Journal of Hydrology, 105 (1989) 157–172
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

[4]

CHANGING IDEAS IN HYDROLOGY — THE CASE OF PHYSICALLY-BASED MODELS

KEITH BEVEN

Centre for Research on Environmental Systems, Institute of Environmental and Biological Sciences, University of Lancaster, Bailrigg, Lancaster LA1 4YQ (U.K.)

(Received March 28, 1988; accepted for publication April 6, 1988)

ABSTRACT

Beven, K., 1989. Changing ideas in hydrology — The case of physically-based models. J. Hydrol., 105: 157–172

This paper argues that there are fundamental problems in the application of physically-based models for practical prediction in hydrology. These problems result from limitations of the model equations relative to a heterogeneous reality; the lack of a theory of subgrid scale integration; practical constraints on solution methodologies; and problems of dimensionality in parameter calibration. It is suggested that most current applications of physically-based models use them as lumped conceptual models at the grid scale. Recent papers on physically-based models have misunderstood and misrepresented these limitations. There are practical hydrological problems requiring physically-based predictions, and there will continue to be a need for physically-based models but ideas about their capabilities must change so that future applications attempt to obtain realistic estimates of the uncertainty associated with their predictions, particularly in the case of evaluating future scenarios of the effects of management strategies.

INTRODUCTION

This essay is concerned with the progress of ideas concerning computer modelling of catchment response. It is not, however, a history but rather a critique. The analysis has been prompted by a number of recent papers outlining recent developments in physically-based models of catchment hydrological response. The comments that follow are based in an active interest in the development and application of this type of model, but particularly in a concern about the dangers of uncritical applications. There has been considerable recent activity in physically-based modelling of catchment hydrology. The intention of this paper is to offer some comments on the nature of the activity and change ideas about how this activity should be pursued in the future. The critique is therefore offered in a spirit of encouraging a realistic attitude towards the abilities and achievements of physically-based models.

The aims of developing a physically-based model are laudable. In Abbott et al. (1986a) the authors point out the limitations of the previous generation of lumped parameter models. They need "sufficiently long" meteorological and hydrological records for their calibration which may not always be available; their calibration involves curve fitting making physical interpretation of the fitted parameter values very difficult; predicting effects of land use change by changing parameter values cannot therefore be done with confidence; and such models have not generally made use of data such as topography, soil type and patterns and changes of vegetation types. Physically-based distributed models, on the other hand "can in principle overcome many of the above deficiencies through their use of parameters which have a physical interpretation and through their representation of spatial variability in the parameter values" (Abbott et al., 1986a).

The words "in principle" are significant here. They have been used many times in writing about physically-based models (e.g., Beven and O'Connell, 1982; Beven, 1983, 1985). The implication is that the theoretical advantages of physically-based models remain unproven in practice. Some of the reasons why this might be so are discussed below. Many authors, however, remain optimistic: "the use of parameters with physical significance and the ability to improve calibrations on the basis of physical reasoning represents an advantage over simpler models and should promote confidence in the modelling system" (Bathurst, 1986a).

I have considerable concern about the practical application of the current generation of physically-based models. We are reaching a time when computing power will no longer be a constraint on the application of this type of model. The Institute of Hydrology Distributed Model (IHDM) is already running (albeit relatively slowly) on a personal computer (see Beven et al., 1987). Software packages will soon be available to consulting engineers to allow such models to be used in a wide range of applications. There is a great danger that the theoretical rigour that underlies these models will engender uncritical belief in their predictions. My belief is that those predictions will often be considerably in error, despite or even because of the so-called "physical basis" of the model. In order to give substance to that belief we must first briefly consider the nature of simulation and the practical application of models in general.

ON MODELS AND SIMULATION IN HYDROLOGY

There are two main aims of simulation models. The first is to explore the implications of making certain assumptions about the nature of the real world system; the second is to predict the behaviour of the real world system under a set of naturally-occurring circumstances. It is very important to distinguish these two objectives. Applications of the so-called physically-based models have to date been primarily concerned with the former with the aim of improving understanding of a particular theoretical construct (see e.g., Freeze, 1972, 1980; Beven, 1977; Smith and Hebbert, 1979). There is an underlying implication that one is by analogy also extending understanding of the real world prototype. Such conclusions must be drawn with great care. Our simulation models, even the most complex available such as the Système Hydrologique Européen (SHE) model (Abbott et al., 1986a, b; Bathurst, 1986a, b), are extreme simplifications of reality. We know that the descriptive equations that underly these models are good descriptors of processes occurring in well defined, spatially homogeneous, structurally stationary *model* catchments and hillslopes in the laboratory. We can feel less assured that those equations may describe the complex three-dimensional spatially heterogeneous and time varying system that is a real catchment. This has been illustrated in the (rather rare) studies that have attempted to compare the predictions of physically-based models with detailed hydrological response data (e.g. Stephenson and Freeze, 1974). There remains much that is only poorly understood about the way that catchments respond to rainfall (see e.g. the discussion in Beven, 1987).

