

Towards a coherent philosophy for modelling the environment

BY KEITH BEVEN

University of Lancaster, Bailrigg, Lancaster LA1 4YW, UK

Received 16 May 2001; revised 15 January 2002; accepted 19 March 2002

The predominant philosophy underlying most environmental modelling is a form of pragmatic realism. The limitations of this approach in practical applications are discussed, in particular, in relation to questions of scale, nonlinearity and uniqueness of place. A new approach arising out of the concept of equifinality of models (structures and parameter sets) in application is outlined in the form of an uncertain ‘landscape space’ to model space mapping. The possibility of hypothesis testing within this framework is proposed as a means of refining the mapping, with a focus on the differentiation of function within the model space. The approach combines elements of instrumentalism, relativism, Bayesianism and pragmatism, while allowing the realist stance that underlies much of the practice of environmental modelling as a fundamental aim. It may be an interim philosophy that is awaiting developments in measurement technique to allow further refinement, but allows some coherent guidance about how to be specific in presenting predictions to end users.

Keywords: Author: please supply up to six keywords/phrases!

1. Why a new philosophy?

When looking for positive guidance from philosophy we must rest content with some vague generalizations about the need to be specific.

Alan Chalmers (1989, p. 23)

It is quite possible to develop and use environmental models without any explicit underlying philosophy (although see, for example, Peters (1991), Abbott (1992), Rhoads & Thorn (1996) and Anderson & Bates (2001) for discussions in different areas of environmental science). Many practitioners do, of course, although most might have the aim of developing and using models that are as ‘realistic as possible’, given the constraints of current knowledge, computing capabilities, observational technologies and available time. This type of implicit or pragmatic realism seems quite natural and appears to be intrinsic to modelling efforts as diverse as the coupled general circulation models (GCMs) being used to predict future climate change and modelling of the biogeochemical processes involved in the transport of radionuclides or persistent organic pollutants in the environment. The philosophical subtleties are not really necessary to the practicing environmental modeller who only needs to know that achieving realism is still difficult in the practical prediction of complex environmental systems. Current implementations of such models are known to have their

Beven 1996a?

limitations (even if the true level of uncertainty in model predictions is not widely appreciated), but it is implicitly accepted that they will continue to evolve towards more realistic representations of the Earth–ocean–atmosphere–biosphere system.

It will be argued in this paper that while this pragmatic realism has served environmental science well, it does have major flaws and should be replaced. Experience suggests that philosophical arguments alone will have little impact on the way in which modelling is actually done, but there is, perhaps, at least the start of a recognition of the need for a more scientifically robust assessment of model capabilities and predictions. The view adopted here that a new philosophical approach to environmental modelling is needed stems not from any deep notions of environmentalism, nor from any higher level holistic principles of nature, nor directly from philosophical arguments about the concept(s) of realism itself, but from experience of practical applications of modelling in various areas of environmental science. In what follows, a coherent philosophical framework for environmental modelling is outlined that might satisfy this need while allowing for model structural error and the incorporation of improved knowledge over time.

In § 2 the theory and practice of environmental modelling are briefly summarized, together with some comments on the impact of complexity and unknowability on the process of formulating and evaluating environmental models. Section 3 addresses the problems posed by uniqueness of place in the application of environmental models. This is an issue that is becoming increasingly important as the rapid increases in computer power allow environmental models of whole regions or countries (or, albeit still crudely, the globe) to become feasible. Section 4 places such applications into a framework of landscape space to model space mapping. Section 5 shows how this can be used to constrain predictive uncertainty, including by a process of hypothesis formulation and testing. Section 6 considers the question of explanatory depth of approximate models in predicting the unexpected, and § 7 returns to the issue of realism in the face of the apparent relativism of landscape to model space mapping.

2. Theory and practice in environmental modelling

Computer simulation has become a common methodology in the environmental sciences. It can be used as a framework for formulating and testing theories. It can be used to make predictions for practical applications in response to demands from policy and decision makers. Particularly interesting examples are the predictions of the impacts of global climate change (e.g. Van Dam 1999, Cameron *et al.* 2000), predictions of groundwater systems (Konikow & Bredehoeft 1992; Anderson & Woessner 1992; Oreskes *et al.* 1994), the long-term predictions required in geomorphology (Richards *et al.* 1995; Beven 1996*a*) and predictions of the impacts of underground repositories for nuclear waste (Sellafield in the UK, Yucca Mountain in the US). Very large research programmes have been funded with a view to improving the accuracy of such predictions. There is an implicit belief underlying these efforts that this is possible, despite the nonlinear and open nature of the systems being studied and the approximations necessary to implement our qualitative understanding of these systems in a working computing program.

The foundation of this belief is the pragmatic realism noted above. I am a hydrologist. As a hydrologist, I intend that the computer simulation models I develop should represent real water; in modelling contaminant transport in flowing water I intend

that the model variables should represent the real contaminant. This is in spite of the fact that, as a hydrologist, I would recognize the varied nomological status of many of the concepts that are used in my models (Beven 1989, 1996b; Shrader-Frechette 1989; Davis et al. 1992; Hofmann & Hofmann 1992). In the same way, the atmospheric modeller will intend that the variables in her computer model should represent the real atmosphere and the aquatic geochemist will intend that the (approximate) computer solutions to systems of stoichiometric equations should represent real aqueous solutions. In time, new understanding and knowledge gained from experiments will be incorporated into improved simulations. In this way, the research programme of environmental science should be progressive (in the sense of Lakatos (1978)).

This common working version of realism is, of course, rather naive. It is naive not only from a philosophical point of view, but also from a scientific point of view, since it requires that the systems under study be knowable. Clearly, for many environmental systems, the perceived complexities are such that all the boundary conditions, auxiliary conditions and system characteristics cannot be knowable given current measurement technologies. Thus, in applying this pragmatic realism to any particular open environmental system, it is clear that we can recognize much more complexity than it is possible to represent in a mathematical model that, implemented as a computer simulation, will make quantitative predictions. We need to differentiate between this qualitative understanding (the *perceptual model*) gained from our training, qualitative observations, quantitative experiment or monitoring, and the practical application of that understanding in a *conceptual or formal model* representation of the system as a set of equations (Beven 2001a). This is not to deny that the perceptual model is not conditioned by fundamental theoretical understanding (even if some of the complexity may be difficult to relate to existing theoretical constructs), but rather to recognize that this understanding may be very difficult to apply at the scales of interest. The perceptual model may only be as realistic as current understanding allows. The conceptual model will, however, be wrong and *will be known to be wrong* (Morton 1993), but will still have the possibility of being *approximately* realistic.

