Hydrology and Earth System Sciences, 4(2), 203–213 (2000) © EGS

Hydrology & Earth System Sciences

Uniqueness of place and process representations in hydrological modelling

Keith J Beven

Environment Lancaster, Institute of Environmental and Natural Sciences, Lancaster University, Lancaster LA1 4YQ, UK
e-mail for corresponding author: k.beven@lancaster.ac.uk

Abstract

This paper addresses the problem of uniqueness of catchment areas in relation to model representations of flow processes. The uniqueness of field measurements as a limitation on model representations is discussed. The treatment of uniqueness as a residual from a modelled relationship may conceal information about the uniqueness of catchments, while the treatment of uniqueness as a set of parameter values within a particular model structure is problematic due to the equifinality of model structures and parameter sets. The analysis suggests that a fully reductionist approach to describe the uniqueness of individual catchment areas by the aggregation of descriptions of small scale behaviour will be impossible given current measurement technologies. A suggested strategy for the representation of uniqueness of place as a fuzzy mapping of the landscape into a model space is suggested. This will lead to a quantification of the uncertainty in predictions of any particular location in a way that allows a conditioning of the mapping on the basis of the available data. This process can incorporate a hypothesis testing approach to model evaluation but the problem of multiple behavioural models may provide an ultimate limitation on the realism of process representations: not on the principle of realism but on the possibility of unambiguous process representations.

The character of uniqueness of place

It is a geographical aphorism that all places are unique. This is one reason why the progression from empirical observation and classification to theorising has been relatively slow in many areas of the geographical and environmental sciences including hydrology. Such theorising has also tended to be somewhat speculative, resulting in conceptual models rather than theory, particularly when applied to real, rather than laboratory, scale systems. In such applications, models have been commonly developed as specific tools for specific purposes, reflecting the demand for quantitative predictions that has grown with the increase in the availability and power of computers. This demand has, however, been largely a sociological response to technological advances in the means of making predictions; it has not necessarily been a response to an increasing scientific basis for those predictions. Why? Primarily, as will become clear, because of uniqueness of place.

It is, of course, not only the technology underlying predictive models that has improved. The technology of observation and of recording and analysis of observations has also improved. There is no doubt that this has led to improved scientific understanding of hydrological processes, at least at the small scale. Most of the measurement techniques available to the hydrologist for studying processes directly are limited to the point, plot or patch scale. Detailed studies at such scales have led to an improved recognition of the heterogeneity and complexity of the flow processes (e.g. Flury *et al.*, 1994; Binley *et al.*, 1996; Henderson *et al.*, 1996). Parameters such as hydraulic conductivity, dispersivity and surface roughness have become scale dependent, if they can be calculated at all at larger scales in media with strong preferential flow components (see discussion in Beven, 1996). Values of parameters calculated by the calibration of models have been recognised as being effective values that may not have a physical interpretation outside of the model structure within which they were calibrated, and may be scale dependent even within that structure (e.g. Bruneau *et al.*, 1995; Franchini *et al.*, 1996; Saulnier *et al.*, 1997). Recognition of such small scale complexity and its implications has not really, however, led to an improved scientific approach to dealing with it.

There have been, certainly, some theoretical developments in addressing the problems of heterogeneity, particularly in the field of groundwater flow and transport. Assuming a statistical structure for the spatial heterogeneity of aquifer properties, the resulting scale dependent effective parameters can be inferred (see for example Dagan, 1986). This is an advance, but at the expense of introducing parameters that must be defined at each (unique) site at which the theory is to be applied. The only way of defining such parameters at present appears to be by calibration against observations, as demonstrated, for example, in applications to the results of some of the large scale tracer tests that have become available in recent years. Even so, attempts to model flow and transport given small scale information about aquifer properties have not generally been very successful without some optimisation of parameter values. Moltyaner *et al.* (1993), for example, concluded that it was necessary to use velocities directly derived from tracer measurements rather than estimated from a hydraulic conductivity and gradient using Darcy's Law to obtain adequate predictions of solute transport. This, however, begs the question of how far such velocities may be extrapolated to other flow conditions and nearby sites.

At the catchment scale, there is the continuing problem of the extrapolation of knowledge to sites where no data are available. This regionalisation problem has largely resisted both empirical and physically-based approaches. It is difficult because catchments are unique in their particular characteristics of topography, soil, rocks, vegetation and anthropogenic modification. It would appear that aquifers and catchments remain unique in a way that transcends the most advanced theorising available. In what follows, this problem will be analysed further with the aim of assessing whether this is inevitable, and if so, what approach should be taken to modelling catchments as unique entities.

Uniqueness in field observations

With a few exceptions, measurement techniques in hydrology are point measurement techniques. Very few techniques give a direct representation of the spatial patterns of hydrological behaviour. A discharge measurement is a point measurement that integrates over all the intra-catchment spatial patterns of hydrological behaviour that affect the response characteristics at that point. Not a lot about the internal dynamics of a catchment; can be learned unambiguously from a discharge measurement; perhaps some mean residence times (Jakeman and Hornberger, 1993; Young and Beven, 1994) may be determined but even these may be subject to some nonstationarity and uncertainty. Raingauges, tensiometers, soil moisture probes, piezometers, soil solution lysimeters, are also point measurements, at scales small enough that multiple measurements may reveal considerable local heterogeneity. All will reflect an integration over some effective volume (Cushman, 1994), but in general this will be small compared with the scale of heterogeneity of the flow domain. In essence, the measurements have a uniqueness that reflects the conditions at that particular location (and that particular time of measurement). The difference in scales of measurement and heterogeneity means that the measurement may not always be "representative" of a larger volume of the flow domain.