Comparison of predicted and observed hydrographs (as for example in Rogers et al., 1985; Bathurst 1986a) is a necessary test but cannot be considered a sufficient test of models that purport to simulate the internal responses of a catchment. To predict the discharge response of a real world catchment to a storm rainfall is not actually very difficult. All that is needed is a loss function and a routing function. A very simple model will satisfy these requirements. The difficulties of hydrological simulation arise due to the non-linearities inherent in these functions, especially in the loss function, as a result of antecedent conditions and the spatial complexities of the catchment topography, soils, vegetation and rainfall inputs. It is therefore much more difficult to build a model capable of accurately simulating the response to a number of different events. This has been amply demonstrated by Loague and Freeze (1985), although these authors allowed only a cursory treatment of initial conditions in their model comparison.

Consider some of the characteristics of hydrological models that can be inferred from experience in using lumped conceptual models. We know that:

(a) The equations of a lumped conceptual model can only be an approximate representation of the real world and must introduce some error arising from model structure.

(b) Spatial heterogeneities in system responses may not be well reproduced by catchment average parameters (see e.g., Sharma and Luxmoore, 1979; Freeze, 1980; Philip, 1980, and others).

(c) The accuracy with which a model can be calibrated or validated is very dependent on errors in the observations of both inputs and outputs (e.g., Ibbitt, 1972; Hornberger et al., 1985). Many input variables, particularly evapotranspiration estimates, may be subject to considerable uncertainty.

(d) There is a great danger of overparameterisation if it is attempted to simulate all hydrological processes thought to be relevant, and fit those parameters by optimisation against an observed discharge record (e.g. Hornberger et al., 1985). It appears that three to five parameters should be sufficient to reproduce most of the information in a hydrological record.

(e) The calibrated parameters of such models may be expected to show a degree of interdependence, so that equally good results may be obtained with different sets of parameter values. This may be true even though a model has only a small number of parameters (e.g., Ibbitt and O'Donnell, 1971; Pickup, 1977; Sorooshian and Gupta, 1983).

The reader may wish to add some other characteristics in the light of her/his own experience. These will be sufficient to evaluate the widely held belief that physically-based models are in some sense superior to simpler lumped conceptual models.

PHYSICALLY-BASED MODELS AND THE LUMPING OF SUBGRID PROCESSES

Many authors do not make clear the distinction between the two aims of simulation outlined above. This is apparent for example in the recent sequence of papers on the SHE model (Abbott et al., 1986a, b; Bathurst 1986a, b). This model (and all other physically-based models) makes a certain set of assumptions about how a hydrological system operates. However, there is a danger of accepting that the equations based on that set of assumptions are good descriptors of reality because they are physically-based. There are a number of reasons to question such a belief. The physics on which the equations are based is the small scale physics of homogeneous systems. In real applications of physically-based models we are forced to lump up the small scale physics to the model grid scale, for example 250 × 250 m in the case of the SHE model as applied to the Wye catchment (Bathurst, 1986a, b). There is no theoretical framework for carrying out this lumping of subgrid processes for spatially heterogeneous grid squares. It is merely assumed that the *same* small scale physical equations can be applied at the model grid scale with the *same* parameters. In doing so we are making a conceptual leap.

It may be argued that such models remain *more* physically-based than simpler models lumped to the catchment scale but it is easy to demonstrate the conceptual nature of current physically-based models. For example, what does a grid square average capillary potential mean physically? There is certainly no way we can go out and measure it. Worse, what does a grid square average capillary potential *gradient* mean when it is calculated from nodes 0.05 m apart in the vertical over an area of 62,500 m^2 (as in Bathurst 1986a, b)? In the model it is used to calculate the grid square vertical flux, but in reality we know that flux must be highly spatially variable due to heterogeneities in both hydraulic gradients and conductivities (see e.g. the problems encountered by Luxmoore et al., 1981). Using grid square effective parameter values implicitly assumes that a grid square is homogeneous. This will not be true and these models must suffer from the same problems that have been demonstrated at the hillslope and catchment scale (point b) above. The model cannot, for example, predict overland flow occurring over part of the grid square, nor the differential velocities of soil water movement towards the water table that may be important in the subsurface response of a grid square of that size.

These examples are used here to illustrate the results of the implicit lumping of subgrid processes inherent in physically-based models. We cannot be sure that the equations will be the same at the grid scale, nor whether effective grid scale parameters can be defined. For now, it is sufficient to conclude that the current generation of distributed physically-based models are *lumped conceptual models*.

EFFECTIVE PARAMETER VALUES AT THE MODEL GRID SCALE

There is no theory for the lumping of subgrid processes in hydrology, although the first steps are being taken in this direction (e.g. Cushman, 1986). A number of recent papers have, however, examined the effect of spatial variability of parameters on hillslope and catchment responses (e.g., Sharma and Luxmoore, 1979; Smith and Hebbert, 1979; Freeze, 1980; Sharma et al., 1987; Binley et al., 1988a, b). There have also been a wealth of papers concerning similar problems in groundwater systems (e.g. Bakr et al., 1978) and some in soil water (e.g., Philip, 1980; Yeh et al., 1985). Many of these studies have concluded that it is not possible to define a consistent effective parameter value to reproduce the response of a spatially variable pattern of parameter values. The primary reason is that a single parameter value cannot reproduce the heterogeneity of responses engendered by the variable catchment characteristics. As an example, if overland flow is generated on part of a soil unit (or grid square) that has relatively low hydraulic conductivity, it might not be predicted at all by a higher effective parameter value for that unit. This suggests therefore that it is not possible to use the small scale physics equations at the grid scale; that we should be developing more complex equations that take account of the effects of such heterogeneity (but in consequence have more parameter values to describe that heterogeneity).