Some of these issues have been explored recently from a philosophical perspective by Cartwright (1999), who suggests that open systems might best be represented in terms of the *capacities* of real entities to respond in a particular way to external influences. She argues that the *representation* of those capacities is only possible within a nomological system with its own defined constraints and limitations. Thus the problem of defining a conceptual model of the system is, within this framework, a matter of defining a nomological system that may not, in itself, be realist in terms of being totally consistent with the perceptual model, but that can be used to produce quantitative predictions within the limits of its own definition. This approach is used very widely in environmental modelling. We should also note here that even having designed a consistent nomological representation of an environmental system, there may be further approximations necessary in implementing that system on a digital computer, such as in the approximate discrete numerical solution of the continuum partial differential equations that are the basis of many environmental models.

In these cases, the need for approximating the perceptual model into a set of mathematical concepts at a certain scale of calculation element is obvious. A good example is the parametrization of sub-grid-scale processes in GCMs that are as much a result of computational constraints as lack of understanding of smaller-scale

Not in
References!
'Davies' in
refs!

processes. In other cases, the process is not so self evident, such as when equilibrium geochemical codes for solutions are applied to mixtures of waters from different sources at field scales (while interacting with a variety of mineral surfaces, organic surfaces and suspended colloids). Even though the principles of the perceptual model may be reasonably well understood, the implementation of those principles in practice may be difficult both for reasons of finding realistic approximations of complex open systems (within the limitations of computational constraints) and also because of lack of knowledge of the local characteristics of the system and its boundary conditions (Beven 2000, 2001b).

This may be, in part, because we still have some surprises to learn about the nature of the processes, but it would seem that even if we had a *perfect* formal description at some practically useful scale, the dominant problem is that the characteristics of the system, which have an important control on how the system works, may be essentially unknowable in detail. Keeping to the sphere of water, there is an analogy here with the study of turbulence. Homogeneous turbulence has received much study and simplified representations of the energy dissipation down to viscosity-dominated scales are used widely in computational fluid dynamics (CFD). It is true that there are still debates about the best closure schemes to use at modelling scales larger than those that are computationally addressable with direct numerical simulation, but for systems that are largely self organizing, CFD models often appear to produce acceptable (publishable) results.

However, there are many fluid-dynamic systems in both laminar and turbulent flow regimes in which the interactions with the boundaries dominate the energy dissipation. This is true for shallow turbulent water flows in streams and rivers (where the aquatic and bankside vegetation may also play an important role in energy dissipation depending on the discharge in the channel); it is true for (mostly) laminar porous media flows in soils and groundwaters. Again, it is not so much that the principles of the interactions are not understood in the perceptual model (though, again, we may still have much to learn), but that it is necessary to characterize the energy dissipation at every point in the flow domain or, more correctly, at the scale of the elements of the approximate representation of that flow domain in the procedural implementation of the conceptual model to produce quantitative predictions.

Scale here is a very significant issue. Its effect is seen in an extreme form in the case of GCMs. Computational constraints mean that global-scale coupled ocean-atmosphere models can still only be run with grid elements that are large relative to the scale of many of the processes being represented, particularly in the sub-grid convection schemes and ‘land-surface parametrizations’ used. Typical spatial grid-scales, for example, in the Hadley Centre CM3 model, runs are 2.5×3.75 degrees for the atmosphere and land surface and 1.25×1.52 degrees for the oceans (see Gordon *et al.* 2000). The same constraints only allow a limited number of layers in the vertical (19 layers for the atmosphere, 20 layers for the oceans). Thus smaller-scale processes (such as smaller-scale turbulence in the atmosphere and oceans) must be represented by sub-grid-scale parametrizations. For the land surface, this has typically meant the implementation of a one-dimensional (in the vertical) model of mass, energy and carbon fluxes in the soil/vegetation/lower atmosphere system (e.g. the SiB2 model of Sellers *et al.* (1996) or the UK Meteorological Office Unified Model MOSES land-surface parametrization (Cox *et al.* 1999)). A single such model is used to represent

Clarify sentence?

each grid square in the GCM (or, more recently, a small number of such models are used to represent different surface covers in each grid square as ‘tiles’).

Since there is no other source of spatial variability within the land-surface parametrization in each grid square, it is clear that the parameter values required in the model must be ‘effective’ values that will ensure that the fluxes from the highly heterogeneous (and time variable) domain that is the real surface will be properly reflected in the model. To reflect the differences between varied types of vegetation (and soil, topography, etc), *different* effective parameter values should, in principle, be specified for each grid square (or tile). A change in grid-scale might also require different effective parameter values because the processes controlling the fluxes between land and atmosphere are nonlinear (which actually implies that different equations might also be required at different scales, as well as different effective parameter values, though this is often ignored). Ways of aggregating local information to suggest grid-scale effective parameter values might also be different in different environments. In some, the surface may be generally wet and the surface cover does not exert the dominant control; in some, the topographic controls on the availability of water may be dominant; in others, the surface may be predominantly dry but there may be irrigated fields that are an important control on the fluxes (an excellent example may be found in Bastiaanssen *et al.* (1994)). Such variability is generally ignored in current GCMs: effective parameters are often taken to be equal to local-scale parameters for the ‘dominant’ cover and soil types. This is clearly one choice of a land-surface parametrization. It is process based in the sense that the processes thought to be important at the local scale are represented in the model; it is very often a nonsense in that the model structures and parameters used take no account of the (nonlinear) sub-grid-scale variability that controls the grid-scale fluxes.

This is, of course, only one of the sub-grid-scale components required by GCMs. Others are required for sub-grid-scale convection, cloud and rain formation, ocean aerosol production, etc. (although, at least with the latest generation of coupled ocean–atmosphere models, there is no longer a need for annual flux corrections (see Gordon *et al.* 2000)). Others still, such as the effect of topography on the availability of water to plants in valley bottoms during dry periods, or the effect of local sub-grid-scale wetting by convective rainstorms on grid-scale latent heat fluxes, are not taken into account at all in the current models. All will require effective parameters, and boundary conditions, at the grid-scale. The more processes that are included then, in general, the more parameter values that will be required before predictions can be made. Similar arguments will apply to all other distributed environmental models that depend on such sub-grid-scale parametrizations. This will be the case even down to quite small grid-scales when current measurement techniques are limited to ‘point’-scale measurements (Beven 2001b).

It could be argued from a pragmatic realistic perspective that the problems outlined in this section are not sufficient, in themselves, to require a new philosophy of environmental modelling, but merely reflect the technological constraints of today (computing limitations, measurement limitations, theoretical limitations, etc.). The expectation is that new technological developments in computing and measurement will reduce the significance of today’s problems. There is always the hope of unforeseen technological innovation, but this argument does not survive a examination

of the effects of trying to apply models to specific locations with their own unique characteristics.

