4 Calibration, Validation and Equifinality in Hydrological Modelling: A Continuing Discussion . . .

KEITH BEVEN
Institute of Environmental and Natural Sciences, Lancaster University, UK

This lunchtime discussion takes place in a university bar following the morning session at a workshop on Validation of Environmental Models.[1]

Taking part in the discussion are a Professor of Hydrology (P), a graduate student (G) who has recently started a research project with P, a visiting scientist from Brazil (B), a consulting engineer with particular interests in the remediation of contaminated land (C), a scientist from the Environment Agency (E) with responsibilities for the licensing of effluent discharges, and a scientist from the UK Government Department of the Environment (D) with interests in the impact of climate change on water resources.

With the exception of the graduate student and the scientist from Brazil, all the others are old friends, having studied together as undergraduates.

D So what did you think of the presentations this morning?

E Well, I got completely lost during those philosophical presentations at the beginning – they really didn't seem to bear much relation to what happens in practice.

D How do you mean?

E With the amount of data that we have for most model applications the problem is not so much whether we can verify or validate or confirm or evaluate a model but rather whether we can gather enough decent data from different sources to make any predictions at all. And we have so many sites to check that we cannot spend too much time on each one.

C We have the same problem in knowing how to characterise a site, especially since I am dealing with what might be happening under the ground. The important thing is to make a prediction quickly, not to worry about validation.

D Does the data available generally affect what type of model you choose to use?

C No, not often – that is usually a question more of what variables are required to be predicted. That might vary for different projects.

P So how do you choose a model to use?

C Well obviously we have a degree of experience in using different types of model, some of which are well established as tools and some of which we have developed or adapted

[1] The audience had heard papers similar to Morton (1993), Oreskes et al. (1994), Schrader-Frechette (1989), Beck (1994), and lastly Beven (2001a).

Model Validation: Perspectives in Hydrological Science. Edited by M.G. Anderson and P.D. Bates.

ourselves – in fact it is important that we should be able to list this experience to the client in tendering for a project.

P But that sort of implies a process of prior evaluation and modification of models over different projects. Your reputation depends on using a model that should give results that are in some way reasonable so that if you were aware that a model does not perform for a given purpose you would have dropped it or modified it. Does that not count as validation or verification?

C It would appear not according to the philosophers. It may be at least a *confirmation*, since at least we certainly do some evaluation of performance in each project but we are not really in the business of showing that a model is a true representation of reality, only that it gives predictions that are practically useful.

G But surely we have to believe that it is possible to validate a model as a representation of reality, otherwise we don't have a science of hydrology.

D I am not sure I agree with you there. There are plenty of examples from the history of science that have shown how scientists have modified models to fit new observations so as to have improved predictive capabilities only to see those theories superseded as a 'true' representation of reality. Indeed there are examples where the new theory is not, initially at least, as good a predictor of the data but appears to have other attractions or potential, if only to act as a competitor to the existing theory.[1] It seems quite possible to go on using and modifying a model or theory, within the current paradigm so to speak, as scientists with the best of intentions but without having to make a strong statement that this model is a true representation of reality. Look at how widely the Horton infiltration excess concept has continued to be used in hydrological prediction, even where overland flow is not the dominant runoff mechanism, by adjusting the infiltration losses so that the effective rainfall has a volume equal to the 'storm runoff'.

P Is the problem in hydrology not more that it is all too easy to recognise the faults or limitations of models, especially of subsurface flow processes, so that if we were really scientists we should be rejecting all the models?

E That wouldn't get us very far – I have to make predictions about the impact of a new effluent on river water quality every day. The models may not be totally realistic in terms of representing all the details of all the processes but I would argue that they can have useful predictive capabilities, so they must be reasonably realistic in that sense.

P But how do you know that – how do you check or validate those predictions?

E Well, very often of course we don't have time to do any checking – we would normally only go back for more detailed studies at a site if a problem occurs, and problems will often be the result of extreme low flow conditions where the model might be less reliable because it is outside its calibrated range.

B That suggests, however, that if a model is only as good as the range of data over which it is calibrated it is not necessarily realistic in terms of reproducing the processes involved.

E But remember there is a historical context here. Today's models are much better than 10 years ago and we can hope that there will be more improvements in the future.