However, we may not have to take such a bleak view of the concept of effective parameter values. Most of the studies of the effects of spatial variability on hydrological responses have used as input data statistical distributions of parameter values based on field measurements. The scale of those field measurements is invariably small relative to model grid scales (at most 2 × 2 m). The flow processes at the scale of model grid squares or elements will naturally integrate over some of the variability inherent at the smaller scales. Surface runoff produced on a small area of low conductivity may infiltrate on an adjacent area of higher conductivity within a grid element.

A *qualitative* impression of the effect of scale on the variability to be expected in parameter values at the model grid scale can be obtained using block kriging (see e.g. Journel and Huijbregts, 1978). The spatial variability of the hydraulic characteristics of soils have been shown by experimental studies to have structure in the sense that measured values are correlated in space (e.g., Viera et al., 1981; Russo and Bresler, 1982). Thus spatially averaged values will also be correlated in space. If the spatial variability of a particular characteristic can be assumed to be a stationary random field, then the spatial structure may be expressed as a point semivariogram, $\gamma(h)$ where h is the

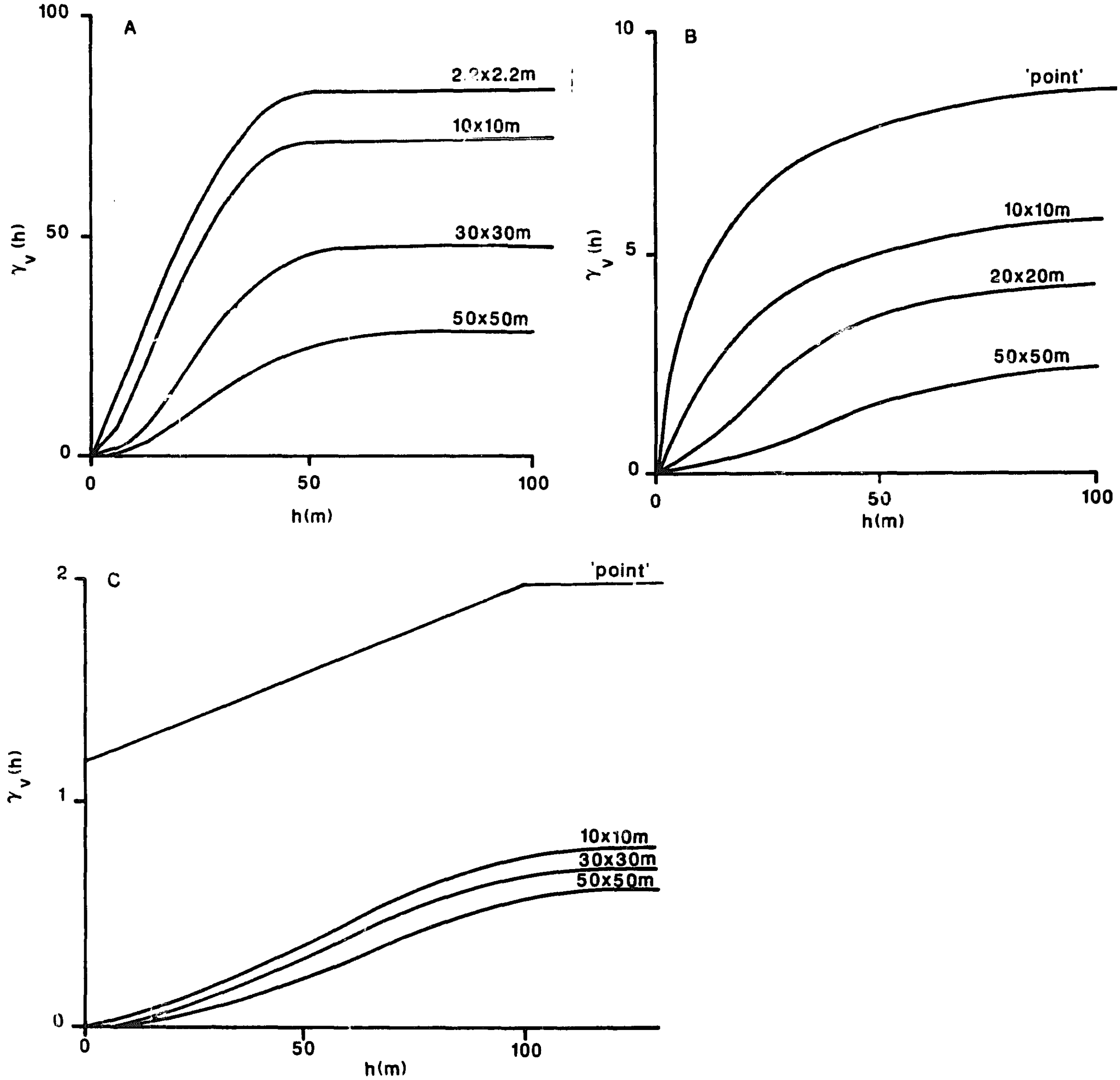

Fig. 1. (A) Variograms for different areal supports derived from hydraulic conductivity data of Russo and Bresler (1982); (B) variograms for different areal supports derived from final infiltration data of Viera et al. (1981); (C) variograms for different area supports derived from Philip "A" parameter data of Sharma et al. (1980).