3. Uniqueness of place in environmental modelling

In any application of a model, it is necessary to specify the boundary conditions and auxiliary conditions before we can implement a quantitative simulation and make predictions for a particular location with its own unique characteristics. The problem arises even in the most trivial ways. In hydrology, for example, the conservation of mass is a basic principle that is accepted by all models. No hydrologist or hydrological modeller would seriously question that principle. However, to demonstrate that principle for a particular catchment area of any useful size is currently beyond our measurement capabilities. Similarly, no hydrologist or hydrological modeller would seriously question the conservation of energy or Newton's second law of motion as principles controlling the movement of water within the landscape, but the application of those principles is even more fraught with difficulty, since energy is dissipated in the system in a way that is controlled by the auxiliary and boundary conditions in (nonlinear and non-stationary) ways that are very difficult to either measure or predict in detail. Formulating the balance equations is itself a significant task (e.g. Reggiani *et al.* 1998, 1999, 2000); defining the coefficients that control the fluxes in those equations to reflect the particular characteristics of particular places with any precision is effectively still beyond us (Beven 2000, 2001b).

Two problems follow from the problems of representing the detailed effects of boundary and auxiliary conditions and their reflection in effective parameter values in a simulation model. The first is that there may be no way of uniquely defining model structures or parameter sets for a particular application, even when significant quantities of data are available for model 'calibration'. This is seen empirically as many different models giving simulation results that are consistent with the available data (e.g. Spear *et al.* 1995; Beven *et al.* 2000). Secondly, it must be appreciated that it is the parameter *set* that gives a good simulation.

Many environmental modellers treat the first problem as a problem of identifiability. This is essentially consistent with the same pragmatic realist view that the model is 'realistic' in its equations (as far as is possible), but we may not have adequate data to properly determine the values of the parameters needed. This underlies all statistical inference about parameter values, as embodied for example in the formulation of likelihood functions (though see Kennedy & O'Hagan (2001) for a recent attempt to include simple model inadequacy effects within a Bayesian statistical framework). Statistical estimation, usually by maximizing the likelihood function, depends on the possibility of defining an appropriate model of the simulation errors. In many cases this is very difficult, particularly in systems with natural (random) boundary forcings, uncertain initial conditions, nonlinear responses and strange attractors, and time variable model structural error. Thus it is common to simply maximize some convenient objective function to calibrate parameters and then proceed to use these 'optimized' values in simulation. In doing so, it is sometimes overlooked that the calibrated values might be dependent on both the chosen objective function and the type and period of calibration data, as well as the model structure itself.

There may be many parameter values to calibrate. Indeed, the calibration problem is often underdetermined with respect to the data available and the type of

Not in
References!
Not in
References!
Not in
References!

'1994' in refs!

'O'Hagen' in
refs!

forcings in the calibration data, leading to problems of non-identifiability (a poorly defined optimum) and/or non-uniqueness (multiple local optima). These problems are endemic to environmental models. Experience suggests that they do not go away with increased physical realism of the model structure, nor with longer calibration datasets or multiple objective functions using different types of data. Longer calibration periods and multiple objectives can help, but increased physical realism can actually make the problem worse if more complex process representations are added with more parameters that cannot be estimated independently. The perfect model, if it could be defined, would be likely to have a very large (semi-infinite) number of parameters, and so would not be immune to such problems in applications to unique places with limited observations.

One interesting issue here is revealed by studies involving replicate measurements at different, similar, sites. This, in many areas of environmental science, is deemed to be good practice (see, for example, the exhortations in Peters (1991)), although it is not always economically or technically feasible to do so. However, for these open heterogeneous systems, even in the best-designed replicate experiments, the measured responses must inevitably reflect uniqueness of inputs, characteristics and measurement artefacts. In some cases, it has been shown that the differences between replicates may be large and inconsistent over time periods (such as shown by the measurements of runoff and erosion from large numbers of replicate plots in Hjelmfelt & Burwell (1984) and Nearing *et al.* (1999)).

This implies that each replicate might require its own set(s) of local model parameters (see, for example, Lamb *et al.* 1998). The aim of the model, however, is not usually to predict the replicate responses except for evaluation purposes. The aim of the model is rather to predict the response of all the unmeasured locations in the landscape, with their own unique characteristics. In this context, we should not, therefore, expect a model to be more precise in its predictions than the expected or observed variability between replicates at the same scale, while the deterministic prediction will, with almost perfect certainty, be wrong in any particular location.

This then has an important bearing on the second problem of finding a parameter *set* to represent a location, since it is quite common in applications of environmental models to take parameter values that have been determined elsewhere under very specific conditions as the basis for estimating the values at a new site or time period. Such values are often reported in the literature without a clear indication of how they were calibrated or back-calculated from data, what other parameters were also involved in the calibration, and what range of conditions were covered by the calibration data. If identifiability is a problem, then there is no guarantee that bringing together parameter *values* from different sources to create a set of parameters for a simulation will lead to acceptable simulations (where acceptability might be defined with respect to potential variability amongst replicates), even if those values of parameters have been shown to produce good results elsewhere. The impact of the unique characteristics of that location on the effective parameters required by a model may be such as to require different sets of values.

Indeed, Beven (1993, 2000) has suggested that a better approach to the identifiability problem is to recognize explicitly that there may be multiple model representations that provide acceptable simulations for any environmental systems. This is the concept of *equifinality* (to distinguish it from non-identifiability or non-uniqueness). The identification problem is then a matter of differentiating between *behavioural*

models expected to give acceptable simulations and *non-behavioural* models that can be rejected. It is a combination of model structure, the boundary conditions and the set of parameter values that gives rise to a behavioural simulation, although the decision about whether a model is acceptable must be dependent on the choice of a criterion or criteria of acceptability in the same way that optimized parameter values depend on the choice of objective function and calibration data. Some data about a site will therefore be necessary to determine whether a model is behavioural or not. This approach has been formalized in the generalized likelihood uncertainty estimation (GLUE) methodology of Beven & Binley (1992; see also Beven *et al.* 2000; Beven & Freer 2001). An approach based on the equifinality concept is intrinsically linked with predictive uncertainty, since there is now no longer just one prediction, but rather there are multiple predictions from all the behavioural models.

GLUE may be considered directly in the line of a Bayesian approach to science (Howson & Urbach 1993). Indeed, in its simplest form, it represents a form of Bayesian averaging over all behavioural models and includes formal Bayesian approaches to mechanistic modelling as special cases (e.g. Krzysztofovitz 1999). In prediction, all the behavioural simulations may be used to make predictions and the predictions of each model can be weighted in accordance with how well that model has done in predicting past data. These likelihood weights can be updated as new evaluation data become available to form a new posterior distribution of behavioural parameter sets. Thus, in this approach, behavioural simulations are only retained as long as they are consistent with the data available for evaluation. Any local or global interactions between parameter values in producing an acceptable simulation should be implicitly reflected in the set of behavioural models retained in this way. The focus in the approach is firmly on the data that will allow the conditioning process by which all feasible models are reduced to the set of behavioural models. The GLUE methodology has now been applied to a wide variety of environmental models (see Beven *et al.* (2000) for a recent summary).