The extent of this problem varies with the process and flow conditions. Measurements of pressure in saturated conditions, for example, will tend to reflect an effective integration over a much larger volume (because of the rapid dissipation of local pressure differences in saturated flow) than for unsaturated conditions where, in structured soils, the integration volume for a pressure measurement may be very small. When the difference in scales is large, a very large number of samples may be required to obtain an adequate characterisation of conditions. A long time ago, Hills and Reynolds (1969), using a gravimetric analysis of near surface soil samples, estimated that 80 samples would be required to estimate mean surface soil moisture of a catchment to 95% confidence (or 490 samples to 98% confidence). Correlation lengths are expected to be small for such measurements. How many studies since have followed such a recommendation, or even made the same calculation from the smaller number of samples collected? And if spatial correlation is to be taken into account, how many samples (and holes) would be needed, given that there is some evidence that the correlation structure of soil variables may be (empirically) described as fractal (see the review and example data set from the Tarrawarra catchment of Western *et al.*, 1998). In addition, it has been shown that, under certain conditions at least, near surface soil moisture exhibits a structure that reflects that of the topography but that under dry conditions this may degrade to a stochastic pattern without strong spatial structure (again, the Tarrawarra catchment provides a good example, see Grayson *et al.*, 1997).

A further question then arises. If it is not feasible to have an adequate characterisation of the spatial heterogeneity of characteristics and flow processes, by means of the detailed measurement techniques available, how is the uniqueness of individual hillslopes or catchments to be assessed? Some broader scale characteristics can be assessed, certainly. The surface topography, vegetation type, geology and soil classification, for example, are widely available now in geographical information system databases. These databases are now being used to define hydrological response units in modelling (e.g. Schultz, 1994; Flügel, 1996; Becker and Braun, 1999). However, these give information that may be only broadly hydrologically relevant. On some hillslopes, bedrock topography may be more significantly related to flow patterns than surface topography. Soil texture derived hydraulic characteristics (pedotransfer functions) may not give good simulations of soil water behaviour where preferential flows are important. There may also be strong differences in transpiration over short distances within a single vegetation type where downslope flow results in strong differences in water availability during dry conditions. Thus, there may be considerable uncertainty associated with the estimation of parameter values to represent a particular hydrological response unit, or hillslope or catchment, on the basis of such data. The question is not yet answered adequately.

There is a reductionist argument that suggests that if better models for the nature and effects of heterogeneity could be developed, then a better representation of individual sites might be possible. This is one of the arguments made by De Marsily (1994). He uses the example of modelling the structure of fluvial deposition as a way of characterising the structure of certain aquifers (e.g. Gross and Small, 1998). For advancing understanding of flow processes under different circumstances, this may be an important path to pursue but in applications to real aquifers and catchments this argument seems flawed. However such models are constructed, they will have some parameters that will need to be defined for each application site. There may, indeed, be many such parameters, regardless of whether a model structure is a perfect physical representation or whether the model is a simplified, largely statistical, representation of the small scale heterogeneity. In both cases, the uniqueness of a site must be reflected in specific parameter values (or distributions of parameter values) within the model structure.

The reductionist argument suggests that because these parameters represent more fundamental descriptions of nature, they should be easier to define or measure. I do not agree with this argument. I believe that, in a measurement technique limited field of study, such as hydrology, the measurement or unambiguous definition of parameters for small scale descriptions will be impossible. I use unambiguous since one consequence of measurement limitations is that that there may be many different models or parameter sets within a model that might be equally acceptable descriptors of the (often limited) data that are available to any study (Beven, 1993, 1996). It is important to note that this will be true in applications even if the perfect model structure was known (which, as yet at least, it is not).

In fact, at the current state of knowledge, the assumption of specific model structures for the description of the spatial variation of parameter values (as, for example, in Dagan, 1986, and others) may be misleading if the resulting theories are constrained by such assumptions in such a way that the scaling behaviour of a particular site will be wrongly predicted. This is not in any way to deny the value of such studies which represent real progress in trying to understand the scaling and aggregation issues. After all, the assumptions can, in principle, be checked for validity in any particular application. The problem is the effort required to do so in that particular application, such that assumptions, originally made for reasons of mathematical tractability as much as for physical realism, might be accepted without proper testing.

This problem of effort is a fundamental limitation on the use of scaling theories in the prediction of responses for real unique sites. With current and foreseeable measurement techniques, the consequence of this limitation is that the aggregation of the small scale to the large scale will prove generally unachievable (for an extended argument see Beven, 1995). Essentially, there is much about the uniqueness of the subsurface flow domain that will remain unknowable, *a fortiori*, and is revealed only in the responses observed at larger scale, with all the limitations of knowledge that those observations imply. Hence, there are difficulties in predicting discharges as well as the element of surprise associated with any large scale tracer experiment. Such larger scale measurements are, however, unambiguous. They represent the real response (albeit still within the limitations of measurement scales, sampling densities, time resolution and measurement accuracy). The difficulty is then extrapolating from those sets of measurements that are available to applications to other conditions and other sites (as noted above in respect of the use of tracer velocities in the simulations of Moltyaner *et al.*, 1993). For this extrapolation, models are needed but perhaps not necessarily a fully reductionist physical description.

Uniqueness as model residuals

The catchment hydrological system is an open, energy dissipating system whose characteristics are the result of the action of nonlinear interacting processes subject to pseudo-random forcing conditions (together with the effects of man) that have changed over time. The response of such systems is known to be sensitive to the initial and boundary conditions, to the extent that in some such systems chaotic behaviour may ensue. If the responses of hydrological systems are not demonstrably chaotic over short periods of time, it is because the responses are strongly constrained by the boundary conditions associated with individual events. In short, the hydrograph resulting from each event cannot have a volume greater than the rainfall inputs. This still leaves scope for the sensitivity of possible hydrograph volumes and shapes to initial conditions and interactions between processes. In the past such variability has been adjudged as stochastic rather than chaotic and, with limited data, it may be impossible to distinguish between these forms of explanation (see the discussion in Wilcox *et al.*, 1991).