distance between two points. The semivariogram at any larger averaging volume (or support) V is then given by:

$$\gamma_V(h) = \bar{\gamma}(V, V_h) - \bar{\gamma}(V, V) \quad (1)$$

where V_h indicates a support V moved over a distance h and the mean value of the semivariogram is given in two dimensions by:

$$\bar{\gamma}(V_1, V_2) = \int_{x_1}\int_{x_2}\int_{y_1}\int_{y_2} \gamma(\{(x_1 - x_2)^2 + (y_1 - y_2)^2\}^{1/2})\, dy_2\, dy_1\, dx_2\, dx_1 \quad (2)$$

For distances h that are large compared with the range of the point variogram:

$$\gamma_V(h) = \gamma(h) - \bar{\gamma}(V, V)$$
$$= \sigma_V^2 \quad (3)$$

which is a constant equivalent to the expected variance of independent samples of support V. The second right-hand side term of eqn. (3) represents the reduction in expected variance due to the averaging process over V. Some example calculations for hydraulic conductivity data are shown in Fig. 1a for an experimental variogram taken from Russo and Bresler (1982). In this case the measurement scale was 2.2 × 2.2 m, so the experimental variogram was reduced to a point variogram by trial and error before the application of eqn. (2). Other examples have been calculated for data on final infiltration capacity (from Viera et al., 1981; Fig. 1b) and on the A parameter of the Philip infiltration equation (data from Sharma et al., 1980; Fig. 1c). In the latter case the definition of the calculated experimental variogram was poor due to a lack of data, and the fitted curve was deliberately chosen to fit the calculated sample variance and demonstrate the effect of including "nugget variance" at the point scale. Integration over a larger area immediately eliminates this nugget variance.

The figures show that increasing the scale of averaging reduces the expected variance of the distribution of parameter values, in some cases drastically. This suggests that effective parameter values at the grid may yet prove to be acceptable in physically-based modelling, subject to three qualifications. The averaging process in the theory of block kriging is a linear process. It may not lead to the same results as the hydraulic averaging of nonlinear processes that occurs in real catchments. Even if we had data on the point variogram of soil hydraulic characteristics at a site, block kriging may not therefore yield the correct effective parameter values at the grid scale. Secondly, the use of effective parameter values presupposes that the model equations are appropriate at the grid scale. As yet we have no proof that this is the case for heterogeneous systems. Thirdly, this type of spatial interpolation and lumping assumes that we are dealing with a stationary random function. This may not be true for field soil properties and non-stationarity of the mean may affect the parameter values at the grid scale.

The use of effective parameter values at the grid scale in the application of SHE and other physically-based models is an understandable pragmatic response to the problem of calibrating such models. It is, however, an act of faith that is not based on sound physical reasoning. Binley et al. (1988b) have shown that this faith may be justified for purely subsurface flow responses, but not for responses in which heterogeneity of soil properties affect both surface and subsurface flows. They also show that, at the hillslope scale, it may be very difficult to relate an appropriate effective parameter value to the moments of the underlying distribution of parameter values, even where that distribution is stationary in space. This is in part due to the fact that a hillslope may represent only a small sample from the range of soil variability, particularly where there is significant spatial correlation. Increasing the size of the sample to the catchment scale may ease this problem, but then introduces additional heterogeneity due to topography and possible rainfalls.

This problem has been investigated by Wood et al. (1988) using a simpler model, who suggest that the differences in response between different heterogeneous areas of the same scale may become less important at a scale of the order of 1 km^2 (for particular rainfall, soil and topographic fields). They suggest this scale is the "representative elementary area" for predicting catchment response. At this scale it may still be necessary to take heterogeneity into account in making predictions, but it is no longer necessary to consider the *pattern* of heterogeneity.

THE CALIBRATION OF PHYSICALLY-BASED MODELS

At this point it may be worth considering physically-based models in the context of the experience with conceptual models lumped at the catchment scale (points a to e above). It should be clear from the discussions above that physically-based models must suffer from the same kind of problems as lumped conceptual models, if not to the same degree. We have already shown how points a and b apply to physically-based models. They cannot avoid the problems associated with errors in observed data (point c). In respect of point d, physically-based models are wildly over parameterised in a systems simulation sense. Not only are there large numbers of parameters associated with the processes simulated in the models (e.g. table 1 in Abbott et al., 1986b) but these parameters may take on different values in different model grid squares. In the application of the SHE model to the Wye catchment, it was necessary to specify approximately 2400 parameter values (not including topographic variables or changes in parameter values over time). This is not a problem if the parameter values can all be adequately estimated a priori or measured economically in the field. It is a problem if it is necessary to optimise those parameter values in any way by comparison of observed and simulated responses. Which ones do we change?

In the recent SHE papers the authors adopt a pragmatic approach to the identification of parameter values. They assume that some parameters, particularly those to which the model predictions are insensitive, can be estimated a priori. Other parameters are assumed to vary only on the basis of soil or vegetation type. The number of parameters actually supplied to the model is therefore much smaller (of the order of 40). They recognise that although "in principle" these values can be measured in the field, in practice (Abbot et al., 1986a):

"Problems such as inadequate representation of the hydrological processes and the possible difference in scale between the measurement and the model grid scale mean that some calibration is likely to continue to be required. In a SHE context, this is regarded as selective improvement of initial parameter estimates by a comparison between observed and simulated hydrological variables· · · · ·carried out on a trial and error basis".

