science (Oreskes and Berlitz, 2001).

Rykiel (1996) provides an overview of how validation has been employed in modelling, and distinguishes (i) operational or whole-model validation (correspondence of model output with real-world observations); (ii) conceptual validation (evaluation of the underlying theories and assumptions); and (iii) data validation (evaluation of the data used to test the model). He suggests that there are at least 13 different sorts of validation procedure that are commonly employed, explicitly or implicitly. These procedures are:

- face validation – the evaluation of whether model logic and outputs appear reasonable;
- Turing tests – where 'experts' are asked to distinguish between real-world and model output (by analogy with the test for artificial intelligence);
- visualization techniques – often associated with a statement that declares how well the modelled results match the observed data;
- comparison with other models – used for example in general circulation model evaluations (although note the high likelihood of developing an argument based on circular logic here especially where different models share a common codebase!);
- internal validity – e.g. using the same data set repeatedly in a stochastic model to evaluate whether the distribution of outcomes is always reasonable;
- event validity – i.e. whether the occurrence and pattern of a specific event is reproduced by the model;
- historical data validation – using split-sample techniques to provide a subset of data to build a model and a second subset against which to test the model results (see also Klemeš, 1983);
- extreme-condition tests – whether the model behaves 'reasonably' under extreme combinations of inputs;
- traces – whether the changes of a variable through time in the model are realistic;
- sensitivity analyses – to evaluate whether changes in parameter values produce 'reasonable' changes in model output (see below);
- multistage validation (corresponding to the stages i, ii and iii noted above);
- predictive validation – comparison of model output with actual behaviour of the system in question; and
- statistical validation – whether the range of model behaviour and its error structure matches that of the observed system (but see the discussion on error propagation below).

Clearly, all of these tests provide some support for the acceptance of a model, although some are more rigorous than others. The more tests a model can successfully pass, the more confidence we might have in it, although there is still no reason to believe it absolutely for the reasons discussed above. But in complex models, validation is certainly a nontrivial procedure – Brown and Kulasiri (1996: 132) note, for example that 'a model can be considered to be successfully validated if all available techniques fail to distinguish between field and model data'. Any model test will in part be evaluating the simplifications upon which the model is based, in part

the reasonableness of its parameterization and in part to what extent the model output and test data are the same property. If we model soil moisture over 1-km^2 pixels and evaluate those on the basis of soil moisture measured at a point in that 1 km^2 to what extent are we comparing apples with apples? However, if our comparators are reasonable and a number of parameterizations fail for a specific model, we might seriously reconsider the model's conceptual basis. As with other aspects of modelling, evaluation is an iterative process.

2.4 Sensitivity analysis and its role

Sensitivity analysis is the process of defining how changes in model input parameters affect the magnitude of changes in model output. Sensitivity analysis is usually carried out as soon as model coding is complete and at this stage it has two benefits: to act as a check on the model logic and the robustness of the simulation and to define the importance of model parameters and thus the effort which must be invested in data acquisition for different parameters. The measurement of the sensitivity of a model to a parameter can also be viewed relative to the uncertainty involved in the measurement of that parameter in order to understand how important this uncertainty will be in terms of its impact on the model outcomes. If sensitivity analysis at this stage indicates that the model has a number of parameters to which the model is insensitive then this may indicate over-parameterization and the need for further simplification of the model.

Sensitivity analysis is usually also carried out when a model has been fully parameterized and is often used as a means of learning from the model by understanding the impact of parameter forcing, and its cascade through model processes to impact upon model outputs (see for example Mulligan, 1996; Burke *et al*., 1998; Michaelides and Wainwright, 2002)). In this way the behaviour of aggregate processes and the nature of their interaction can be better understood. After calibration and validation, sensitivity analysis can also be used as a means of model experiment and this is very common in GCM studies where sensitivity experiments of global temperature to greenhouse forcing, to large-scale deforestation or large-scale desertification are common experiments. Sensitivity analysis is also used in this way to examine the impacts of changes to the model structure itself, its boundary or initial conditions or the quality or quantity of data on its output (for example, May and Roeckner, 2001).

The sensitivity of model parameters is determined by their role in the model structure and, if this role is a reasonable representation of their role in the system under study, then there should be similarities between the sensitivity of model output to parameter change and the sensitivity of the real system response to manipulation. Nevertheless one must beware of attributing the model sensitivity to parameter change as equivalent to the sensitivity of the real system to similar changes in input (see Baker, 2000).

The methods of sensitivity analysis are covered in some detail by Hamby (1994) and, more recently, by Saltelli *et al*. (2000) and will not be outlined in detail here. In most sensitivity analyses a single parameter is varied incrementally around its normal value, keeping all other parameters unaltered. The model outputs of interest are monitored in response to these changes and the model sensitivity is usually expressed as the proportional change in the model output per unit change in the model input. In Figure 2.2 we show an example sensitivity analysis of a simple soil-erosion model, first in terms of single parameters and then as a multivariate sensitivity analysis. The former demonstrates the relative importance of vegetation cover, then slope, runoff and finally soil erodibility in controlling the amount of erosion according to the model. The multivariate analysis suggests that spatially variable parameters can have significant and sometimes counterintuitive impacts on the sensitivity of the overall system. A sensitive parameter is one that changes the model outputs of interest significantly per unit change in its value and an insensitive parameter is one which has little effect on the model outputs of interest (though it may have effects on other aspects of the model). Model sensitivity to a parameter will also depend on the value of other model parameters, especially in systems where thresholds operate, even where these remain the same between model runs. It is important to recognize the different propensities for parameter change in sensitivity analysis – that is, a model can be highly sensitive to changes in a particular parameter but if changes of that magnitude are unlikely ever to be realized then the model sensitivity to them will be of little relevance. In this way, some careful judgement is required of the modeller to set the appropriate bounds for parameter variation and the appropriate values of varying or non varying parameters during the process of sensitivity analysis.