Bayesian approaches have previously been interpreted in terms of anti-realist and instrumentalist philosophies. The consequences of model equifinality clearly do create problems for a realist stance. It is also clear, however, that the recognition of equifinality will not result in environmental modellers giving up their pragmatic realism. The fact that we cannot decide which is *the* model of the system with all the idiosyncracies that are reality can be considered as only a reflection of decidability. New data sources should help make parameter values more identifiable (or constrain the set of behavioural models). One source of new data that has the potential to be particularly important in dealing with the problem of uniqueness of place is remote-sensing data. Remote sensing does, after all, yield time sequences of spatial images of unique places (albeit with either limited spatial resolution or limited time resolution). Space, time and spectral resolutions will improve with the new generation of EOS satellites being launched. Yet it does not appear that remote sensing has yet had much impact on the uniqueness problem. It is interesting to consider why.

Spell out?

One reason is that remote sensing rarely gives results of interest to the environmental modeller directly. The digital numbers of the images require interpretation. In general, this will require a separate model with its own parameters (and, indeed, before the modeller even sees the digital numbers they will generally have been subjected to a corrective model with its own parameter values). These models are generally used as if the parameters were known *a priori*, but such models will also

not be exempt from time and space variability of parameter values and, indeed, the problem of equifinality. An example is given by the identification of different bidirectional reflectance function (BDRF) coefficients by Lewis *et al.* (1999) on the basis of high-resolution multilook multispectral sensor images. The results showed that the optimal model structure, as well as the optimal model parameter values, could change from pixel to pixel in the images. This gives a simple indication of the potential for equifinality in this type of interpretative process with consequent uncertainty. Such uncertainties are frequently ignored on the basis that the values of parameters that have given acceptable results elsewhere must have some physical significance. This takes no account of either the uniqueness of place or the importance of parameter *sets* in inducing behavioural simulations.

It is true, however, that significant information on spatial patterns can be seen in remote-sensing images and direct measurement of spatial patterns of (point-scale) observables (e.g. Grayson & Blöschl 2000). Such spatial information should prove useful in many types of applications but, again, may then require local parameter values within a model structure to represent those local observations, creating further problems for model identifiability. It is in a more explicit consideration of the inevitable uncertainties that arise from such considerations of uniqueness of place that the foundations of a new philosophy might lie.

4. Environmental modelling as a landscape space to model space mapping

The nature of environmental systems is such that they are not easily closed for direct scientific experimentation. Where direct experimentation is possible, it is normally only at small scales and it has often proven difficult to take relationships and theories developed at these scales and apply them in the field where so much of the characteristic nature of the real system remains unknowable. Thus resort is normally made to experimentation in the form of monitoring (perhaps in response to experimental forcings) of the system in all its complexity and as open to external influences. However, to develop and apply environmental models, some form of closure must be enforced. Thus, on the one hand, we have the real ‘landscape space’, and on the other a closed ‘model space’ that is constructed so as to have (hopefully) an adequate range of functional responses. The landscape space cannot be known perfectly, since even our perceptual model may be, as yet, inadequate and subject to change over time. The model space, however, can be known perfectly for deterministic models with specified boundary conditions and its response to different types of forcing can be explored perfectly (within the limitations of computer resources). This can be extended to models with stochastic inputs and outputs, at least in principle, but at the expense of more computer resources. It is, again, analogous to the nomological system of Cartwright (1999) in that it is a formal definition of the capacities of the processes described in the model and within which the model constructs have validity (here, by definition). It must be recognized that the model space may not cover all eventualities that might arise in reality, but that if information becomes available about such eventualities, then it might lead to a revision of the feasible model space.

Thus the modelling problem can be formulated in terms of a mapping of the ‘landscape’, the prototype system or part of nature of interest, into a high-dimensional

model space, where the dimensions might include various parameters, various boundary conditions and, in the general case, various competing model structures. This mapping will be necessarily uncertain (or fuzzy). Initially, it can be done on the basis of prior estimates of the appropriate model structure(s), boundary conditions and parameter values. The choice, for example, of a single model structure, single set of boundary conditions and single set of parameter values for a ‘landscape unit’ (such as a GCM grid-scale land-surface parametrization), is then equivalent to mapping that landscape unit to a single point in the model space. If the model is run deterministically, then there is only one possible set of outcomes. This is a very common modelling practice, but the possibilities for uncertainty in such a restricted mapping have been revealed, for example, in the project for the intercomparison of land surface parametrizations (PILPS) for GCMs, which reveals considerable variation between groups of modellers and models in trying to represent the same landscape units (see, for example, Lohmann *et al.* 1998). The implication is that the particular choices made, for example, in running GCM simulations might be, to some extent, unrealistic.

It really does seem that this mapping is generally being carried out in an overly deterministic way in many such modelling studies. This is understandable with the current generation of GCMs because computing constraints mean that there is no possibility of allowing for uncertainty in the mapping for every process and every location (even though this carries the risk of the simulations being in error). However, for many less computationally demanding models, it would be possible to allow for an uncertain mapping. Defining the likelihood function in a model parameter space in statistical modelling is one such mapping, within which the covariation of the parameters in fitting the data can be estimated. The GLUE methodology can also be interpreted in this way. It is normally applied using Monte Carlo simulation to sample the model space and some likelihood measure is calculated for each modelled sample by a comparison of observed and predicted responses. The values of the likelihood measure for each behavioural model representation of a landscape unit can then be used as weights in the mapping process (Beven 2000, 2001*b*). Those models that have done well in simulating that landscape unit (in so far as it is possible to test performance given the available observations) will have the greatest weight. The set of weights represents the mapping of the landscape unit into the model space. Other, density-dependent, sampling methods, such as Latin hypercube methods or Monte Carlo Markov chain methods, can be used when there is strong prior evidence for the form of the mapping, though such strong prior assumptions are often difficult to justify in applications to unique places.

An uncertain mapping will mean that a given (real) landscape unit might be represented by a cloud of sampled points in the model space rather than a single point. It might even be represented by several non-contiguous clouds of points in the model space where local interactions between parameter values (and possibly model structures) give rise to locally behavioural models (e.g. Spear *et al.* 1994). Some features of the modelling process, such as the fact that quite different model structures might be equally consistent with the observations and that parameters with the same name might require different values in different model structures, are easily handled within this framework.

This mapping also has the features of a Bayesian methodology (Howson & Urbach 1993). It can be formulated purely on the basis of prior information. If some obser-

variations on the prototype become available, they can be used to update the mapping and form a posterior distribution of the behavioural parameter sets (and possibly model structures). Depending on the performance of the models in reproducing the functionality of the prototype, the posterior mapping may have less uncertainty than the prior or sometimes more. In some circumstances where the data show that the model does not have the correct functionality, then the posterior set may be null. However, in such cases, the modeller should have learned from such a process of rejection, as in any hypothesis testing methodology.