In the event about four parameter values (as well as four initial condition values) were calibrated in this way (Bathurst, 1986a). This is of the same order as was suggested in point d above as being appropriate to calibration against a discharge record, except that in this case the parameter values are all of the same type (hydraulic conductivity) and only the information available in one storm is being used. The authors place great emphasis on the use of physical reasoning to guide the trial and error calibration, but it is difficult to be convinced. The very fact that parameter values of the same type are being used suggests that there will be some interaction between the effects of the different values (vide point e above). It is surely easier to use physical reasoning to calibrate the residence time parameter of a linear store, than it is to obtain a correct combination of interacting parameters.

The problem is much greater if an attempt is made to selectively improve more than one parameter value for each soil type. Indeed physical reasoning would suggest that there *must* be interaction between parameter values. The authors recognise that for the type of saturation excess runoff generation simulated in the Wye, several parameters can influence the volume and timing of runoff generated. An increase in hydraulic conductivity or slope angle can be offset by a decrease in available storage which will be controlled by the porosity, moisture characteristics and initial conditions.

Parameter interaction is inherent in the physics of hydrological systems and given the number of parameter values available in physically-based models, any optimisation of parameters must be subject to far greater problems of interaction than simpler lumped models. This problem might be offset to some extent by making use of measured internal state variables in the calibration. However, it is difficult to see at present how this might be achieved. How can we compare an average grid square moisture content or water table level predicted by the model with a "point" neutron probe or piezometer measurement? What if we are fortunate enough to have two or more measurement sites within a grid square? We do not currently have anything better than an ad hoc methodology to make use of such data.

Some indication of the problems that will arise, given limited time and resources in calibrating this type of model, is evident from the application of the SHE model to the Wye catchment (Bathurst, 1986a). Discharge data were available for tributaries of the Wye, allowing calibration checks against internal variables. For storm 1, two of the tributaries show significant errors, although these errors effectively cancel each other in the simulation of the whole catchment. Physical reasoning did not appear to throw light on the cause of the errors and "it is difficult to improve one of the simulations without further increasing the error in the other" (Bathurst, 1986a). In essence, the calibrated model is not extracting all the information available from the data in either a physical or a systems simulation sense. I do not mean to imply any particular criticism of the model, or of this particular application. I do feel, however, that it is indicative of the type of problem that will arise in the practical calibration of physically-based models *even by expert users*. Despite the underlying theory and the possibility of obtaining first estimates of parameter values a priori, they are far more difficult to calibrate than simpler models, and physical reasoning can do little to mitigate that.

Much will therefore depend on the possibility of specifying parameter values either by measurement or by a priori estimation. Bathurst (1986b) suggests that measurements at a "few representative sites" may be sufficient to obtain an initial calibration of the model. It is not suggested how a representative site should be chosen, nor what measurement techniques might be appropriate to obtain the required effective grid-scale parameters. He has taken the estimates of soil parameters from the work of Knapp (1970). Knapp used two different methods to estimate hydraulic conductivity. The first was a laboratory method based on small samples which yielded a range for saturated O/A horizon samples of 0.241 to 0.396 m h^{-1}. The second was based on outflows from 1 m wide throughflow troughs at different points on a hillslope, which yielded a range of 14.12–0.01 m h^{-1}. There must therefore be a high degree of uncertainty in estimating the grid square values from this data.

Clearly, in most applications the modeller will not have the benefit of such field measurements. His choice is then to go out and measure parameter values directly (which is expensive), or make a priori estimates of the values required from other physical information about the catchment. To illustrate what might happen, let us again consider the hydraulic properties of soils. In some applications we may have information about soil textures. It is known that soil hydraulic characteristics vary with soil texture. Recent work in the U.S. has attempted to quantify these relationships on the basis of measurements on some 5000 soil samples collected by the U.S.D.A (see e.g. Brakensiek et al., 1981; Cosby et al., 1984). It is consequently very tempting to use these relationships to estimate saturated hydraulic conductivity and other soil parameters.

There are a number of dangers in this. There is considerable uncertainty associated with these estimates. The variance associated with the estimates of hydraulic conductivity in each textural class is not small relative to the mean. Data in Cosby et al. (1984) show a range of variances from 0.33 to 0.69 log units. Even more important is the nature of the original measurements on which these estimates are based. At least most of the hydraulic conductivity measurements were made on "fist-sized fragments" of the soil taken back to the laboratory (Holtan et al., 1968). The measurements are therefore estimates of matrix conductivities and take little account of the effects of structural and biotic voids which must surely be important in determining the grid scale parameters. In addition, such estimates take no account of the spatial correlation that has been demonstrated in detailed field measurements of soil hydraulic parameters (e.g. Russo and Bresler, 1982) and that must influence the pattern of variability to be expected at the grid scale. It must consequently be concluded that such a priori parameter estimates, although physically-based, might give very misleading results.

It has also been suggested that future developments in remote sensing might also provide a means of obtaining "average parameter values on a grid basis"

(Abbott et al., 1986a). Such optimism again needs qualifying. We will have great difficulty in relating the imaging sensor average at the pixel resolution scale, to the required hydrologically effective parameter (or state variable) values at the model grid scale (if such effective values indeed exist). This is not a way around the lumping of subgrid processes problem, since in order to develop such a relationship we will still need a theory of spatial averaging.