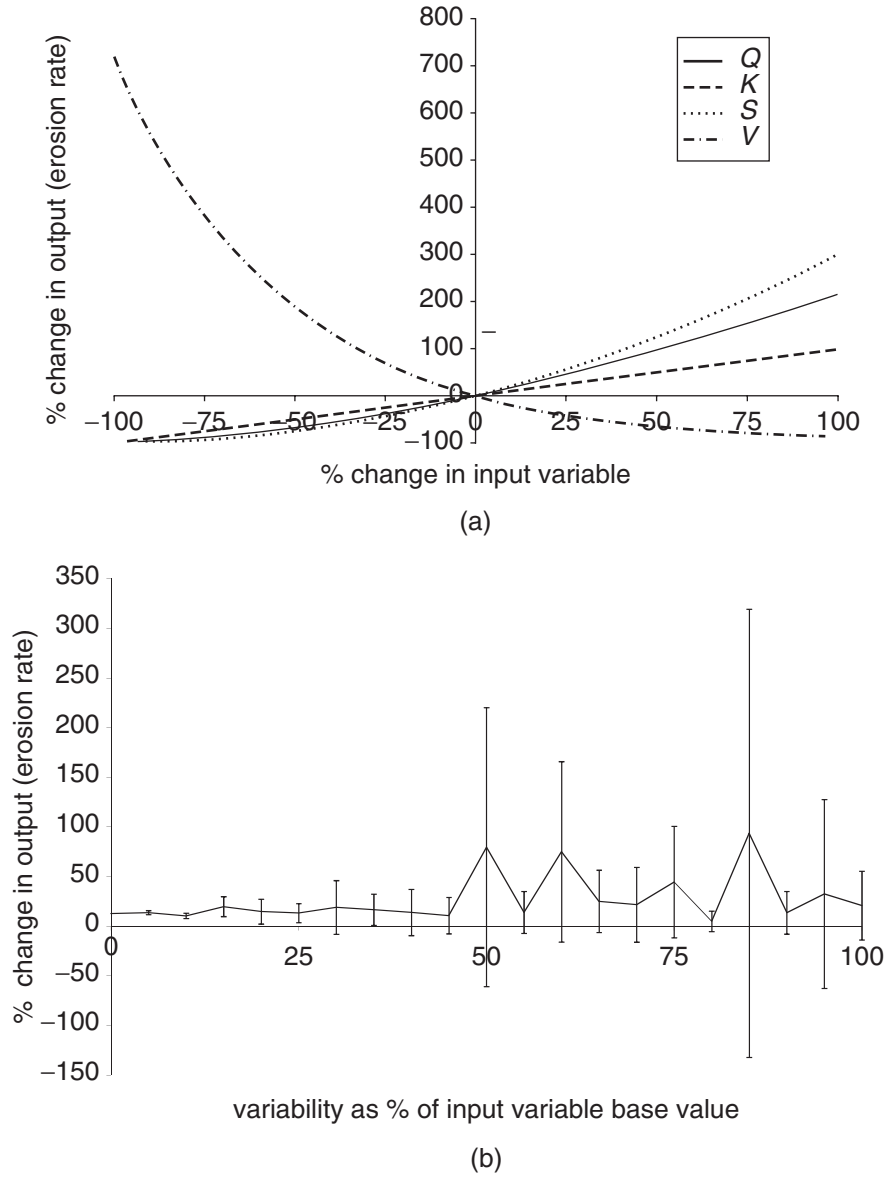


Figure 2.2 Example sensitivity analysis of the simple erosion model $E = kQ^m S^n e^{-iV}$ (where: E = erosion [mm month^{-1}], k = soil erodibility, Q = overland flow [mm month^{-1}], m = flow power coefficient [1.66], S = tangent of slope [m m^{-1}], n = slope constant [2.0], V = vegetation cover [%], i = vegetation erosion exponential function [dimensionless]): (a) univariate sensitivity analysis. Base values are $Q = 100$ mm month^{-1}, $k = 0.2$, $S = 0.5$, $m = 1.66$, $n = 2.0$, $i = 0.07$ and $V = 30\%$. The variables Q, k, S and V are varied individually from −100% to +100% of their base values and the output compared. Note that k has a positive linear response; Q a nonlinear response faster than k; S a nonlinear response faster than Q (because Q is raised to the power $m = 1.66$ while S is raised to the power $n = 2$); and V a negative exponential response. The order of parameter sensitivity is therefore $V > S > Q > k$.; and (b) Multivariate sensitivity analysis of the same model, where normally distributed variability is randomly added to each of the parameters as a proportion of the base value. Note the large fluctuations for large amounts of variability, suggesting that the model is highly sensitive where variability of parameters is >50% of the mean parameter value. No interactions or autocorrelations between parameter variations have been taken into account.