5. Hypothesis testing and model validation in the face of equifinality

The idea of a landscape space to model space mapping outlined in the previous section requires an ability to define those areas of the model space where behavioural models occur. It also seems to be based on a relativist philosophy, since there is an implication that any ranking of the models in terms of some likelihood measure (or true likelihood where this can be properly formulated) means that some models are rejected and, of the remainder, some are more likely than others and can therefore be given more weight in prediction. The behavioural models could also be viewed as multiple working hypotheses, some of which we have (slightly) more confidence in than others on the basis of the empirical evidence.

However, within this framework, there is scope for refining the mapping with a view to imposing additional constraints on the predictive uncertainty that is the consequence of an uncertain mapping. There are two ways in which this might be done. The first is in a purely Bayesian updating sense, waiting for additional observations or different types of observation to become available and then evaluating all the behavioural models in the model space to see if they remain consistent with the new information (see, for example, the effect of adding a new type of information in Franks *et al.* (1998), in constraining the feasible parameter sets in a hydrological modelling example).

The second might be more active. Since the model space can actually be explored in terms of the responses of all the behavioural models, these can be evaluated *deductively* in terms of function (e.g. the predicted values of a particular output variable). It might then be possible to formulate hypotheses that differentiate behavioural models in terms of how they function, quite independently of any experimentation on the prototype. If testable hypotheses can be formulated, and the required observations are made, then it will be possible to reject or falsify a subset of the behavioural models and thereby refine the predictive uncertainty. Of course, not all the potentially differentiating hypotheses will be testable with current measurement technologies.

Models that survive such a test (i.e. continue to be considered behavioural) can be considered to have an increased level of confirmation as a result of this demonstration of empirical adequacy (Van Fraassen 1980). This is not the same as verification, as is evident from the fact that in most applications of environmental models there will continue to be a set of behavioural models (or none if we look too closely). The problem of verification in respect of environmental models has been discussed in detail by Oreskes *et al.* (1994) in a contribution from professional philosophers as a response to the evaluation of the predictions of groundwater models by practicing environmental scientists (Konikow & Bredehoeft 1992; Anderson & Woessner 1992). Oreskes *et al.* suggest that verification and validation of models of open natural systems is

impossible, despite the widespread use of those words in the modelling literature. Models of such systems may be non-unique and only a conditional confirmation is possible. It is conditional because it may depend on errors in the model structure(s), the calibration of parameters or other auxiliary conditions and may also depend on the period of data used in evaluation. The approach proposed here is fully consistent with this conditionality while retaining the possibility of a realist perspective.

6. Explanatory depth and expecting the unexpected

The unknowability inherent in uniqueness of place is a constraint on the explanatory depth of models as used in prediction about real places. These models may aim to simulate the correct capacities of the underlying, more fundamental, processes, but there will be a limit on how far those capacities can be reproduced when they must be represented in terms of effective parameter values at the scale of the model, even if the fundamental principles embodied in the model equations were correct. Again, the conceptual or formal model will be wrong and will be known to be wrong.

This then poses two interesting questions. The first is how approximate can a model be, and still retain an element of realism in explaining and predicting the quantities and fluxes of interest? The user of the model predictions (and the modeller as instrumentalist) is not necessarily so interested in the explanatory *depth* of the model as in its explanatory *power* for the phenomena being predicted; that is, the power to predict states and fluxes of real quantities, especially under different scenarios for the future. The view is often expressed that a realist philosophy implies a reductionist progression to ever more fundamental levels of epistemological and ontological levels of explanation. This implies that realism then *requires* explanatory depth, even if the stopping rules may be problem dependent. It is not clear that this is actually a simple progression in the case of environmental prediction, since it is necessarily linked to the problems of formal model complexity and knowability of characteristics at different scales. Explanatory depth, however complex, does not necessarily imply explanatory power if there is a conflict between the formal model and the processes or characteristics of a particular application. In such a case, approximate models for which the parameters may be more easily identified may be advantageous.

The second question is how far can models that are necessarily approximate be expected to predict the unexpected, i.e. modes of response of the system that have not yet been observed? A fully realistic model should do this correctly. An argument is therefore often used in model development that additional processes and complexity perceived as having an effect in the real system should be represented in the model, even if only as phenomenological relationships of no great explanatory depth, because those processes might be important in predicting future responses. Such an argument has led to the great vertical complexity of land-surface parametrizations used in GCMs, noted in § 2 above, at the expense of neglecting sub-grid-scale spatial heterogeneity. The explanatory depth of the vertically complex description may be greater for predictions at a point, but the explanatory power at the scale of the GCM grid (the scale at which predictions are required) has been significantly compromised.

Similar issues arise in a variety of other fields. There is a description of mixing in river channels based on turbulent shear in the vertical velocity profile that leads to a one-dimensional advection–dispersion equation (ADE). The approach has been

shown to give accurate predictions in laboratory flume experiments and has been applied very widely in natural channels (Rutherford 1994). In many natural channels, however, the ADE will give incorrect predictions. While it has depth of interpretation in terms of fundamental mixing processes, it completely neglects the transverse mixing and internal shear zones that are important in controlling mixing in natural channels of irregular three-dimensional geometries and boundary conditions (Young & Wallis 1994).

‘1993’ in refs!

The ADE has also been applied to the problem of predicting transport of solutes in porous media. Again, it has been shown to give good predictions of solute breakthrough curves in laboratory columns of homogeneous media. Adding terms to the ADE allows processes such as adsorption/desorption to be included, so that a wider range of solute behaviours can be predicted. It has been used very widely to predict transport of contaminants in real soil and groundwater systems, after some calibration of the parameters involved. A interesting example of its application has been in the prediction of the movement of highly sorptive contaminants (pesticides, herbicides, phosphorus) applied at the surface of the soil. The predictions are that such contaminants will be retained in the near-surface soil. Sampled vertical concentration profiles at study sites confirm this impression, with a rapid decline in measured concentrations down from the soil surface. However, gradually, other types of measurements revealed that this was not the whole story. Unexpectedly, pesticides, herbicides, phosphorus and other contaminants were being found widely in field-drainage waters, stream waters and at depth in groundwaters. The ADE, for which the advective term is based on a mean velocity, neglects the possibility of flow in preferential flow pathways in the natural soil structure that allow at least small amounts of contaminant, possibly sorbed to colloidal material, to move rapidly to depth during rainfalls. Without expensive high-frequency sampling in the field during storm periods, this process was not evident from the measurements.