THE ACCEPTABILITY OF PHYSICALLY-BASED MODELS

Many modelling studies demonstrate the acceptability of the calibrated model structure by applying the model to a number of further storms. In the case of physically-based models, this may require the calibration of the initial conditions for each storm (e.g., Rogers et al., 1985; Bathurst 1986a). This is not a proper validation. Stephenson and Freeze (1974) note that the calibration–validation sequence ideally requires perfect knowledge of the initial and boundary conditions. If these must be calibrated for a validation event "the resulting flexibility almost ensures that a satisfactory validation will be obtained" (Stephenson and Freeze, 1974, p. 289). This suggests that it may be very difficult to determine the acceptibility of a physically-based model without long period simulations in which the effect of the initial conditions is negligible, even given good boundary conditions and input data. This is not currently computationally practicable.

There are, however, more fundamental problems concerning the acceptibility of physically-based models. In an earlier assessment of distributed modelling (Beven, 1985) the danger of a subversion of hydrological reasoning by ill-conceived applications of physically-based models was discussed. We now have an example from the research community to support this opinion. In the paper of Loague and Freeze (1985) a "quasi-physically-based" model based on predictions of Hortonian overland flow with kinematic routing is applied to a large number of storms on several different catchments. Model predictions were compared for both calibration storms and validation storms. They conclude that the physically-based model has little predictive value. That may be true, but one of the catchments to which the model is applied is one of the Hubbard Brook experimental catchments which has a tabulated soil hydraulic conductivity of 559 mm h^{-1}. Freeze (1972) himself pointed out the low probability of obtaining overland flow in the Northeast U.S. Loague and Freeze (1985) recognise the incongruence of the model and the data in this case but it is surely an indictment of hydrological practice to even consider applying such a model to such a catchment.

A more subtle example concerns the application of different physically-based models to the same catchment. In studies using the IHDM on the Wye and Severn catchments, storm responses have been simulated by an infiltration excess mechanism (Morris, 1980; Rogers et al., 1985). The argument is that infiltration capacities are exceeded at the base of the peat top soil (see Bathurst, 1986a, for a physical description of the Wye catchment), and that the surface flow component (which is based on equations describing sheet flow) subsumes both rilled and unrilled surface flows and flows through natural pipes in the peat. The SHE model makes a similar assumption in representing surface and pipe flows as an "equivalent" sheet flow, but these flows are produced in the model as a result of saturation of the peat layer. Both models are "physically-based", both mimic part of what is thought to happen in reality, both make gross assumptions about the nature of the "surface flow" component. Is one model more physically-based than the other, or are both equally incorrect? On what rational basis can we decide on the adequacy of one model vis-a-vis another, especially when we have little detailed knowledge of the flow processes occurring in the field, as will be true in so many applications? Consequently can we use such physically-based models to infer the processes operating in a particular environment? I believe that any application in which physically correct predictions are considered important must involve a close cooperation between field observation and modelling. The nature of this cooperation, and in particular the nature of appropriate field observation and parameter measurement techniques, requires considerable further research if physically-based models are to realise the advantages that are currently being claimed for them (e.g. Abbott et al., 1986a).

MULTISCENARIO MODELLING AND PREDICTIVE UNCERTAINTY

The discussion above has demonstrated the many limitations of the current generation of physically-based models. Problems associated with errors in model structure, estimation of parameter values and the specification of initial and boundary conditions must result in a significant degree of uncertainty associated with the model predictions. Yet there are some hydrological problems demanding predictions which at present can only be provided by physically-based models (see Beven and O'Connell, 1982; Abbott et al., 1986a). Predicting the effects of land use changes is a particular example that has often been used to justify the effort of developing physically-based models. No calibration of parameter values by comparison with observed data is possible in such circumstances, so that the model predictions must rely on a priori estimation of parameter values. Abbott et al. (1986a) rightly point out that lack of such data does not prevent planning decisions from being made and argue that: "the value of SHE in such a case resides precisely in... its ability to examine the uncertainty attached to a predicted outcome as it is affected by uncertainty in the database". These authors suggest that this will be achieved through multiscenario modelling, in which a number of different models (differing in estimated parameter values or boundary conditions) are prepared, and their predictions compared to the historical record. Simulations that satisfactorily reproduce the historical behaviour may then be used to evaluate the outcomes of future changes. They suggest that usually no more than four scenarios are prepared (presumably all historically acceptable).

It would appear that some clarification is in order here. Abbott et al. (1986a) do not make clear the difference between evaluation of different scenarios, and evaluation of the predictive uncertainty associated with every simulation. The authors appear to be suggesting that these scenarios be prepared in a deterministic simulation mode. This is a very limited evaluation of uncertainty in outcomes and allows no evaluation of the probability associated with each possible outcome. The discussion above would suggest that there may be a great many combinations of parameter values and boundary conditions that might be consistent with historical behaviour, some more or less probable than others. Equally, predictions into the future should also be probabilistic in that we can only estimate changes in parameter values with a significant degree of uncertainty.