Sensitivity analysis is a very powerful tool for interacting with simple or complex models. Sensitivity analysis is used to:

(a) better understand the behaviour of the model, particularly in terms of the ways in which parameters interact;
(b) verify (in the computer-science sense) multicomponent models;
(c) ensure model parsimony by the rejection of parameters or processes to which the model is not sensitive;
(d) targeting field parameterization and validation programmes for optimal data collection focusing on the most sensitive inputs; and
(e) provide a means of better understanding parts of or the whole of the system being modelled.

Another form of sensitivity analysis is changing the model instead of changing the parameter values. A modeller might 'play' with different models or model configurations and thereby better understand the dynamics of the system before committing to a more sophisticated model-building adventure. This approach tends to be most effective with very simple (sometimes called 'toy') models or model components. An excellent example is described in detail in Chapter 16.

2.5 Errors and uncertainty

2.5.1 Error

No measurement can be made without error. (If you doubt this statement, get ten different people to write down the dimensions in mm of this page, without telling each other their measurements, and compare the results.) Although Heisenberg's uncertainty principle properly deals with phenomena at the quantum scale, there is always an element of interference when making an observation. Thus, the act of observation perturbs what we are measuring. Some systems may be particularly sensitive to these perturbations, for example when we introduce devices into a river to measure patterns of turbulence. The very act of placing a flow meter into the flow causes the local structure of flow to change. If we were interested in the river section discharge rather than the local changes in velocity, our single measuring device would have less significant impacts of perturbation, but the point measurement would be a very poor measurement of the cross-section flow. To counter this problem, we may repeat the measurement at a number of positions across the cross section and provide an average discharge (usually weighting by the width of flow represented by each measurement). But this average will only be as good as the choice of positions taken to represent the flow. Sampling theory suggests that a greater number of measurements will provide a better representation, with the standard error decreasing with the square root of the number of measurements made. However, a larger number of samples will take a longer time to make, and thus we have possible temporal changes to contend with in giving added error. Clearly, this approach is impractical when flows are rapidly changing. If we require continuous measurements, we may build a structure into the flow, such as a flume or weir (e.g. French, 1986) which again perturbs the system being measured (possibly with major effects if the local gradient is modified significantly, or if sediment is removed from the system). Furthermore the devices used to measure flow through these structures will have their own uncertainties of measurement, even if they are state-of-the-art electronic devices. In the flume case, the measurement is usually of a depth, which is then calibrated to a discharge by use of a rating curve. The rating curve itself will have an uncertainty element, and is usually proportional to depth to a power greater than one. Any error in measurement of depth will therefore be amplified in the estimation of discharge. Although it might be possible to measure flow depth non-invasively (e.g. using an ultrasonic detector), unless the channel section is well controlled (by invasive means), there is likely to be an even larger error in the rating curve used (e.g. because the channel cross-section changes as the flow gets deeper due to erosion or deposition of sediment). Such measurements also tend to be costly, and the cost and disturbance therefore prevents a large number of measurements being taken in a specific area, which is problematic if we are interested in spatial variability or if we wish to simulate at larger (more policy and management relevant) scales.

Other environmental modelling questions might require even further perturbation of the system. For example, soil depth is a critical control of water flow into and through the soil and thus has an impact on other systems too, such as vegetation growth or mass movement. In a single setting, we might dig a soil pit. Even if we try to replace the soil as closely as possible in the order in which we removed it, there will clearly be a major modification to the local conditions (most notably through bulk-density changes, which often mean a mound is left after refilling; a lower bulk density means

more pore space for water to flow through, so we have significantly altered what we intended to observe). Even if we were foolish enough to want to use this approach at the landscape scale, it is clearly not feasible (or legal!), so we might use a spatial sample of auger holes. However, the auger might hit a stone and thus have difficulty penetrating to the soil base – in stony soils, we usually only reach into the upper part of the C horizon. We might therefore try a non-invasive technique such as ground-penetrating radar (GPR). GPR uses the measured reflections of transmitted radio-frequency waves (usually in the 25–1,200 MHz range) to provide 'observations' of the changes in dielectric constant in engineering structures such as building and bridges or in the ground. Dielectric constant is determined by changes in material density and water content. Where transmitted waves encounter a change in dielectric constant some energy passes through the interface and some is reflected. The reflected energy from a GPR transmitter is recorded on a nearby receiver with the time delay (in nanoseconds) between the transmission of the pulse and its receipt indicating the distance of the reflecting object from the transmitter-receiver array. In this way GPR can image the subsurface and has found application in archaeology (Imai *et al.*, 1987), hydrology (van Overmeeren *et al.*, 1997), glaciology (Nicolin and Koffman, 1994) and geology (Mellett, 1995). The difficulty with electromagnetic non-invasive techniques is that whilst dielectric discontinuities can be fairly obviously seen, the techniques provide little information on what these discontinuities are (rocks, roots or moisture for example). Thus, non-invasive techniques are also subject to significant potential error. The implication is that all measurements should have their associated error cited so that the implications can be considered and due care be taken in interpreting results. Care is particularly necessary in the use of secondary data, where one may have very little idea about how the data were collected and quality controlled.

Field measurements are often particularly prone to error, because of the difficulty of collecting data. We may choose to use techniques that provide rapid results but which perhaps provide less precise measurements, because of the high costs involved in obtaining field data. Note the difference between error and precision (sometimes called the tolerance of a measurement) – the latter relates only to the exactness with which a measurement is taken. A lack of precision may give very specific problems when measuring features with fractal characteristics, or when dealing with systems that are sensitive to initial conditions. Thus, a consideration of the modelling requirements is often important when deciding the precision of a measurement.