B Where do you think that the improvements will come from? New theory? New measurement techniques? Or just more computer power?

E I think new measurement techniques will be very important – we are only just starting to make use of remote sensing data and continuous water quality measurements. These should eventually lead to better understanding and new theoretical developments.

P But I cannot see how this will lead to improved confirmation or validation of models in

[1] See the discussions in, for example, Chalmers (1990) and Feyerabend (1978), or, for pure entertainment, Feyerabend (1991).

general practical applications. Think of a real catchment, with its patchwork of different hydrological conditions on the hillslopes. I was out in the field with some students last week, trying to help them relate what they could observe with the textbook diagrams they have seen about hydrological processes. To do so you have to filter out all the complexity. That is perhaps useful as a teaching tool but, if you dig a soil pit in wet conditions or do a surface or subsurface tracing experiment, you realise how the complexity dominates the nature of the flow processes. The water flow is usually dominated by very discrete preferential flow pathways.[1] Emmett showed the same thing for surface runoff back in the 1970s,[2] while tracer experiments in rivers show how complex the flow structures, mixing processes and exchanges within the hyporheic zone are.[3] In the field, it is only too obvious that the hydrological theory of the textbooks has its limitations. Even Darcy's law has been shown to fail a (stringent) validation test when applied to Darcy's original data set.[4]

G But what is wrong with Darcy's law? Surely that is fundamental to hydrological theory? It is used in all the physically based models.

D I thought the arguments were well rehearsed in that talk this morning about the sociological context of theory development.

C Oh that! I went for coffee as soon as she mentioned Derrida and the signification of signs.

D Well it wasn't easy to follow, for sure, but I think a crude interpretation of what she was trying to get over was that the post-modernists regard a theory as a social construct or sign that, over time, many people agree to recognise. As such, they are then reluctant to give up that theory, despite the fact that it is difficult to demonstrate its adequacy.

G But how do we throw away Darcy's law? Surely it has been used successfully and besides what is there to put in its place?

D Quite – that was the point (and a whole lot more about adequacy and misapplication of theory at different scales). There isn't really anything to put in its place at its correct scale of application (which would appear to be very small), and at larger scales we can get away with calibrating the parameters. Remember she posed the question as to whether a hydraulic conductivity calibrated by some inverse method could be a 'real' hydraulic conductivity.

G How could you show it was?

C Go and take measurements.

D But wasn't that the point she was trying to make? The calibrated value is a large-scale effective value that also depends, through model calibration, on the model structure, the surrounding values and responses over the calibration period. The measured value is some local value in both space and time that depends on the measurement technique used. The two may not be the same quantity.[5]

C But surely the very process of taking the measurement adds new information that can be used in calibration – using kriging of the measured values for example, or as prior information in a Bayesian inverse method.[6]

D But there is a danger of treating that measured value as true when it may not be at the right scale for the effective values required by the model. There might not be a simple functional link between them.

P In fact the problem may be worse than that. I remember a paper by one of David Lerner's

---

[1] See, for example, Flury et al. (1994), Binley et al. (1996), Henderson et al. (1996).
[2] See Emmett (1978).
[3] See Harvey et al. (1996).
[4] See Davis et al. (1992).
[5] See also discussion in Beven (1996b).
[6] See, for example, McLaughlin and Townley (1996).

students in WRR a while back[1] where they tried all sorts of different optimisations and came up with completely different best fit parameter sets.

B Perhaps they just hadn't found the global optimum. We have seen that when trying to do groundwater calibrations in the past, but recently we have had a lot of success with a simulated annealing method.[2] It seems to be much more effective in converging to a consistent optimum parameter set. It does take a lot of runs but there is not such a problem with computer power these days.

G That also depends on what you are optimising. There is a recent paper in WRR using multi-criterion optimisation that tries to take account of the fact that different parameter sets perform best on different criteria. I think they came up with what they called the Pareto optimal set of models.[3]

E Is this not getting away from the point? Ok, so we have used optimisation as a way of fitting models to data in the past, but isn't that just like using a regression analysis? The parameters are only going to compensate for errors in the model, boundary conditions and observations. It does not have much to do with whether the model is realistic or not, just whether it has the right sort of functionality to reproduce the data. In regression analysis, it is generally easy to see if you have the right sort of functionality. If not, you try a different sort of transform. In hydrological modelling it is not nearly so easy to reject a model if you have calibrated the parameters. There are usually enough degrees of freedom to fit the data reasonably well.