All these examples show that when taking formal models into the environment and applying them to unique places, depth of interpretation should not be confused with explanatory depth of the real phenomena, nor with explanatory power in prediction. All that can be demanded is consistency of explanation at the scale of application, including consistency with perceptual understanding and observations at larger and smaller scales. The importance of preferential flow in contaminant transport should not have been a surprise, for example. It had already been recognized in experimental work by Lawes *et al.* (1882) at Rothamsted and the translocation of clay particles in large pores had been recorded from soil-thin section work for over 100 years (see the review of preferential flow in Beven & Germann (1982)). The unexpected may not always be so unexpected when we are forced to review a model because of a failure in prediction.

In the context of the landscape to model space mapping approach proposed here, ‘unexpected’ predictions in the model space can be evaluated, in so far as they are inherent in the model structure(s) considered. This suggests that it might be dangerous to exclude models that are consistent with observational data but lead to ‘unexpected’ conclusions in prediction. (Is it still possible, for example, that increasing greenhouse gas concentrations might accelerate the onset of a new ice age? Certain model parametrizations in the past predicted this. They do not now form part of the consensus view of climate change (IPCC 2001), but is it that they have they now been invalidated, rejected as non-behavioural for good reasons, or merely qui-

etly forgotten?) Indeed, such unexpected predictions might prove to be a useful basis for the type of hypothesis testing described above, since they can suggest ideas for experimentation where that is possible. Where it is not possible, then it may be just a question of monitoring and waiting for the unexpected. All the time that the range of behavioural models remains behavioural, then the predictions will appear to be acceptable. If some models can be rejected on the basis of new, perhaps unexpected, data, then perhaps the set of behavioural models can be refined. When the truly unexpected occurs and all the behavioural models must be rejected, then the science will truly progress (even if only by considering more realistic auxiliary conditions, as noted by Morton (1993)).

The unexpected may be the effects of human interference, either inadvertent or as a result of policy and management. Anthropogenic effects may change either the characteristics of the system or its boundary conditions at the scale of closure. Often, these impacts are not well quantified, even in hindcasting and history matching (the construction of records of inputs of greenhouse gases into the atmosphere since pre-industrial times is a good example of such uncertainties). Thus it may only be possible to treat such impacts as potential scenarios, giving potential outcomes but without any objective estimation of the associated uncertainty. Most predictions of the impacts of climate change are of this type and can be handled within the framework suggested here, which can allow explicit subjective evaluations of possible outcomes if necessary.

However, it has often been argued in predictive studies of the impacts of change that even if our models are not strictly realistic in modelling current conditions, they can still be used to predict the relative magnitude of change for different scenarios (for an example in the context of global environmental change, see IPCC (2001)). This would appear to be an extremely tenuous argument in the context of nonlinear systems. It assumes, rather blithely, that although model structural error may mean that not all the nonlinear interactions in the modelled system are represented accurately under current conditions, the potential for nonlinear change in those interactions under changed conditions will be represented accurately. In the words of the IPCC (2001, box 3, p. 49):

The differencing technique removes most of the effects of any artificial adjustments in the model as well as systematic errors that are common to both runs.

They do, however, go on to recognize the problem of model structural error immediately:

However, a comparison of different model results makes it apparent that the nature of some errors still influences the outcome.

Thus, in the context of the philosophy espoused here, there is uncertainty with respect to model structure that should lead to uncertainty in the predicted change under different scenarios. It is, in fact, quite possible that the distribution functions for predicted variables under different scenarios may be overlapping. This was shown by Cameron *et al.* (2000) for the case of changing flood frequencies under different future climate scenarios. They could, however, still interpret their results in terms of a change in the *risk* of a given flood magnitude being exceeded.

In the Cameron *et al.* study, only the inputs to the model were being changed, not the model parameter sets. Consideration of change, for example, in land-management and land-surface responses to change in climate, might also require the estimation of changes in parameter values. This is now much more problematic, since the point has already been made that the value of a parameter is strictly only significant in the context of the model structure and parameter *set* within which it was found to provide behavioural simulations. Changing m different parameter values of a model within a set that is behavioural under past conditions to reflect a new set of conditions might not be so simple. What is needed is an m -dimensional transposition of that behavioural model in the model space to a new position (or multiple potential positions) that will properly reflect the new set of conditions, *including* the potential for different types of parameter interaction in the model structure.

7. Realism, relativism and the limits of predictability

The arguments presented above may be summarized in terms of the following principles.

- (1) That a formal environmental model can only ever be an approximation to the perceptual model of the complex processes governing the response to some forcing.
- (2) That places are unique in their characteristics and boundary conditions and their uniqueness is inevitably to some extent unknowable.
- (3) That there will then always be the possibility of equifinality in model representations of a system, in that many different structures or parameter sets may give simulations that are acceptable representations of the observations available.
- (4) That the range of behavioural models is best represented in terms of an uncertain 'landscape space' to model space mapping.
- (5) That the range of predictive uncertainty associated with the set of behavioural models should therefore be explored.
- (6) That observations will be crucial in constraining the set of behavioural models and gradually refining understanding of the response of particular places.

It has been suggested that these principles provide the basis for a coherent philosophy for environmental modelling that allows for both the difficulties of decidability between different competing model descriptions *and* the possibility of a scientific hypothesis-testing approach to refining the set of feasible models. They imply that a realist approach to environmental modelling is problematic, although they do not preclude it.

They also imply some important practical problems for the application of environmental models. They imply an evaluation of all feasible models, making use of all valuable information available about the prototype, including prior experience. They imply predictions carried out using multiple behavioural models. They imply explorations of the model space with a view to finding potentially testable hypotheses.

They imply lots of computer resources (although taking large numbers of samples from the model space is feasible now for at least some models (see the applications in Beven *et al.* (2000))). They imply that it may not always be possible to differentiate between competing models or, alternatively, that if we look in sufficient detail, it might be possible to reject all the necessarily approximate models in the model space (see the example in Zak & Beven (1999) in the context of modelling critical loads for acid deposition). It has already been noted that even if we knew *a priori* the structure of the perfect model, it would not be immune to these problems because of difficulties in specifying parameter values and boundary conditions in any application.

What then are the implications for modelling practice? Clearly, many pragmatic realists and instrumentalists continue to use a ‘fit-for-purpose’ argument and retain only one ‘optimal’ model as being the ‘most realistic’ available for the conditions under study. It is relatively rare at present that applications take any account of modelling uncertainties. They tend to be based rather on obtaining a ‘best estimate’, perhaps with a certain amount of sensitivity analysis. Where uncertainties are taken into account (such as in the estimation of risk in the regulation of the nuclear industry), then the number of degrees of freedom involved means that usually only *a priori* estimates of uncertainty are included and that the possibility of covariation and interaction among parameters and other factors is necessarily treated only very approximately, if at all. Where it is possible to make statistical inferences about model parameter values, applications often do not take account of model structural error or consider the possibility of quite different predictions arising from different behavioural model structures (but, see, for example, Draper (1995) for some interesting examples of doing so).