Specification errors can arise when what is being measured does not correspond to the conceptualization of a parameter in the model. This problem may arise if a number of processes are incorporated into a single parameter. For example, if erosion is being considered as a diffusion process in a particular model, diffusion may occur by a number of processes, including rainsplash, ploughing, animal activity and soil creep. The first two might be relatively easy to measure, albeit with their own inherent problems (e.g. Torri and Poesen, 1988; Wainwright *et al.*, 2008), while the latter may be more difficult to quantify either because of inherent variability in the case of bioturbation or because of the slow rate of the process in the case of creep. Interactions between the different processes may make the measurement of a compound diffusion parameter unreliable. It is also possible that different ways of measurement, apparently of the same process, can give very different results. Wainwright *et al.* (2000) illustrate how a number of measurements in rainfall-runoff modelling can be problematic, including how apparent differences in infiltration rate can be generated in very similar rainfall-simulation experiments. Using pumps to remove water from the surface of the plot led to significant overestimation of saturated infiltration because of incomplete recovery of ponded water, when compared to direct measurement of runoff from the base of the plot (which itself incorporates ponding into the amount of infiltration and thus also overestimates the real rate). The pumping technique also introduces a significant time delay to measurements, so that unsaturated infiltration is very poorly represented by this method. Differences between infiltration measured using rainfall simulation, cylinder infiltration and the falling-head technique from soil cores for the same location can be orders of magnitude (e.g. Wainwright, 1996) because each are representing infiltration in different ways. Different infiltration models may be better able to use measurements using one technique rather than another. Such specification errors can be very difficult to quantify, and may in fact only become apparent when problems arise during the modelling process. It should always be borne in mind that errors in model output may be due to incorrect parameter specification. When errors occur, it is an important part of the modelling process to return to the parameters and evaluate whether the errors could be caused in this way.

Environmental models operate in a defined space in the real world. However, the representation of that space will always be some form of simplification. At the

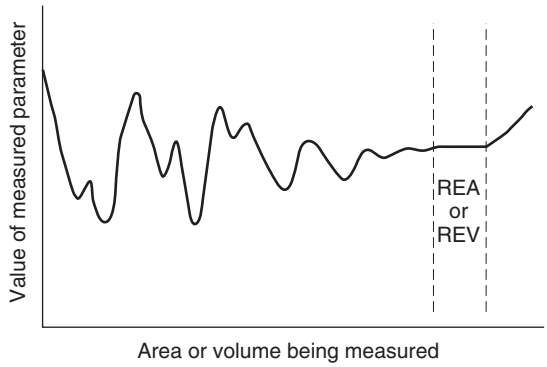


Figure 2.3 Definition of the representative elemental area (REA) or volume (REV) concept.

extreme case, the system will be completely lumped, with single values for parameters and each input and output. Such models can be a useful generalization, for example in the forecasting of flooding or reservoir filling (e.g. Blackie and Eeles, 1985 – another example is the population models discussed above, although these can be spatialized as shown by Thornes, 1990). However, the definition of each parameter may be non-trivial for all but the simplest of catchments. Wood *et al.* (1988) used the term representative elemental area (REA) to evaluate the scale at which a model parameter might be appropriate (Figure 2.3). At the opposite end of the spectrum is the fully distributed model, in which all parameters are spatialized. There still remains the issue of the REA in relation to the grid size used (distributed applications may still have grid sizes of kilometres – or hundreds of kilometres in the case of General Circulation Models). However, in addition there is the issue of how to estimate parameters spatially. Field measurements are costly so that extensive data collection may be impossible financially, even if the people and equipment were available on a sufficient scale. Therefore, it is usual to use some sort of estimation technique to relate parameters to some easily measured property. For example, Parsons *et al.* (1997) used surface stone cover to estimate infiltration rates, finding that the spatial structure provided by this approach gave a better solution than simply assigning spatial values based on a distribution function of measured infiltration rates, despite the relatively high error in the calibration between stone cover and infiltration rate. Model sensitivity to different parameters (see above) may mean that different techniques of spatialization are appropriate for the parameters of the model in question. It is important that distributed models are tested with spatially distributed data, otherwise it is possible to arrive at completely different conclusions about their reliability (Grayson *et al.*, 1992a, b). Similarly, if the scale of the output is not related to the scale of the test data, errors in interpretation can arise. It is for this reason that techniques of upscaling or downscaling of model results are important (see chapter 20). For further details on the technicalities of different types of error, refer to Engeln-Müllges and Uhlig (1996) and Mulligan and Wainwright (2012).