To use an example from their paper, we can certainly play "what-if games" to assess the effects of different levels or patterns of logging in a catchment, but our predictions would depend not only on the pattern of removal of the forest cover, but also on our estimates of the effect of the logging operation on the hydraulic conductivity of the topsoil. Because of the uncertainty in such parameters (and initial and boundary conditions) it is important to recognise that every scenario will be associated with a probabilistic range of possible behaviours. It is an interesting question as to whether this uncertainty, if incorporated into the simulations, would allow us to distinguish statistically between the different logging scenarios, especially given that every simulation is ultimately constrained by the requirements of mass balance.

There are no ideal methods for assessing such predictive uncertainty (and no methods at all for assessing the uncertainty resulting from errors in model structure). The method of Rosenblueth used by Rogers et al. (1985) has the advantage of requiring only a small number of simulations if the number of uncertain parameters is small. However, it is essentially an approximate linear method applied to a nonlinear problem and as such can only give approximate estimates. For a larger number of parameter values, Monte Carlo methods based on conditional simulations (e.g., Delhomme, 1979; Clifton and Neuman, 1982) may become competitive, but as yet remain very expensive in computer time.

TOWARDS A FUTURE FOR PHYSICALLY-BASED MODELS

This comment has attempted to take a realistic look at the problems involved in physically-based modelling. While specific reference has been made to the SHE and IHDM models, the comments apply equally to all current physically-based models. These models are founded in small-scale laboratory physics and are best suited to the essentially research task of exploring the implications of making specific sets of assumptions about the operation of hydrological systems. They are not well suited to applications to real catchments. They do not overcome the various disadvantages of the "previous generation" of lumped conceptual models in this respect. They are themselves lumped conceptual models, in the sense of relying heavily on "effective" grid scale values of parameters and variables, and are subject to the same disadvantages. The ability to use physical reasoning in model calibration is not unique to physically-based models and does little to overcome the difficulties of calibrating such models. Physically-based a priori estimation of grid scale effective parameter values may be subject to considerable uncertainty.

There are problems demanding physically-based predictions, and there will continue to be a need for physically-based models. Future developments in physically-based modelling must take account of the need for a theory of the lumping of subgrid scale processes; for closer correspondence in scale between model predictions and measurements; for closer correspondence between model equations and field processes; and for the rigorous assessment of uncertainty in model predictions. In the worst cases, explicit evaluation of such uncertainty might reveal that the predicted effects of different strategies cannot be distinguished statistically.

The application of current physically-based models might best be carried out in conjunction with a programme of field measurements, to ensure consistency between model predictions and real world processes. In the prediction of the effects of future changes, a measurement programme should be initiated and maintained to allow continual reassessment of model predictions. Only in this way will physical hydrology progress, rather than being undermined by the misplaced application of inadequate models, albeit "physically-based". Both field measurements and the application of current physically-based models are demanding of staff time and consequently expensive. It needs to be shown that such effort can be justified on a cost–benefit basis before these models are widely applied beyond the research applications for which they are best suited.

ACKNOWLEDGEMENTS

The ideas expressed in this paper have evolved over a long period of time during which many discussions with colleagues involved on the SHE and IHDM projects at the Institute of Hydrology have been an important influence on my own changing ideas. Support from NERC grant GR3/6264 is also gratefully acknowledged.

REFERENCES

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J., 1986a. An introduction to the European Hydrological System — Système Hydrologique Européen, "SHE", 1. History and philosophy of a physically-based, distributed modelling system. J. Hydrol., 87: 45–59.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J., 1986b. An introduction to the European System — Système Hydrologique Européen, "SHE", 2. Structure of a physically-based, distributed modelling system. J. Hydrol., 87: 61–77.

Bakr, A.A., Gelhar, L.W., Gutjahr, A.L. and Macmillan, J.R., 1978. Stochastic analysis of spatial variability in subsurface flows, 1. Comparison of one and three dimensional flows. Water Resour. Res., 14: 263–271.

Bathurst, J.C., 1986a. Physically-based distributed modelling of an upland catchment using the Système Hydrologique Européen. J. Hydrol., 87: 79–102.

Bathurst, J.C., 1986b. Sensitivity analysis of the Système Hydrologique Européen for an upland catchment. J. Hydrol., 87: 103–123.

Beven, K.J., 1977. Hillslope hydrographs by the finite element method. Earth Surf. Processes, 2: 13–28.

Beven, K.J., 1983, Surface water hydrology — runoff generation and basin structure. Rev. Geophys. Space Phys., 21: 721–730.

Beven, K.J., 1985. Distributed models. In: M.G. Anderson and T.P. Burt, (Editors), Hydrological Forecasting. Wiley, New York, N.Y., pp. 405–435.

Beven, K.J., 1987. Towards a new paradigm in hydrology. IASH Publ. No. 164: 393–403.

Beven, K.J. and O'Connell, P.E., 1982. On the role of physically-based models in hydrology. Inst. Hydrol. Rep. No. 81, Wallingford.

Beven, K.J., Calver, A. and Morris, E.M., 1987. The Institute of Hydrology Distributed Model. Inst. Hydrol. Rep. No. 98, Wallingford.

Binley, A.M., Elgy, J. and Beven, K.J., 1988a. A physically-based model of heterogeneous hillslopes. I. Runoff production. Water Resour. Res. (in press).