Over longer periods of time, such nonlinear systems have a tendency to display emergent characteristics associated with the concept of self-organised criticality. This appears to be true of hydrological systems in that the continual flux of water, solutes and sediment through the system results in drainage networks that have some (generally dendritic) similarities and are, at least over a certain range of scales, scaling (see, for example, Rodriguez-Iturbe and Rinaldo, 1997). This is one area where theories of scaling have been proposed (e.g. Gupta and Dawdy, 1995). It suggests that there may be a way out of the uniqueness problem if this self-organisation allows the establishment of criteria for similarity and scaling of catchment responses.

The study of scaling behaviour across a wide range of catchments, as in the studies of Gupta *et al.* (1994) and Gupta and Dawdy (1995) is one example of a regionalisation model that seeks to find structure in catchment responses, at least for regions that are "homogeneous" in some sense. As

Gupta *et al.* (1994) note, there may still be a wide degree of variability about the postulated structural relations in the data for individual catchments (their example was for catchments in Applachia; see also the example of instantaneous unit hydrograph (IUH) parameters for Bavarian catchments in Becker and Braun, 1999). Similar variability has been found in the past in attempts to regionalise hydrological responses using multiple regressions against indices of catchment characteristics (e.g. NERC, 1975; Pilgrim, 1987; Post and Jakeman, 1994). The uniqueness of the response of individual catchments in this context, as measured, appears as a residual from a model structure.

Traditional statistical analysis, of course, assigns the residuals to purely stochastic effects but, from a physical viewpoint, there may still be information in the residuals. Why do some catchments tend on average to produce higher peak flows than the regional modal behaviour; why do some catchments tend on average to produce lower peak flows, even after taking account of either catchment characteristics (in so far as they can be represented by the available indices) or network scale transformations (see Acreman and Sinclair, 1986 for a sub-regional examination of this type)? Uniqueness of individual catchments may be one reason, but the phrase "as measured" is also important here, since there may also be effects associated with error in the measurements, particularly for peak flows, the spatial variability of rainfalls in individual events, and the length and period of the record for different catchments which may not be consistent across the set of catchments analysed.

Perhaps because of the implicit assumptions of traditional statistical analysis, there appear to have been few detailed studies of the information content of such model residuals. A detailed examination of the original historical data sources and catchment responses would clearly be far more time consuming than the simple piecewise regression of available variables traditionally undertaken, but it may be in the evidence of differences between catchments due to different initial conditions and trajectories of development, rather than in their similarities, that the hope for improving predictability may lie. The uniqueness of individual catchments lies in these differences. It may be possible to learn from studies of similarity in structure and behaviour at the catchment scale: it may be possible to learn more from the deviations or residuals of individual catchments away from some ideal self-organised structure and behaviour at the catchment scale.

Investigation of these residuals requires some degree of reductionism in the sense that an explanation must be sought in the intra-catchment characteristics and processes of a study basin. This is where the limitations of the available measurement techniques in hydrology become critical, since there is then a jump in what can be properly quantified down to the point scale, as discussed above. The proper characterisation of intra-catchment hydrological response units remains an open problem, especially when each is itself unique. It seems clear to me that characterising this uniqueness will not be possible by the aggregation of small scale process descriptions (though other authors might disagree, e.g. Becker and Braun, 1999).

Uniqueness as model parameters

Another widespread use of models as a means of extrapolating understanding from catchment to catchment is in the form of simulation models. A model structure, found to be acceptable in one catchment, is applied to another catchment. This, however, usually requires the calibration of parameter values for each application site. The uniqueness of the catchment will then be reflected in the parameter values that represent it, whether those parameters be specified by measurement, estimation *a priori*, or calibration against some variables observed at the site. In some studies, attempts have been made to combine parameter estimation with regionalisation of calibrated parameter values for gauged sites by regression against indices of catchment characteristics, with the aim of predicting parameter values for ungauged sites (e.g NERC, 1975; Post and Jakeman, 1994). Such an approach is fraught with difficulties, not least because (again) there may be no single model structure, or parameter set within a model structure, that is an optimum representation of a catchment. Most models have sufficient degrees of freedom that they can be calibrated to give acceptable simulations of the discharge responses of gauged catchments at the catchment scale, at least if acceptable is not defined too tightly! Indeed, there is evidence that many different parameter sets might be considered acceptable if the concept of finding an optimum is relaxed in favour of a concept of equifinality of models (see discussions in Beven, 1993, 1996). Thus, there is an implication that the uniqueness of a particular catchment may not be unequivocally represented by a set of calibrated parameters.

Interpretation of the values of individual parameters must also be made very carefully; the values may have meaning only in the context of a particular model structure and values of the other parameters within that structure. Extrapolation of parameter values to other catchments may then be highly uncertain. Note that in what follows the word model will be used to indicate a particular combination of model structure and set of parameter values within that model structure.

Dealing with uniqueness in the face of uncertainty: beyond a reductionist approach

The discussion above has suggested that:

- As a result of measurement constraints, there may be limitations to the possibility of dealing with uniqueness

of individual catchment areas through a fully reductionist approach and it may be necessary to distinguish between the representation of the effects of small scale heterogeneity for understanding and the description of uniqueness at larger scales for the purposes of prediction.
- It may be as yet very difficult, if not impossible, to describe the uniqueness of individual catchment areas by the aggregation of descriptions of small scale process behaviour.
- There may be information in the interpretation of the uniqueness of individual catchments areas as residuals from regionalised structural relationships or regressions, but there may also be significant residual uncertainty in such relationships and potential information in the residuals.
- It may not be possible to represent the uniqueness of individual catchment areas unambiguously by the calibration of model parameters.