2.5.2 From error to uncertainty

Zimmerman (2000) defines six causes of uncertainty in the modelling process: lack of information, abundance of information, conflicting evidence, ambiguity, measurement and belief. *A lack of information* requires us to collect more information, but it is important to recognize that the *quality* of the information also needs to be appropriate. It must be directed towards the modelling aims and may require the modification of the ways in which parameters are conceived of and collected. *Information abundance* relates to the complexity of environmental systems and our inability to perceive large amounts of complex information. Rather than collecting new data, this cause of uncertainty requires the simplification of information, perhaps using statistical and data-mining techniques. *Conflicting information* requires the application of quality control to evaluate whether conflicts are due to errors or are really present. Conflicts may also point to the fact that the model being used is itself wrong, so re-evaluation of the model structure and interaction of components may be necessary. *Ambiguity* relates to the reporting of information in a way that may provide confusion. Uncertainty can be removed by questioning the original informant, although this approach may not be possible in all cases. *Measurement uncertainty* may be reduced by invoking more precise techniques, although it must be done in an appropriate way. There is often a tendency to assume that modern gadgetry will allow measurement with fewer errors. It must be noted that other errors can be introduced (e.g. misrecording of electronic data if a data logger gets wet during a storm event) or that the new measurement may not be measuring exactly the same property as before. *Beliefs* about how data are to be interpreted can also cause uncertainty because different outcomes can result from the same starting point. Overcoming this uncertainty is a matter of iterative testing of the different belief structures. Qualitative assessment is thus as much an aspect of uncertainty assessment as qualitative analysis.

The quantitative evaluation of uncertainty has been discussed above in detail. Error-propagation techniques can be used in relatively simple models (or their subcomponents) to evaluate the impact of an input error (measurement uncertainty) on the outcome. In more complex scenarios, Monte Carlo analysis is almost certainly necessary. If sufficient runs are performed, then probability estimates can be made about the outputs. Carrying out this approach on a model with a large number of parameters is a nontrivial exercise, and requires the development of appropriate sampling designs (Parysow *et al.*, 2000). The use of sensitivity analysis can also be used to optimize this process (Klepper, 1997, Bärlund and Tattari, 2001). Hall and Anderson (2002) note that some applications may involve so much uncertainty that it is better to talk about *possible* outcomes rather than give specific probabilities. Future scenarios of climate change evaluated using General Circulation Models (see Chapters 9 and 18) is a specific case in point here. Another approach that can be used to evaluate the uncertainty in outcome as a function of uncertain input data is fuzzy set theory. Torri *et al.* (1997) applied this approach to the estimation of soil erodibility in a global dataset and Özelkan and Duckstein (2001) have applied it to rainfall-runoff modelling. Because all measurements are uncertain, the data used for model testing will also include errors. It is important to beware of rejecting models because the evaluation data are not sufficiently strong to test it. Monte *et al.* (1996) presented a technique for incorporating such uncertainty into the model-evaluation process.

Distributed models may require sophisticated visualization techniques to evaluate the uncertainty of the spatial data used as input (Wingle *et al.*, 1999). An important consideration is the development of appropriate spatial and spatio-temporal indices for model evaluation, based on the fact that spatial data and their associated errors will have autocorrelation structures to a greater or lesser extent. Autocorrelation of errors can introduce significant nonlinear effects on the model uncertainty (Henebry, 1995).

Certain systems may be much more sensitive to the impacts of uncertainty. Tainaka (1996) discusses the problem of spatially distributed predator-prey systems, where there is a phase transition between the occurrence of both predator and prey, and the extinction of the predator species. Such transitions can occur paradoxically when there is a rapid increase in the number of the prey species triggered by instability in nutrient availability, for example. Because the phase transition represents a large (catastrophic) change, the model will be very sensitive to uncertainty in the local region of the parameter space, and it can thus become difficult or impossible to interpret the cause of the change.

2.5.3 Coming to terms with error

Error is an important part of the modelling process (as with any scientific endeavour). It must therefore be incorporated within the framework of any approach taken, and any corresponding uncertainty evaluated as far as possible. A realistic approach and a healthy scepticism to model results are fundamental. It is at best misleading to present results without corresponding uncertainties. Such uncertainties have significant impacts on model applications, particularly the use of models in decision making. Large uncertainties inevitably lead to the rejection of modelling as an appropriate technique in this context (Beck, 1987). Recent debates on possible future climate change reinforce this conclusion (see the excellent discussion in Rayner and Malone, 1998).

In terms of modelling practice, it is here that we come full circle. The implication of error and uncertainty is that we need to improve the basic inputs into our models. As we have seen, this improvement does not necessarily just mean collecting more data. It may mean that it is better to collect fewer samples, but with better control. Alternatively, it may be necessary to collect the same number of samples, but with a more appropriate spatial and/or temporal distribution. Ultimately, the iterations involved in modelling should not just be within computer code, but also between field and model application and testing.

2.6 Conclusions

Modelling provides a variety of tools with which we can increase our understanding of environmental systems. In many cases, this understanding is then practically applied to real-world problems. It is thus a powerful tool for tackling scientific questions and answering (green!) engineering problems. But its practice is also something of an art, which requires intuition and imagination to achieve appropriate levels of abstraction from the real world to our ability to represent it in practical terms. As a research tool, it provides an important link between theory and observation, and provides a means of testing our ideas of how the world works. This link is important in that environmental scientists generally deal with temporal and spatial scales that are well beyond the limits of observation

of the individual. It is important to recognize that modelling is not itself a 'black box' – it forms a continuum of techniques that may be appropriately applied in different contexts. Part of the art of modelling is the recognition of the appropriate context for the appropriate technique.

Hopefully, we have demonstrated that there is not one single way of implementing a model in a specific context. As we will see in the following chapters, there are many different ways to use models in similar settings. How the implementation is carried out depends on a range of factors including the aims and objectives of the study, the spatial and temporal detail required, and the resources available to carry it out. This chapter has presented the range of currently available techniques that may be employed in model building and testing. More details of these techniques are found on the book website and in the numerous references provided. We have also addressed a number of the important conceptual issues involved in the modelling process, again to suggest that the modelling process is not as uniform as is often assumed, and that debate is often fruitful within and between disciplines in order to tackle some of these issues.