C No, surely it is not so different. If a model produces accurate predictions in some sense is that not sufficiently realistic or valid or, at the very least, fit for purpose. If it does not produce sufficiently accurate predictions then you should change the model. It is allright for you academics to argue about these things but I sometimes have to defend fitness for purpose of the model I have chosen at a public inquiry or in court.

P And what argument do you use?

C Past experience of successful predictions together with any validation tests if there are data available to do so.[4]

E But if success in prediction is the only criterion, surely we have a problem – we could have as many calibrated models as there are hydrologists (it was almost like that in the days of conceptual rainfall–runoff models). If we are scientists should we not take a viewpoint that we can at least falsify some of the possible models?

G Falsification is not the problem if you are thinking in terms of process representation. Unfortunately it is all too easy to falsify all our models in applications to real sites, so I do not think that hydrologists can afford to take a critical rationalist position here.[5] I learned a bit about Popper's ideas on falsification in my MSc course last year but I did not see how it related to the type of modelling I do. As Morton said this morning,[6] I *know* that the available process theories are not true descriptions of a catchment but I do not reject them (or I would be left with nothing to predict with) but try to improve them. That is why I am doing a PhD project.

[1] See Brooks et al. (1994).
[2] See Sen and Stoffa (1995).
[3] See Gupta et al. (1998).
[4] See Bair (1994).
[5] For a simple introduction to Popper, critical rationalist and modern realist philosophies, see Chalmers (1990). For a Bayesian perspective on falsification see Howson and Urbach (1993). For a discussion with specific reference to geomorphological explanation, see Richards (1990, 1994), Bassett (1994), Rhoads (1994), Beven (1996a).
[6] See Morton (1993).

D How far are improvements possible, though? Do we not need a radically different, more realistic theory in hydrology that will explain rather than simply predict. That would give me much more faith in the predictions that are being made about the hydrological impact of a changed climate in Britain for example. At the moment, since calibration with future conditions is not possible, I am very dubious about whether the models that are being used will produce adequate predictions.

P That is an interesting example. In fact, we generally have quite a good *qualitative* theory of what is going on in a catchment in our heads. Even if we cannot be sure about all the details of the spatial and temporal variability in the flow pathways, especially through the deeper subsurface pathways, we can at least be aware of the possibilities. In that sense we do, already, have a good theory of hydrology. The difficulty is putting that qualitative theory into a quantitative mathematical description.

E But will such a quantitative theory ever be possible? When it comes down to real hydrological applications, the problem is not one of generalisations, it is about the hydrology of specific places where we will inevitably have imperfect knowledge of the inputs, outputs and processes over a range of hydrologically relevant scales. Every place and every application is unique in that sense.

P That is true and, in part, what I meant about being unsure of the details of the flow pathways. Listening to the discussion this morning, I was thinking how does this relate to *my* catchment. I am not only interested in the theory, but in its description of my unique catchment. I do not see how we can get away from calibration in the end.

B And think of our example of the Parana catchment in Brazil, 1.5 million km^2. But we still need to make predictions for water resource management and to evaluate the impacts of land use and climate change. And we are not even sure of the pattern of rainfall inputs under current conditions. My own feeling is that it may not be necessary to have a completely accurate description of the processes to have a model that is useful in prediction, as long there are some data available to calibrate the particular characteristics of a catchment.

P But if we are forced always to calibrate in some way there is an implication that there will always be the possibility of multiple descriptions that will do equally well in calibration. After all, the recent trend is for models to include more processes which means more parameters which means more degrees of freedom in calibration.

E That was what I understood this morning from Beck's use of *non-identifiability* and Beven's *equifinality*.[1] Did they mean the same thing?

P I am not sure that they did. I think that non-identifiability was intended to be in respect of some 'true' description of the system, whereas equifinality accepted that we may *only* be able to find a variety of acceptable descriptions given the data available so that we are unlikely ever to be able to say that we have the true description.