Binley, A.M., Beven, K.J. and Elgy, J., 1988b. A physically-based model of heterogeneous hillslopes. II. Effective hydraulic conductivities. Water Resour. Res. (in press).

Brakensiek, D.L., Engleman, R.L. and Rawls, W.J., 1981. Variation within texture classes of soil water parameters. Trans. Am. Soc. Agric. Eng., 24: 335–339.

Clifton, P.M. and Neuman, S.P., 1982. Effects of kriging and inverse modelling on conditional simulation of the Avra Valley aquifer in Southern Arizona. Water Resour. Res., 18: 1215–1234.

Cosby, B.J., Hornberger, G.M., Clapp, R.B. and Gunn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res., 20: 682–690.

Cushman, J.H., 1986. On measurement, scale and scaling. Water Resour. Res., 22: 129–134.

Delhomme, J.P., 1979. Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. Water Resour. Res., 15: 269–280.

Freeze, R.A., 1972. Role of subsurface flow in generating surface runoff. 2. Upstream source areas. Water Resour. Res., 8: 1272–1283.

Freeze, R.A., 1980. A stochastic-conceptual analysis of rainfall–runoff processes on a hillslope. Water Resour. Res., 16: 391–408.

Holtan, H.N., England, C.B., Lawless, G.P. and Schumaker, G.A., 1968. Moisture-tension data for selected soils on experimental watersheds. Rep. No. ARS 41–144, Agric. Res. Serv., Beltsville, Md., 609 pp.

Hornberger, G.M., Beven, K.J., Cosby, B.J. and Sappington, D.E., 1985. Shenandoah Watershed Study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. Water Resour. Res., 21: 1841–1850.

Ibbitt, R.P., 1972. Effects of random data errors on the parameter values for a conceptual model. Water Resour. Res., 8: 70–78.

Ibbitt, R.P. and O'Donnell, T., 1971. Designing conceptual catchment models for automatic fitting methods. IASH Publ., 101: 461–475.

Journel, A.G. and Huijbregts, Ch. J., 1978. Mining Geostatistics. Academic Press, New York, N.Y., 600 pp.

Knapp, B.J., 1970. Patterns of water movement on a steep upland hillside, Plynlimon, Central Wales. Unpubl. Ph.D. Thesis, University of Reading.

Loague, K.M. and Freeze, R.A., 1985. A comparison of rainfall–runoff modelling techniques on small upland catchments. Water Resour. Res., 21: 229–240.

Luxmoore, R.J., Grizzard, T. and Pastterson, M.R., 1981. Hydraulic properties of Fullerton Cherty Silt Loam. Soil Sci. Soc. Am., J., 45: 692–698.

Morris, E.M., 1980. Forecasting flood flows in grassy and forested basins using a deterministic distributed mathematical model. In: Hydrological Forecasting. IAHS Publ., 129: 247–275.

Philip, J.R., 1980. Field heterogeneity: some basic issues. Water Resour. Res., 16: 443–448.

Pickup, G., 1977. Testing the efficiencies of algorithms and strategies for automatic calibration of rainfall–runoff models. Hydrol. Sci. Bull., 22: 257–274.

Rogers, C.C.M., Beven, K.J., Morris, E.M. and Anderson, M.G., 1985. Sensitivity analysis, calibration and predictive uncertainty of the Institute of Hydrology Distributed Model. J. Hydrol., 81: 179–187.
Russo, D. and Bresler, E., 1982. Soil hydraulic properties as stochastic processes: II Errors of estimates in a heterogeneous field. Soil Sci. Soc. Am., J., 46: 20–26.
Sharma, M.L., Gander, G.A. and Hunt, C.G., 1980. Spatial variability of infiltration in a watershed. J. Hydrol., 45: 101–122.
Sharma, M.L. and Luxmoore, R.J., 1979. Soil spatial variability and its consequences on simulated water balance. Water Resour. Res., 15: 1567–1573.
Sharma, M.L., Luxmoore, R.J., DeAngelis, R., Ward, R.C. and Yeh, G.T., 1987. Surface water flow simulated for hillslopes with spatially dependent soil hydraulic characteristics. Water Resour. Res., 28: 1523–1530.
Smith, R.E. and Hebbert, R.H.B., 1979. A Monte-Carlo analysis of the hydrologic effects of spatial variability of infiltration. Water Resour. Res., 15: 419–429.
Sorooshian, S. and Gupta, V.K., 1983. Automatic calibration of conceptual rainfall–runoff models: the question of parameter observability and uniqueness. Water Resour. Res., 19: 260–268.
Stephenson, G.R. and Freeze, R.A., 1974. Mathematical simulation of subsurface flow contributions to snowmelt runoff, Reynolds Creek Watershed, Idaho. Water Resour. Res., 10: 284–298.
Viera, S.R., Nielsen, D.R. and Biggar, J.W., 1981. Spatial variability of field-measured infiltration rates. Soil Sci. Soc. Am., J., 45: 1040–1048.
Wood, E.F., Sivapalan, M., Beven, K.J. and Band, L., 1988. Effects of spatial variability and scale with implications to hydrologic modelling. J. Hydrol., 102: 29–47.
Yeh, T.-C., Gelhar, L.W. and Gutjahr, A.L., 1985. Stochastic analysis of unsaturated flow in heterogeneous soils, 3 papers. Water Resour. Res., 21: 447–471.