References

Aber, J. (1997) Why don't we believe the models? *Bulletin of the Ecological Society of America*, **78**, 232–3.

Aber, J. (1998) Mostly a misunderstanding, I believe. *Bulletin of the Ecological Society of America*, **79**, 256–7.

Baker, V.R. (1998) Paleohydrology and the hydrological sciences, in *Palaeohydrology and Environmental Change* (eds G. Benito, V.R Baker and K.J. Gregory), John Wiley & Sons, Ltd, Chichester, pp. 1–10.

Baker, V.R. (2000) South American palaeohydrology: future prospects and global perspective. *Quaternary International*, **72**, 3–5.

Barahona, E. and Iriarte, U.A. (2001) An overview of the present state of standardization of soil sampling in Spain. *The Science of the Total Environment*, **264**, 169–74

Bärlund, I. and Tattari, S. (2001) Ranking of parameters on the basis of their contribution to model uncertainty. *Ecological Modelling*, **142**, 11–23.

Beck, M.B. (1987) Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research*, **23**, 1393–442.

Beven, K. (1992) Future of distributed modelling (editorial). *Hydrological Processes* **6**, 279–98.

Blackie, J.R. and Eeles, C.W.O. (1985) Lumped catchment models, in *Hydrological Forecasting* (eds M.G. Anderson and T.P. Burt), John Wiley & Sons, Ltd, Chichester, pp. 311–45.

Botterweg, P. (1995) The user's influence on model calibration results: an example of the model SOIL, independently calibrated by two users. *Ecological Modelling*, **81**, 71–81.

Brown, A.G. (2011) The Anthropocene – a sedimentary perspective. *Geoscientist*, July, 22.

Brown, T.N. and Kulasiri, D. (1996) Validating models of complex, stochastic, biological systems. *Ecological Modelling*, **86**, 129–34.

Bueso, M.C., Angulo, J.M. and Alonso, F.J. (1998) A state-space model approach to optimum spatial sampling design based on entropy. *Environmental and Ecological Statistics*, **5**, 29–44.

Burke, S., Mulligan, M. and Thornes, J.B. (1998) Key parameters that control recharge at a catchment scale: using sensitivity analysis and path analysis to define homogenous recharge classes, in *Hydrology in a Changing Environment. Volume II* (H.S. Wheater and C. Kirby), John Wiley & Sons, Ltd, Chichester, pp. 229–43.

Chao, C.T. and Thompson, S.K. (2001) Optimal adaptive selection of sampling sites. *Environmetrics* **12**, 517–38.

Cheng, C.T., Ou, C.P. and Chau, K.W. (2002) Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology*, **268**, 72–86.

Cohn, J.P. (2002) Biosphere 2: turning an experiment into a research station. *BioScience*, **52**, 218–23.

Collins, L. (1975) *An Introduction to Markov Chain Analysis*. Geo Abstracts, Norwich.

Cross, M. and A.O. Moscardini, (1985) *Learning the Art of Mathematical Modelling* John Wiley & Sons, Chichester.

Crutzen, P.J. (2002) Geology of mankind. *Nature*, **415**, 23.

Dale, V.H. and Van Winkle, W. (1998) Models provide understanding, not belief. *Bulletin of the Ecological Society of America*, **79**, 169–70.

Eckhardt, K. and Arnold, J.G. (2001) Automatic calibration of a distributed catchment model. *Journal of Hydrology*, **251**, 103–9.

Fishman, G.S. and Kiviat, P.J. (1968) The statistics of discrete–event simulation. *Simulation*, **10**, 185–91.

French, R.H. (1986) *Open–Channel Hydraulics*, McGraw Hill, New York.

Govindasamy, B., Caldeira, K. and Duffy, P.B. (2003) Geoengineering Earth's radiation balance to mitigate climate change from a quadrupling of CO_2. *Global and Planetary Change*, **771**, 1–12.

Grayson, R.B., Moore, I.D. and McMahon, T.A. (1992a) Physically based hydrologic modelling. 1. A terrain-based model for investigative purposes. *Water Resources Research*, **28**, 2639–58.

Grayson, R.B., Moore, I.D. and McMahon, T.A. (1992b) Physically based hydrologic modelling. 2. Is the concept realistic? *Water Resources Research*, **28**, 2659–66.

Hall, G.W (2001) Elevated atmospheric CO_2 alleviates drought stress in wheat. *Agriculture, Ecosystems and Environment*, **87**, 261–71.

Hall, J. and Anderson, M.G. (2002) Handling uncertainty in extreme or unrepeatable hydrological processes – the need for an alternative paradigm. *Hydrological Processes*, **16**, 1867–70.

Hamby, D.M. (1994) A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, **32**, 135–54.

Hansen, D.P., Ye, W., Jakeman, A.J. (1996) Analysis of the effect of rainfall and streamflow data quality and catchment dynamics on streamflow prediction using the rainfall-runoff model IHACRES. *Environmental Software*, **11**, 193–202.

Harding, S. (1976) *Can Theories Be Refuted? Essays on the Duhem–Quine Thesis*. Reidel, Dordrecht.