Such practices seem to have important limitations. However, taking more realistic account of the inherent uncertainties in modelling the environment is not only a matter of devoting greater effort and computer resources but involves a variety of other issues in the sociology of science, policy and decision making. These elements are often very closely linked (for discussions, see Gare 1995; Ravetz 1996). A very brief summary of the issues in this context could perhaps be presented in the form that there will be no requirement to take a more realistic account of uncertainty until the users (policy formulators and decision makers) demand it, and that such users are concerned that ‘too much uncertainty’ in the science will not be helpful in either policy formulation or decision making (especially if the scientific models are considered to be only socially conditioned constructs anyway (see, for example, Ravetz 1996; Schrader-Frechette 1985)). The science must evolve in this context. The evolving uncertainty estimates in measures of climate change as summarized by the IPCC are an interesting case in point here (see Shackley *et al.* 1997). Such predictions are necessarily the result of only a few GCM scenarios, but in other application areas it is clear that the possibility of doing many different model runs is now technically feasible. They are just not usually demanded.

The philosophy for environmental modelling that is being suggested here aims to extend the pragmatic realism that underlies much of the current work in both research and practical prediction with a more realistic account of uncertainty, formulated here in terms of a concept of landscape to model space mapping. As such, it is a reflection not only of practical modelling experience, but also of some of the current discussions of uncertainty that have arisen out of concepts of nonlinear dynamics

‘Shrader-Frechette’ in refs!

and complexity in open systems (e.g. Ruelle 1991; Prigogine 1997; Rescher 1998), in which chance effects and the self organization of (simple) nonlinear dynamic systems interact to produce the observed complexity of the real world emergent at the scales of interest. Uncertainty then arises from the sensitivity of such systems to initial conditions (especially where there is a tendency to chaotic behaviour) and the problem of trying to mimic the real attractor of the system (Smith 2001). The types of parameter-estimation problem described above, that are endemic to models of open environmental systems in particular places, will then only add to the uncertainties. Such concepts, together with stochastic modelling approaches and ‘dominant mode’ data-based modelling (e.g. Young 2000, 2001), are easily encompassed within the philosophy outlined, albeit that there may be practical limitations in realizing all the possibilities involved.

Thus a formulation in terms of a landscape to model space mapping is not inconsistent with a realist stance (in the sense, for example, of Bhaskar (1980, 1989) and Chalmers (1989)), or of a relaxed form of Popperian falsificationism (Popper 1959, 1979), while recognizing that the approach towards a *single* realist description of any environmental system will be fraught with difficulty, since, in most practical applications, it is all too easy to falsify *all* the available models if they are evaluated in detailed. Less severe criteria for falsification then results in an equifinality of model representations, which may be viewed as a variation on the underdetermination thesis (Van Fraassen 1980; Newton-Smith 1981, 2000), applied to a restricted range of potentially realistic models rather than theories. It would seem that the attention given to particular places in the landscape to model space mapping makes such a realist interpretation more easily tenable, since the aim of the modeller is then to achieve a progressively realistic representation of those places within the limitations of currently available modelling and observational techniques. However, the approach is also flexible enough to allow relativist or instrumentalist interpretations if it is believed that a realist description will prove ultimately impossible. Consequently, it reflects the immaturity of the environmental sciences that should, we hope, evolve with the development of new techniques for observing and characterizing the systems and places of interest. Yet it does not seem that this would dramatically change the underlying philosophical approach suggested here, which does, at least, give some guidance about the need to be specific and explicit in the form of mapping used. In this respect, it provides a scientific basis for the discussion and comparison of environmental models in applications to unique places, without too much need for self delusion in presenting the value of model representations of complex environmental systems.

The ideas in this paper have benefited from discussions with a wide variety of colleagues and students over a long period of time. Earlier versions of the manuscript have been improved as a result of many comments and suggestions, particularly from Rodger Grayson, Rob Lamb, Adam Morton and Thorsten Wagener, but any inconsistencies remain, of course, mine.

References

- Abbott, M. B. 1992 The theory of the hydrological model, or: the struggle for the soul of hydrology. In *Advances in theoretical hydrology* (ed. J. P. O’Kane), pp. 237–254. Elsevier.
- Anderson, M. G. & Bates, P. D. (eds) 2001 *Model validation: perspectives in hydrological science*. Wiley.

- Anderson, M. P. & Woessner, W. W. 1992 The role of the post-audit in model validation. *Adv. Water Resources* **15**, 167–174.
- Bastiaanssen, W. G. M., Hoekman, D. H. & Roebing, R. A. 1994 *A methodology for the assessment of surface resistance and soil water storage variability at mesoscale based on remote sensing measurements. A case study with the HAPEX-EFEDA data*. IAHS Special Publication No. 2. Wallingford: IAHS Press.
- Beven, K. J. 1993 Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Resources* **16**, 41–51.
- Beven, K. J. 1995 Linking parameters across scales: sub-grid parameterisations and scale dependent hydrological models. *Hydrol. Process.*, **9** 507–526. Not cited in text!
- Beven, K. J. 1996a Equifinality and uncertainty in geomorphological modelling. In *The scientific nature of geomorphology* (ed. B. L. Rhoads & C. E. Thorn), pp. 289–313. Wiley.
- Beven, K. J. 1996b A discussion of distributed modelling. In *Distributed hydrological modelling* (ed. J.-C. Refsgaard & M. B. Abbott), ch. 13A, pp. 255–278. Denter: Kluwer.
- Beven, K. J. 2000 Uniqueness of place and the representation of hydrological processes. *Hydrol. Earth Syst. Sci.* **4**, 203–213.
- Beven, K. J. 2001a *Rainfall-runoff modelling—the primer*. Wiley.
- Beven, K. J. 2001b How far can we go in distributed hydrological modelling? *Hydrol. Earth Syst. Sci.* **5**, 1–12.
- Beven, K. J. & Binley, A. M. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* **6**, 279–298.
- Beven, K. J. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *J. Hydrol.* **249**, 11–29.
- Beven, K. J. & Germann, P. F. 1982 Macropores and water flow in soils. *Water Resources Res.* **18**, 1311–1325.
- Beven, K. J., Freer, J., Hankin, B. & Schulz, K. 2000 The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems. In *Nonlinear and nonstationary signal processing* (ed. W. J. Fitzgerald, R. L. Smith, A. T. Walden & P. C. Young), pp. 115–151. Cambridge University Press.
- Bhaskar, R. 1980 *A realist theory of science*. Brighton: Harvester.
- Bhaskar, R. 1989 *Reclaiming reality*. London: Verso.
- Cameron, D., Beven, K. J. & Naden, P. 2000 Flood frequency estimation by continuous simulation under climate change (with uncertainty). *Hydrol. Earth Syst. Sci.* **4**, 393–405.
- Cartwright, N. 1999 *The dappled world: a study of the boundaries of science*. Cambridge University Press.
- Chalmers, A. 1989 Is Bhaskar’s realism realistic? *Radical Phil.* **49**, 18–23.
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R. & Smith, J. 1999 The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynam.* **15**, 183–203.
- Davies, P. A., Olague, N. E. & Goodrich, M. T. 1992 Application of a validation strategy to Darcy’s experiment. *Adv. Water Resources* **15**, 175–180. ‘Davis’ in text!
- Draper, D. 1995 Assessment and propagation of model uncertainty. *J. R. Statist. Soc. B* **37**, 45–98.
- Franks, S. W., Gineste, Ph., Beven, K. J. & Merot, Ph. 1998 On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Res.* **34**, 787–797.
- Gare, A. E. 1995 *Postmodernism and the environmental crisis*. London: Routledge.
- Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B. & Wood, R. A. 2000 The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynam.* **16**, 147–168.