Thus, if a description on the basis of small scale parameterisations is currently impossible, and large scale representations are uncertain, how to proceed in the future to address the requirement of prediction of the process controls and hydrological responses of unique areas? One suggestion is to do so within a framework of model conditioning and rejection. Assume that one or more models have been chosen as potential predictors for an application. The type of model(s) chosen will depend on the aims of the application but might include, for example, empirical regionalisation models for the case of estimating the response of an ungauged catchment, or a more physically-based structure for predicting the response of an area to significant changes. In each case, there will be some parameters involved, but it may be possible to estimate values of those parameters only within some limits of uncertainty. Those limits might specify only some feasible range of parameter values; some estimators might allow the definition of joint prior distributions for parameter values. Running the model or models, taking account of such uncertainty in the parameter values will then give a range of predictions that should reflect the possible effects of uniqueness of a catchment on any variation from an expectation or mean prediction.

It is, of course, possible to stop at that point. The range of predictions, conditioned on the prior knowledge embodied in the prediction methodology, can be used directly in any risk based decision analysis associated with the application. The range of predictions may be expected to be quite wide; however, it might be possible to represent the unique response of the area being studied more closely by further conditioning based on data collected at the site. Different types of data might have different values in such a conditioning process and this focuses attention on what type of data might best be used for conditioning for different types of application or desired predictive capability. It will be demonstrated below that this problem can be posed within a scientific hypothesis testing framework.

Uniqueness as a mapping from the landscape into a model space

If the idea of a single model representation of a unique hydrological system is rejected (the logical consequence of the arguments presented above), then an acceptance of uncertainty in the representation of that system must follow. Traditionally, such uncertainty has been assessed by exploring the sensitivity of predictions to parameter variations around some "optimal" model. The type of conditioning discussed in the previous section is, however, a more flexible and realistic representation of such uncertainty. It also has an interesting interpretation as a mapping of the landscape into a specified model space.

This interpretation follows in a straightforward way from the explicit recognition of the problem of equifinality in representing the uniqueness of individual sites. Accept, for the moment, that the limitations of hydrological process theory are such that it is not possible to identify a single model (model structure or set of parameter values) as a unique representation of a unique catchment area. This may be regarded as a fundamental limitation or as only a temporary limitation resulting from the inadequate means of investigation today that will be overcome in the future: it does not matter to the discussion below.

If, then, it is accepted that it is impossible to identify a single model of the catchment of interest, there must be many models (or parameter sets) that must necessarily be considered the "behavioural" in that they are compatible with the (limited) observations available for that catchment and the perceptions of the processes controlling the response of that catchment. These "behavioural" models will occupy part of what might be called the "model space". The model space will be a high-dimensional space with dimensions for different parameters and, possibly, for different model structures. Behavioural models will not fill the model space (in that some and, perhaps, the majority of models, will be considered as non-behavioural) but may be scattered widely throughout the space rather than restricted to some local region of the model space (see for example, Duan *et al.*, 1992; Beven, 1993; Freer *et al.*, 1996; Franks and Beven, 1997a). The feature that behavioural models have in common is similarity in function; function that is, in some sense, also similar to the real catchment. It is important to note that the fact that there may be many models of similar function, is not only a product of limitations of current knowledge. Even if a perfect model structure that would be an ideal representation of hydrological reality could be defined, that model structure would still have parameters that would need to be specified to represent a particular unique catchment area. For the perfect model structure, the number of parameter values might, indeed, be semi-infinite. There will then be plenty of opportunity for interactions between parameter values such that many different parameter sets will produce similar (perhaps behavioural) functions and predictions within some limit of acceptability. Thus, unless those parameter values can be identified accurately by some independent means, there may be many behavioural models even for the perfect model structure.

Thus, the cloud (or many clouds) of models that are behavioural in representing a catchment under study in effect represents a mapping of the catchment into the model space (Fig. 1). Such a mapping, using fuzzy weights, has been demonstrated for the case of predicting landscape scale latent heat fluxes by Franks and Beven (1997b). The unique catchment or landscape now no longer occupies a unique location in the model space but a fuzzy set of locations. The mapping can, however, be used in predicting the responses under different conditions or for different periods, since the predictions of individual models within the behavioural set can be used to form a cumulative distribution function of predictions from which any quantiles can be extracted to calculate prediction bounds in a way similar to the Monte Carlo sampling of the Generalised Likelihood Uncertainty Estimation (GLUE) methodology of Beven and Binley (1992).

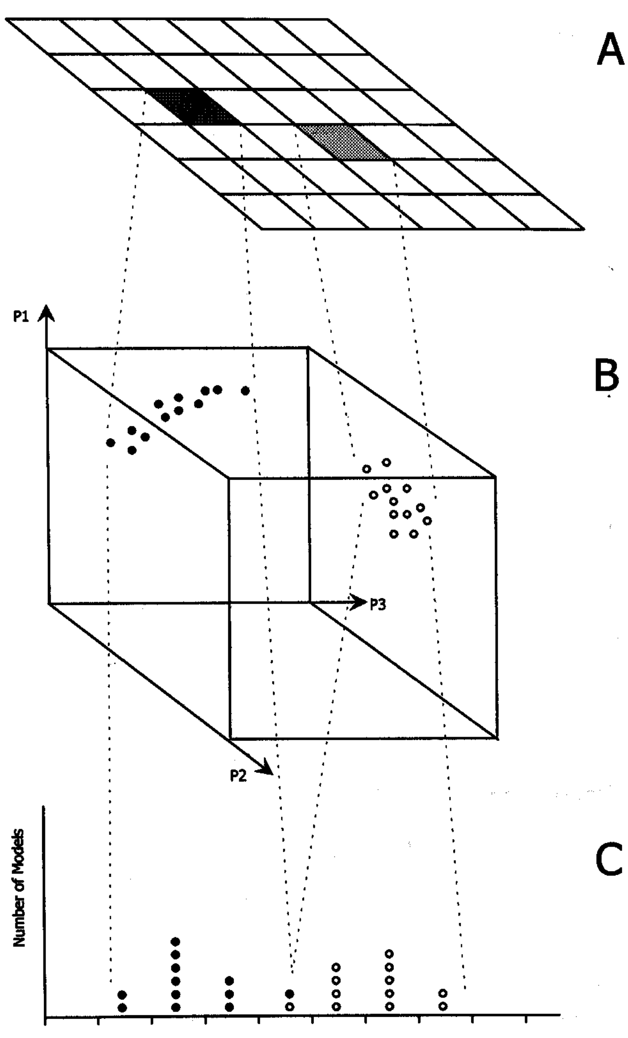

Fig. 1. A. *A landscape space represented as pixels or patches.* B. *Representation of the pixels or patches in a model space with N parameter dimensions (here N = 3). Each dot represents a model with a set of parameter values for which the simulated variables are consistent with the data available for the pixel.* C. *Grouping of models into functional classes according to similarity of predictions for some variable of interest. Note that models compatible with the data for a pixel may map into different functional classes (see, for example, Franks and Beven, 1997b).*