Harte, J. (1985) *Consider a Spherical Cow. A Course in Environmental Problem Solving*. William Kaufmann Inc., Los Altos, CA.

Haykin, S. (1999) *Neural Networks. A Comprehensive Foundation*, 2nd edn, Prentice Hall, Upper Saddle River, NJ.

Henebry, G.M. (1995) Spatial model error analysis using autocorrelation indices. *Ecological Modelling*, **82**, 75–91.

Hofstadter, D.R. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*, Penguin Books, Harmondsworth.

Imai T., Sakayama, T. and Kanemori, T. (1987) Use of ground-probing radar and resistivity surveys for archaeological investigations. *Geophysics*, **52**, 137–50.

Kirkby, M.J., Naden, P.S., Burt, T.P. and Butcher, D.P. (1992) *Computer Simulation in Physical Geography*. John Wiley & Sons, Chichester.

Klemeš, V. (1983) Conceptualization and scale in hydrology. *Journal of Hydrology*, **65**, 1–23.

Klemeš, V. (1997) Of carts and horses in hydrological modelling. *Journal of Hydrologic Engineering*, **1**, 43–9.

Klepper, O. (1997) Multivariate aspects of model uncertainty analysis: tools for sensitivity analysis and calibration. *Ecological Modelling*, **101**, 1–13.

Lane, S.N. and Richards, K.S. (2001) The 'validation ' of hydrodynamic models: some critical perspectives, in *Model Validation: Perspectives in Hydrological Science* (eds M.G. Anderson and P.D. Bates), John Wiley & Sons, Chichester, pp. 413–38.

Lark, R.M. (2001) Some tools for parsimonious modelling and interpretation of within-field variation of soil and crop systems. *Soil and Tillage Research*, **58**, 99–111.

Madsen, H. (2000) Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. *Journal of Hydrology*, **235**, 276–88.

Madsen, H. (2003) Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, **26**, 205–16.

May, W. and Roeckner, E. (2001) A time-slice experiment with the ECHAM 4 AGCM at high resolution: the impact of horizontal resolution on annual mean climate change. *Climate Dynamics*, **17**, 407–20.

Mellett, J.S. (1995) Ground-penetrating radar applications in engineering, environmental management and geology. *Journal of Applied Geophysics*, **33**, 157–66.

Michaelides, K. and Wainwright, J. (2002) Modelling the effects of hillslope-channel coupling on catchment hydrological response. *Earth Surface Processes and Landforms*, **27**, 1441–58.

Mode, N.A., Conquest, L.L. and Marker, D.A. (1999) Ranked set sampling for ecological research: accounting for the total costs of sampling. *Environmetrics* **10**, 179–94.

Mode, N.A., L.L. Conquest and Marker, D.A. (2002) Incorporating prior knowledge in environmental sampling: ranked set sampling and other double sampling procedures. *Environmetrics*, **13**, 513–21.

Monte, L., Håkanson, L., Bergström, U. *et al.* (1996) Uncertainty analysis and validation of environmental models: the empirically based uncertainty analysis. *Ecological Modelling*, **91**, 139–52.

Morton, A. and Suárez, M. (2001) Kinds of models, in *Model Validation: Perspectives in Hydrological Science* (eds M.G. Anderson and P.D. Bates), John Wiley & Sons, Ltd, Chichester, pp. 11–21.

Mulligan, M. (1996) Modelling the complexity of landscape response to climatic variability in semi arid environments, in *Advances in Hillslope Processes* (eds M.G. Anderson and S.M. Brooks), John Wiley & Sons, Ltd, Chichester, pp. 1099–149.

Mulligan, M. (1998) Hydrology in a varying as well as changing environment, in *Hydrology in a Changing Environment. Volume II* (eds H.S. Wheater and C. Kirby), John Wiley & Sons, Ltd, Chichester, pp. 591–604.

Mulligan, M. and Wainwright, J. (2012) *Building Environmental Models: A Primer on Simplifying Complexity*, John Wiley & Sons, Ltd, Chichester.

Musters, P.A.D. and Bouten, W. (2000) A method for identifying optimum strategies of measuring soil water contents for calibrating a root water uptake model. *Journal of Hydrology*, **227**, 273–86.

Nicollin, F. and Kofman, W. (1994) Ground-penetrating radar sounding of a temperate glacier – modelling of a multilayered medium. *Geophysical Prospecting*, **42**, 715–34.

Oreskes, N. and Berlitz, K. (2001) Philosophical issues in model assessment, in *Model Validation: Perspectives in Hydrological Science* (eds M.G. Anderson and P.D. Bates), John Wiley & Sons, Ltd, Chichester, pp. 23–41.

Oreskes, N., Shrader-Frechette, K. and Bellitz, K. (1994) Verification, validation and confirmation of numerical models in the Earth sciences. *Science*, **263**, 641–6.

Özelkan, E.C. and Duckstein, L. (2001) Fuzzy-conceptual rainfall-runoff models. *Journal of Hydrology*, **253**, 41–68.

Parsons, A.J., J. Wainwright, Abrahams, A.D. and Simanton, J.R. (1997) Distributed dynamic modelling of interrill overland flow. *Hydrological Processes*, **11**, 1833–59.

Parysow, P., Gertner, G. and Westervelt, J. (2000) Efficient approximation for building error budgets for process models. *Ecological Modelling*, **135**, 111–25.