- Grayson, R. B. & Blöschl, G. (eds) 2000 *Spatial patterns in catchment hydrology: observations and modelling*. Cambridge University Press.
- Hjelmfelt, A. T. & Burwell, R. E. 1984 Spatial variability of runoff. *J. Irrig. Drain. Engng* **110**, 46–54.
- Hofmann, J. R. & Hofmann, P. A. 1992 *Darcy's law and structural explanation in hydrology*, vol. 1, pp. 23–35. Philosophy of Science Association.
- Howson, C. & Urbach, P. 1993 *Scientific reasoning: the Bayesian approach*, 2nd edn. Chicago, IL: Open Court.
- IPCC 2001 *Climate change 2001: the scientific basis*. Cambridge University Press.
- Kennedy, M. C. & O'Hagan, A. 2001 Bayesian calibration of mathematical models. *J. R. Statist. Soc. D* **63**, 425–450. 'O'Hagan' in text!
- Knikow, L. F. & Bredehoeft, J. S. D. 1992 Groundwater models cannot be validated. *Adv. Water Resources* **15**, 75–83.
- Krzysztofowicz, R. 1999 Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Res.* **35**, 2739–2750.
- Lakatos, I. 1978 *Philosophical papers, vol. 1. The methodology of scientific research programmes* (ed. J. Worrell & G. Curry). Cambridge University Press.
- Lamb, R., Beven, K. J. & Myrabø, S. 1998 Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Adv. Water Resources* **22**, 305–317.
- Lawes, J. B., Gilbert, J. H. & Warrington, R. 1882 On the amount and composition of the rain and drainage water collected at Rothamsted. London: Williams, Clowes and Sons Ltd.
- Lewis, P., Disney, M. I., Barnsley, M. J. & Muller, J. P. 1999 Deriving albedo maps for HAPEX-Sahel from ASAS data using kernel driven BRDF models. *Hydrol. Earth Syst. Sci.* **3**, 1–13.
- Lohmann, D. (and 28 others) 1998 The project for intercomparison of land-surface parameterization schemes (PILPS), phase 2c, Red-Arkansas River basin experiment. Part 3. Spatial and temporal analysis of water fluxes. *Global Planet. Change* **19**, 161–179.
- Morton, A. 1993 Mathematical models: questions of trustworthiness. *Br. J. Phil. Sci.* **44**, 659–674.
- Nearing, M. A., Govers, G. & Norton, L. D. 1999 Variability in soil erosion data from replicated plots. *Soil Sci. Soc. Am. J.* **63**, 1829–1832.
- Newton-Smith, W. H. 1981 *The rationality of science*. London: Routledge and Keegan Paul.
- Newton-Smith, W. H. 2000 Underdetermination of theory by data. In *A companion to the philosophy of science* (ed. W. H. Newton-Smith), pp. 532–536. Oxford: Blackwell.
- Oreskes, N., Shrader-Frechette, K. & Belitz, K. 1994 Verification, validation and confirmation of numerical models in the Earth science. *Science* **263**, 641–646.
- Peters, R. H. 1991 *A critique for ecology*. Cambridge University Press.
- Popper, K. R. 1959 *The logic of scientific discovery*. London: Hutchinson.
- Popper, K. R. 1979 *Objective knowledge*, 2nd edn. Oxford University Press.
- Prigogine, I. 1997 *The end of certainty*. New York: The Free Press.
- Ravetz, J. R. 1996 *Scientific knowledge and its social problems*. London: Brunswick.
- Rescher, N. 1998 *Complexity: a philosophical overview*. London: Brunswick.
- Richards, K. S., Brooks, S. M., Clifford, N. J., Harris, T. R. J. & Lane, S. N. 1995 Theory, measurement and testing in 'real' geomorphology and physical geography. In *Process and form in geomorphology* (ed. D. R. Stoddart). Oxford: Blackwell.
- Ruelle, D. 1991 *Chance and chaos*. Princeton, NJ: Princeton University Press.
- Rutherford, J. C. 1994 *River mixing*. Wiley.
- Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D. & Bounoua, L. 1996 A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I. Model formulation. *J. Climate* **9**, 676–705.

- Shackley, S., Young, P. C., Parkinson, S. & Wynne, B. 1997 Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools? *Climatic Change* **38**, 159–235.
- Schrader-Frechette, K. 1985 *Risk analysis and scientific method: methodological and ethical problems with evaluating societal hazards*. Dordrecht: Reidel. ‘Schrader-Frechette’ in text!
- Schrader-Frechette, K. 1989 Idealized laws, antirealism and applied science: a case in hydrogeology. *Synthese* **81**, 329–352.
- Smith, L. A. 2001 Of climate model products whereof must we remain silent. Paper presented at the XXVI General Assembly, European Geophysical Society, Nice.
- Spear, R. C., Gried, T. M. & Shang, N. 1994 Parameter uncertainty and interaction in complex environmental models. *Water Resources Res.* **30**, 3159–3170.
- Van Dam, J. C. (ed.) 1999 *Impacts of climate change and climate variability on hydrological regimes*. Cambridge University Press.
- Van Fraassen, B. C. 1980 *The scientific image*. Oxford: Clarendon.
- Young, P. C. 2000 Stochastic dynamic modelling and signal processing: time variable and state dependent parameter estimation. In *Nonlinear and nonstationary signal processing* (ed. W. J. Fitzgerald, R. L. Smith, A. T. Walden & P. C. Young), pp. 74–114. Cambridge University Press.
- Young, C. C. 2001 Data-based mechanistic modelling and validation of rainfall-flow processes. In *Model validation: perspectives in hydrological science* (ed. M. G. Anderson & P. D. Bates), pp. 117–162. Wiley.
- Young, P. C. & Wallis, S. G. 1993 Solute transport and dispersion in stream channels. In *Channel network hydrology* (ed. K. J. Beven & M. J. Kirkby), pp. 128–173. Wiley. ‘1994’ in text!
- Zak, S. K. & Beven, K. J. 1999 Equifinality, sensitivity and uncertainty in the estimation of critical loads. *Sci. Total Environ.* **236**, 191–214.