Different (unique) catchments will generally map into different fuzzy sets in the model space. The intersection of the sets associated with different catchments will reflect their similarities or differences in function in a far more realistic way than an assessment based on "optimal" parameter values.

The concept of such a mapping is not difficult to grasp but requires a recognition that the ultimate realist goal of a "true" representation of hydrological processes and catchment responses may not be achievable. The hydrologist may have a good qualitative perception of the physics of the processes involved, but the heterogeneities and complexities of the flow domain and the resulting process interactions are such that a fully realist mathematical description seems likely to remain elusive for the foreseeable future (see discussions in Morton, 1993; Oreskes *et al.*, 1994; Beven, 1995, 1996). The proposed approach is not inconsistent with a realist philosophy or with the ultimate aim of a "true" representation of hydrological processes. In fact, despite the fuzziness of the mapping being advocated here, the approach can be formulated within a fully scientific hypothesis testing framework (see below).

Mapping into a model space by monte carlo simulation

It might be accepted that, in principle, there are many models within the model space that are compatible with the available knowledge and observations of a particular hydrological system, but there is still then a problem of identifying where in the model space those models lie. A general, if not always practical, approach to this problem is the type of uniform Monte Carlo simulation approach used in the GLUE methodology of Beven and Binley (1992). For each parameter (within each model structure if more than one is considered) a range of feasible values for that parameter is specified. Random sets of parameters are generated uniformly within the model space and run within the appropriate model structure. The results are assessed to see whether that model will be accepted as behavioural.

This type of sampling of the model space is easy to implement and requires minimal prior assumptions about the distribution and covariation of the parameters. The practical difficulty of this simple approach is that, as the complexity of the model and number of parameter dimensions increases, very large numbers of simulations are required to achieve more than a very coarse sampling of the model space. Ideally, the most efficient sampling strategy would be to sample only where behavioural simulations would be expected in the model space, but this may be very difficult to achieve in practice since experience suggests that behavioural simulations may be scattered through the model space as a result of local parameter interactions.

If there is some prior knowledge about the nature of parameter interactions, or combinations of parameter values that should not be considered feasible, it can be used to guide the sampling but in general this will not be the case. Methods of importance sampling, such the various methods of Monte Carlo Markov Chain (MC^2) sampling, can be used to refine the sampling as more is learned about the nature of the model space (see, for example, Sen and Stoffa, 1996). Most MC^2 methods, however, require restrictive assumptions about the nature of the space (such as multivariate normal parameter distributions) that may not be borne out by a detailed exploration of the space that often shows highly complex variations in the behavioural measures. There are other methods of refining the sampling strategy, such as the Tree Structured Density Estimation approach of Spear *et al.* (1994) that are essentially assumption free. These use an initial exploration of the pattern of behavioural simulations in model space to guide later sampling iterations to increase the density of sampling in the areas where there is the greatest probability of finding behavioural simulations. In this way, the number of simulations run that are found to be non-behavioural may be greatly reduced (but at the risk of leaving some areas of behavioural simulations undetected). Such methods require further exploration.

However, with simple models it is now possible to make many thousands of simulations runs, particularly using distributed memory parallel computers. Providing a single run of the model will fit on a single processor, these systems are ideally suited to Monte Carlo simulation. At Lancaster, a 20 processor parallel PC system running under Linux, uses Pentium II processors each with 64 Mbyte of memory linked by a fast 100 MHz Ethernet network. This system is capable of running at speeds of 2.5 Gflops. At less than £1000 stg. per node, such a system is highly cost effective for this type of calculation. It does, however, produce a very large volume of results to be analysed!!

Fuzzy mapping into the model space and prediction uncertainty

Once such a mapping has been achieved it can be used for prediction. Interestingly, if the chosen model is deterministic, the deterministic nature of the model predictions is retained i.e. for any input sequence, the outputs from any parameter set in the model space are known precisely. This has been used by Franks and Beven (1997b) to reduce the number of model prediction runs required by analysing the model space to identify different functional behaviours (Beven and Franks, 1999). It is, after all, the prediction of the functioning of a particular unique part of the landscape that we are trying to represent and it may be possible to define similarity in function within the model space. The fuzzy weights can then be used in a manner analogous to that of the GLUE methodology of Beven and Binley (1992) to weight the predictions of the associated models, forming a cumulative weight distribution function of the predicted variables of interest. Modal and quantile values from this distribution can then be used to represent the uncertainty in the predictions.

Uniqueness, process representations, data collection and hypothesis testing in hydrology

It will generally be found in any such exploration of the model space that there are many different behavioural models that are, in some sense, acceptable in describing and predicting the responses of the hydrological systems in which we are interested. We have noted that the set of behavioural models is necessarily defined in terms of describing a particular hydrological system and consequently must share a similarity in function to that real system (as defined in some practically useful way). The many behavioural models may, however, produce similar functionality in different ways, for example with different proportions of surface and subsurface flow pathways in predicting stream discharges. Thus, there is the opportunity to examine the collection of behavioural models and propose testable hypotheses based on process implications. The behavioural models could, in this sense, be considered as multiple working hypotheses. A critical test, involving the collection of additional data or additional types of data, might then, at least for well formulated tests, eliminate some of those possibilities as incompatible with the new knowledge of the processes.