Ratto, M., Tarantola, S. and Saltelli, A. (2001) Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communications*. **136**, 212–24.

Rayner, S. and Malone, E.L. (1998) *Human Choice and Climate Change. Volume 1*. Batelle Press, Columbus, OH.

Refsgaard, J.C. and Storm, B. (1996) Construction, calibration and validation of hydrological models, in *Distributed Hydrological Modelling* (eds M.B. Abbott and J.C. Refsgaard), Reidel, Dordrecht, pp. 41–5.

Ruddiman, W.F. (2007) *Earth's Climate: Past and Future*, 2nd edition, W.H. Freeman, New York.

RykielJr, E.J. (1996) Testing ecological models: the meaning of validation. *Ecological Modelling*, **90**, 229–44.

Saltelli, A., Chan, K. and Scott, M. (2000) *Sensitivity Analysis*, John Wiley & Sons, Ltd, Chichester.

Schaetzen, W.B.F. de, Walters, G.A. and Savic, D.A. (2000) Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms. *Urban Water*, **2**, 141–52.

Seibert, J., Uhlenbrook, S., Leibundgut, C. and Halldin, S. (2000) Multiscale calibration and validation of a concenptual rainfall runoff model. *Physics and Chemistry of the Earth B*, **25**, 59–64.

Senarath, S.U.S., Ogden, F., Downer, C.W. and Sharif, H.O. (2000) On the calibration and verification of two-dimensional, distributed, Hortonian, continuous watershed models. *Water Resources Research*, **36**, 1495–510.

Steffen, W., Crutzen, P., McNeill, J.R. (2007) The Anthropocene: are humans now overwhelming the great forces of Nature? *Ambio*, **36**, 614–21.

Tainaka, K.-I. (1996) Uncertainty in ecological catastrophe. *Ecological Modelling*, **86**, 125–8.

Thornes, J.B. (1990) The interaction of erosional and vegetational dynamics in land degradation: spatial outcomes, in *Vegetation and Erosion* (ed. J.B. Thornes), John Wiley & Sons, Ltd, Chichester.

Torri, D. and Poesen, J. (1988) The effect of cup size on splash detachment and transport measurements. Part II: theoretical approach, in *Geomorphic Processes in Environments with Strong Seasonal Contrasts. Volume 1 Hillslope Processes* (eds A. Imeson and M. Sala), Catena Supplement 12 Catena, Cremlingen, pp. 127–37.

Torri, D., Poesen, J. and Borselli, L. (1997) Predictability and uncertainty of the soil erodibility factor using a global dataset, *Catena*, **31**, 1–22.

Van Overmeeren, R.A., Sariowan, S.V., and Gehrels, J.C. (1997) Ground penetrating radar for determining volumetric water content; results of comparative measurements at two test sites. *Journal of Hydrology*, **197**, 316–38.

Van Wesemael, B., Mulligan, M. and Poesen, J. (2000) Spatial patterns of soil water balance on intensively cultivated hillslopes in a semi-arid environment: the impact of rock fragments and soil thickness. *Hydrological Processes*, **14**, 1811–28.

Van Winkle, W. and Dale, V.H. (1998) Model interactions: a reply to Aber, *Bulletin of the Ecological Society of America*, **79**, 257–9.

Wainwright, J. (1996) A comparison of the infiltration, runoff and erosion characteristics of two contrasting 'badland' areas in S. France. *Zeitschrift für Geomorphologie Supplementband*, **106**, 183–98.

Wainwright, J. (2009) Earth-system science, in *Blackwell Companion to Environmental Geography* (eds N. Castree, D. Liverman, B. Rhoads and D. Demerritt), Blackwell, Oxford, pp. 145–67.

Wainwright, J., Mulligan, M. and Thornes, J.B. (1999a) Plants and water in drylands, in Baird, A.J. and Wilby, R.L. (eds) *Ecohydrology*, Routledge, London, pp. 78–126.

Wainwright, J. and Parsons, A.J. (2002) The effect of temporal variations in rainfall on scale dependency in runoff coefficients. *Water Resources Research*, **38** (12), 1271. Doi: 10.1029/2000WR000188.

Wainwright, J., Parsons, A.J. and Abrahams, A.D. (2000) Plot-scale studies of vegetation, overland flow and erosion interactions: case studies from Arizona and New Mexico. *Hydrological Processes*, **14**, 2921–43.

Wainwright, J, Parsons, A.J., Müller, E.N. *et al*. (2008) A transport-distance approach to scaling erosion rates: 3. Evaluating scaling characteristics of Mahleran, *Earth Surface Processes and Landforms*, **33**, 1113–28.

Wingle, W. L., Poeter, E.P. and McKenna, S.A. (1999) UNCERT: geostatistics, uncertainty analysis and visualization software applied to groundwater flow and contaminant transport modelling. *Computers and Geosciences*, **25**, 365–76.

Wood, E.F., Sivapalan, M., Beven, K. and Band, L. (1988) Effects of spatial variability and scale with implications to hydrologic modelling. *Journal of Hydrology*, **102**, 29–47.

Zalasiewicz, J, Williams, M., Steffen, W. and Crutzen, P. (2010) The new world of the Anthropocene. *Environmental Science and Technology*, **44**, 2228–31.

Zimmermann, H.-J. (2000) An application-oriented view of modelling uncertainty. *European Journal of Operational Research*, **122**, 190–8.