The key, then, is the collection of data. Data that will allow some of the potential responses to be eliminated as unlikely and will condition the possibility or probability weights associated with others in the model space mapping. Either Bayesian or Fuzzy conditioning procedures could be used at this point (e.g. Freer *et al.*, 1996; Franks *et al.*, 1997a; Franks *et al.*, 1998). The value of additional data in conditioning the feasible responses can be evaluated using different uncertainty measures (see for example Beven and Binley, 1992). Different types of data may be more or less effective in such conditioning and, in particular, in the rejection of different models. Indeed, this could be formulated as an interactive process, where the range of model predictions is used to decide the most cost-effective data collection/hypothesis testing strategy for conditioning the prior predictions to a particular site. This would be expected to lead to some of the models considered behavioural until the new data are available being rejected or falsified.

It is not often, however, such a controlled process. More usually opportunist use will be made of whatever data are available within the time and financial constraints available. Only rarely is it possible to plan experiments with such hypothesis testing in mind and, even then, the measurement techniques available may not support the critical tests needed. However, there is some evidence that much might be gained from looking at different types of data. Franks *et al.* (1998), for example, in an application of TOPMODEL to the Naizin catchment in Brittany were able to use the capability of TOPMODEL to predict saturated contributing areas in conjunction with some very imprecise data on contributing areas obtained from multiple ERS-1 radar data together with some local ground observations. The result of using this additional data was to reject many of the models considered behavioural only on the basis of a comparison of observed and predicted discharges. A transmissivity parameter, in particular, was dramatically constrained in its behavioural range in the model space.

Thus, it would appear that there is value in gaining additional information from more detailed process studies in constraining the possible behavioural models. There are, however, two further considerations that make such constraints within a hypothesis testing framework difficult. The first is again concerned with the uniqueness of catchments and of locations within catchments. Collecting data about processes generally means restricting measurements to a number of locations that may not have the same characteristics as other similar locations within a catchment. Thus, if the interest is in testing the feasible range of a particular parameter, such as transmissivity, at a site then it does not follow that the same range will apply at other sites. There is the potential for local variations in such values (as is expected from our perceptual model of the catchment system). Local variations in parameter values, however, essentially add dimensions (and the potential for additional interactions between parameters) to the model space. Thus, although information is being added, the value of that data will not necessarily have other than local value. The use of the remotely sensed contributing area information by Franks *et al.* (1998) was successful in this respect because it provided a global constraint, despite the limited accuracy in estimating the catchment scale contributing areas.

The second problem is that testing of models with respect to more detailed process information might lead to the rejection of all the available models as non-behavioural. The point has been made many times in the literature that the calibration of catchment rainfall–runoff models to reproduce catchment discharges to some level of acceptability is generally not a difficult problem. Even lumped models with 3 or 4 parameters can often achieve this. Reproduction of catchment discharges does not imply an adequate reproduction of flow processes, however, even in terms of a gross differentiation between surface and subsurface runoff generation (which is, after all, not always necessarily easy to observe in the field). One example is provided by the attempts by Lamb *et al.* (1998) to model the behaviour of the small Saeternbekken catchment in Norway, again using a variant on TOPMODEL. This catchment has observations on over 100 piezometers made at a number of different flow stages. The model does well in reproducing the catchment discharges using catchment scale effective parameters but not so well in reproducing the spatial patterns of water table measurements. Using one of the sets of piezometer measurements to derive local transmissivity parameters for those points, helps improve the predictions at many of those points but there are still some points that are not well predicted, and there is still a significant part of the catchment for which no measurements were made and where there will remain significant uncertainty in transmissivity. The additional conditioning on the local water table measurements in this catchment in fact makes little difference to the uncertainty in predicting the catchment discharges. Thus, while the model has important predictive capability and, while allowing for local transmissivity values gives an important improvement in predicting the pattern of water table variations, all the models still fail to reproduce all the available data satisfactorily. In one sense, therefore all the models are non-behavioural.

In fact, we would surely expect that a catchment scale model is unlikely to predict the hydrology at all points in the catchment precisely. In the Saeternbekken study, TOPMODEL is recognised as representing only an approximate description of the flow processes but it is difficult to envisage even the most detailed physically based models faring any better in this respect. Even our most detailed models, when applied at the catchment scale, have limitations in their process descriptions. The retention of some models will, however, be necessary to retain some predictive capability. The definition of a behavioural model will often, therefore, be a compromise between the success of the model in reproducing some observations and the failure to reproduce other observations, where different observations might have different degrees of importance depending on measurement accuracy or local rather than global scope or relevance. This implies that hypothesis testing might be used as a methodology, but that a declaration of success in such tests may be based on a relative rather than absolute measure of acceptability and might not always be as rigorous as would be liked. The question of what constitutes a behavioural model has, as yet, not been addressed in a satisfactorily rigorous way. There are some analogies here with some of the philosophical debate over the issue of falsification in theory testing. There is no doubt that in any environmental modelling exercise, if a single false prediction were to be grounds for falsification, then no predictive models would be left. Some more relaxed criteria for acceptability are necessary. The type of hypothesis testing approach outlined here at least represents a framework for addressing this type of problem.

Implications for process studies

Some rather important implications for process studies arise from this discussion. Field process studies are by their nature unique in both space and time; they cannot be repeated under exactly the same boundary and initial conditions. Much may be learned from process studies in terms of the qualitative nature of hydrological responses. The difficulty comes in using that knowledge in quantitative predictions under different conditions or for different sites. The usual route for doing so involves the intermediate step of a model with its parameters to be determined for the study site and estimated for any other site.

In the past, however, the reporting of field process studies in the literature has generally involved the reporting of single values of parameters. These will have been inferred from some calibration process, either by back-calculation directly from observations (such as the calculation of a hydraulic conductivity from a known flux and two hydraulic potential measurements to calculate a gradient) or inferred from the optimisation of a model structure. The former can be used only in very simple cases and, when applied in the field rather than the laboratory, can be fraught with difficulties (such as the not uncommon case of apparent negative hydraulic conductivities where a downwards vertical flux is expected following rainfall but two tensiometers are still recording an upward potential gradient). The latter will be subject to the type of equifinality of parameter sets that has been discussed in this paper. An optimum value of each parameter can always be found, but they may not be very robust with respect to either different data sets, changes in the values of other parameters or variations on the model structure.

The result is some doubt about the physical meaning of parameter values, if they are depending on model structure, model scale, observational period or the values of other parameters. The implication is that great care needs to be taken in adopting values of individual parameters from other sources without assessing their means of derivation. One good example of this is the type of pedotransfer function regressions that are now being used to provide estimates of soil hydraulic characteristics on the basis of more easily measured textural information (e.g. Rawls and Brakensiek, 1989). The data on which these regressions are based were originally collected on small samples, with hydraulic conductivity measurements being made on "fist-sized fragments" in the laboratory. Usually, only one measurement was recorded for each depth in the profile at a measurement site. Thus, the values derived may not be appropriate to be used as the "hydraulic conductivities" required in distributed hydrological models that have an element scale of metres or more since the resulting values cannot reflect the effects of larger scale flow pathways. In some soils, this may not be important, in others, it will be critical to a proper prediction of the flow processes. Similar considerations apply to other "physical" hydrological or hydraulic parameters such as roughness, porosity and dispersivity.

Some types of process predictions require parameter values to be taken from a wide variety of published and other sources to be used in a model structure that may be different from that in which the parameter values were originally determined. The type example is a land surface to atmosphere transfer parameterisation for use within atmospheric circulation models. These models have become more and more complex as computing constraints have become less limiting. The more complex models have introduced more and more parameters that must be specified before the model can be run. Computer limitations are still such that for each grid square in the atmospheric model, not all the variations in the land surface can be represented. Typically, the current generation of models will represent the variability of the surface in terms of a small number of tiles or patches of different vegetation types (parameter sets) but for each vegetation type a single parameter set must be specified. Estimates of parameter values come from many different sources and, in general, there will be no observations available with which to check the success of the model over each patch in each (very large) grid square. There are no practical measuring techniques for doing so.

Even if the model is a reasonable (or perfect) representation of the physics involved, putting together a parameter set from many different sources is analogous to choosing a single point in the model space. Monte Carlo explorations of the parameter space suggest that for any single parameter value, there will be good simulations and there will be poor simulations depending on the values of other interacting parameter values. There is, therefore, in using parameter estimates taken from different studies at different sites under different conditions and applying them at a larger scale, a distinct possibility that a non-behavioural model will result. This expectation is confirmed by the results of the series of PILPS intercomparisons of many different land surface parameterisations. Running each parameterisation (model structure) with estimated parameter values results in a very wide range of predictions; some "behavioural", some not. In this case, the parameterisations vary in structure as well as parameter values but the same effect should be expected once a "Community Land Model" parameterisation is available (as has been proposed by the NCAR CLM working group in the USA).

We are, of course, in a transitory phase here, gradually learning more about appropriate model structures and appropriate model parameter values as more studies are made. The general feeling is still that with hard won experience there will be a gradual improvement in the realism of model parameterisations, in the choice of parameter values and in predictive capability. The estimation of a "good" set of parameter values will then gradually become less problematic. A science of hydrology is otherwise impossible. But is this really true, or is the problem of equifinality and multiple behavioural models endemic to hydrological and other types of environmental modelling? It has already been stressed that even a perfect model which has parameter values that must be specified for each unique location, will be subject to equifinality and uncertainty of parameter values arising from the uniqueness of place, scale effects and measurement errors. In short, uniqueness of place would appear to impose limitations on realism; not on the principle but on the possibility of an unambiguous realistic representation. The major question for the future is then how far data and new measurement techniques to investigate the characteristics and responses of a particular place can overcome this problem.

Acknowledgements

The ideas on which this paper were based were explored with Professeur Invité at the Ecole Polytechnique Fédéral de Lausanne. The hospitality and support provided by André Musy and the HYDRAM group are greatly appreciated. Some of the results presented are from the work carried out by Stewart Franks and Rob Lamb while they were at Lancaster. Invitations to make presentations at EGS, Nice 1998 from Michel Vauclin and Sue White and comments from Günter Blöschl and an anonymous referee helped clarify some of the issues but I claim sole responsibility for the confusions that remain.

References

Acreman, M.C. and Sinclair, C.D., 1986. Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland, *J. Hydrol.*, **84**, 365–384.
Becker, A. and Braun, P., 1999. Disaggregation, aggregation and spatial scaling in hydrological modelling, *J. Hydrol.*, **217**, 239–252.
Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modeling, *Adv. Water Resour.*, **16**, 41–51.
Beven, K.J., 1995. Linking parameters across scales: subgrid parameterisations and scale dependent hydrological models, *Hydrol. Process.*, **9**, 507–525.
Beven, K.J., 1996. A discussion of distributed modelling, Chapter 13A, In: *Distributed Hydrological Modelling*, J-C. Refsgaard and M.B. Abbott (Eds.) Kluwer, Dordrecht, 255–278.
Beven, K.J. and Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, **6**, 279–298.
Beven, K.J. and Franks, S.W., 1999. Functional similarity in landscape scale SVAT modelling, *Hydrol. Earth Sys. Sci.*, **3**, 85–94.
Binley, A.M., Henry-Poulter, S. and Shaw, B., 1996. Examination of solute transport in an undisturbed soil column using electrical resistance tomography, *Wat. Resour. Res.*, **32**, 763–769.
Bruneau, P., Gascuel-Odoux, C., Robin, P., Merot, Ph. and Beven, K.J., 1995. Sensitivity to space and time resolution of a hydrological model using digital elevation data, *Hydrol. Process.*, **9**, 69–82.
Cushman, J.H., 1994. On unifying the concepts of scale, instrumentation, and stochastics in the development of multiphase transport theory, *Wat. Resour. Res.*, **20**, 1668–1676.
Dagan, G., 1986. Statistical theory of groundwater flow and transport: pore to laboratory; laboratory to formation and formation to regional scale, *Wat. Resour. Res.*, **22**, 120–135.
De Marsily, G., 1994. Quelques réflexions sur l'utilisation des modéles en hydrologie, *Rev. des Sciences de l'Eau*, **7**, 219–234.
Duan, Q., Sorooshian, S. and Gupta, V., 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models, *Wat. Resour. Res.*, **28**, 1015–1031.
Flügel, W.A., 1996. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany, *Hydrol. Process.*, **9**, 423–436.
Flury, M., Flühler, H., Jury, W.A. and Leuenberger, J., 1994. Susceptibility of soils to preferential flow of water: a field study, *Wat. Resour. Res.*, **30**, 1945–1954.
Franchini, M., Wendling, J., Obled, C. and Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL, *J. Hydrol.*, **175**, 293–338.
Franks, S.W. and Beven, K.J., 1997a. Bayesian estimation of uncertainty in land surface-atmosphere flux predictions, *J. Geophys. Res.*, **102**, 23991–23999.
Franks, S.W. and Beven, K.J., 1997b. Estimation of evapotranspiration at the landscape scale: a fuzzy disaggregation approach, *Wat. Resour. Res.*, **33**, 2929–2938.
Franks, S.W., Gineste, Ph., Beven, K.J. and Merot, Ph., 1998. On constrainting the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process, *Wat. Resour. Res.*, **34**, 787–797.
Freer, J., Beven, K.J. and Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff production and the value of data: an application of the GLUE approach, *Wat. Resour. Res.*, **32**, 2161–2173.
Grayson, R.B., Western, A.W., Chiew, F.H.S. and Blöschl, G., 1997. Preferred states in spatial soil moisture patterns: local and nonlocal controls, *Wat. Resour. Res.*, **33**, 2897–2908.
Gross, L.J. and Small, M.J., 1998. River and floodplain process simulation for subsurface characterisation, *Wat. Resour. Res.*, **34**, 2365–2376.
Gupta, V.K., Mesa, O. and Dawdy, D.R., 1994. Multiscaling theory of flood peaks: regional quantile analysis, *Wat. Resour. Res.*, **30**, 3405–3421.
Gupta, V.K. and Dawdy, D.R., 1995. Physical interpretation of regional variations in the scaling exponents of flood quantiles, *Hydrol. Process.*, **9**.
Henderson, D.E, Reeves, A.D. and Beven, K.J., 1996. Flow separation in undisturbed soil using multiple anionic tracers (2) steady state core scale rainfall and return flows, *Hydrol. Process.*, **10**, 1451–1466.
Hills, R.D. and Reynolds, S.G., 1969. Illustration of soil moisture variability in selected areas and plots of different size, *J. Hydrol.*, **8**, 27–47.
Jakeman, A.J. and Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model?, *Wat. Resour. Res.*, **29**, 2637–2650.
Lamb, R., Beven, K.J. and Myrabø, S., 1998. Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, *Adv. Water Resour.*, **22**, 305–317.
Moltyaner, G.L., Klukas, M.H., Wills, C.A. and Killey, R.W.D., 1993. Numerical simulations of Twin Lake natural-gradient tracer tests: a comparison of methods, *Wat. Resour. Res.*, **29**, 3433–3452.
Morton, A., 1993. Mathematical models: questions of trustworthiness, *Brit. J. Phil. Sci.*, **44**, 659–674.
NERC, 1975. *The Flood Studies Report* (5 volumes), Wallingford.
Oreskes, N., Schrader-Frechette, K. and Belitz, K., 1994. Verification, validation and confirmation of numerical models in the earth sciences, *Science*, **263**, 641–464.
Pilgrim, D.H., 1987. *Australian Rainfall and Runoff, A Guide to Flood Estimation*, The Institution of Engineers, Australia.

Post, D.A. and Jakeman, A.J., 1994. Relationships Between Catchment Attributes and Hydrological Response Characteristics in Small Australian Mountain Ash Catchments, *Hydrol. Process.*, **10**, 877–892.

Rawls, W.J. and Brakensiek, D.L., 1989. Estimation of Soil Water Retention and Hydraulic Properties, in *Unsaturated Flow in Hydrologic Modeling: Theory and Practice*, Morel-Seytoux, H.J. (Ed.), 275–300, Kluwer Academic, Dordrecht.

Rodriguez-Iturbe, I. and Rinaldo, A., 1997. *Fractal River Basins: Chance and Self-Organisation*, Cambridge University Press, 588pp.

Saulnier, G-M., Beven, K.J. and Obled, Ch., 1997. Digital elevation analysis for distributed hydrological modelling: reducing scale dependence in effective hydraulic conductivity values, *Wat. Resour. Res.*, 33, 2097–2101.

Schultz, G.A., 1996. Meso-scale model of runoff and water balance using remote sensing and other GIS data, *Hydrol. Sci. J.*, **39**, 121–142.

Spear, R.C., Grieb, T.M. and Shang, N., 1994. Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, **30**, 3159–3170.

Western, A.W., Blöschl, G. and Grayson, R.B., 1998. Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment, *J. Hydrol.*, **205**, 20–37.

Wilcox, B.P., Seyfried, M.S. and Matison, T.H., 1991. Searching for chaotic dynamics in snowmelt runoff, *Wat. Resour. Res.*, **27**, 1005–1010.

Young, P.C. and Beven, K.J., 1994. Data-based mechanistic modelling and the rainfall-flow non-linearity, *Environmetrics*, V. 5, 335–363.