B What do you mean by description? Getting the qualitative description of the dominant *processes* right, or getting accurate predictions of the volume and timing of runoff and evapotranspiration *fluxes*. It is not absolutely necessary to do the first in order to achieve the second.

G But that is to accept the traditional split between engineering hydrology and hydrological science. Surely we not only want the get the flux predictions right, but also the process descriptions.

E But I cannot see that this will be possible without a dramatic change in measurement techniques that would enable us to determine fluxes or parameters more accurately. Until

[1] See Beck (1987), Beck et al. (1990), Beven (1993, 1996a).

then we may have to accept that the equifinality problem is endemic to environmental models.

C I was not happy about this equifinality idea. Surely such a relativistic viewpoint is untenable for a scientist and there are examples from the environmental sciences that suggest that real advances can be made. Look at how numerical weather prediction using Global Circulation Models has improved in recent years. They don't worry about equifinality.

D Not in terms of parameter values no, but in weather forecasting they do now make 'ensemble' runs with different initial conditions to test the sensitivity of the predictions. And there are some very conceptual sub-grid parameterisations in those models – look at how they represent the land surface hydrology, for example – it would not be considered acceptably scientific to a hydrologist doing catchment modelling, since they are mostly one-dimensional vertical representations averaged over the GCM grid square that cannot reproduce discharges correctly.[1]

P And when Global Circulation Models are used for climate prediction into the next century? These are nonlinear dynamic models and the predictions should be expected to diverge using different initial conditions and parameterisations if they were not constrained by the boundary conditions of the seasonal energy budget. The idea of different nonlinear trajectories leading to much the same answer only reinforces the possibility of having different acceptable or feasible model structures or parameter sets.

C But they do allow that there is some degree of uncertainty in the results produced by different models.

D Yes, and surely this is a situation crying out for a critical rationalist approach. We can surely falsify some of these descriptions by some well chosen experiments in the classical scientific way – that is set up some hypotheses that will distinguish between different representations and test them. That way we are sure to move towards some more realistic models and theories as Popper suggested – and thereby reduce the resulting uncertainty.

E That may be true up to a point but how far is it possible with current measurement techniques to make such tests and what if, when different models or parameter sets are evaluated in this way, there are several or many that pass the test of acceptability. Rainfall–runoff models are a classic example of this, if they are only evaluated on the basis of their success in predicting discharges.

P You can set the test up within a Bayesian framework, that is you start with some prior beliefs and after testing have some posterior distribution of models or parameter sets. The test can then even be subjective.

C But isn't that only a justification for the same relativistic idea posing as science – either a model passes the test or not, either it is acceptable or not, either it is realistic or not.

P It is not always so easy, surely. Take all those odd dotty plots in Beven's talk this morning of goodness of fit against parameter values.[2] How does anyone decide where the level of acceptability is? Is it possible that the one that appears best is not necessarily the most 'realistic' because of the effects of errors in the data against which the model is being tested – or that none of them are 'realistic' despite fitting the data quite well.

C That was where I started to get confused – especially when he started to talk about transforming a landscape space into model space. I didn't see what that had to do with practical prediction.

---

[1] See Lohmann et al. (1998).

[2] Figure 4.1, see also the applications of the GLUE methodology in Beven (1993), Freer et al. (1996), Romanowicz et al. (1996), Franks and Beven (1997a), Zak et al. (1997), Franks et al. (1998), Romanowicz and Beven (1998), Beven (2001b).

B I think I saw what he was getting at and its application to the type of large scale problem I have to deal with. What I understood of the idea was this: let us accept for the moment that there may be many models that have similar function to what is happening in the landscape in the sense that they can be shown to predict similar fluxes to those measured. The important thing in prediction is that the functioning should be represented correctly. So the modeller would be trying to create a map of a landscape or catchment in the model space in a way that properly reflects the uncertainties associated with predicting those fluxes at a point. You can probably envisage that most easily if you think of the model space as represented by a number of parameter axes, with each patch of the landscape being represented by a rather fuzzy region in the parameter space represented by parameter sets that, to the best of our knowledge, have the same functioning as the real patch. The uncertainty could be represented by weights or fuzzy measures associated with the mapping of individual patches to different parameter sets, but the predictions, of course, are done with the different parameter sets. The transformation is a sort of conditioning of the parameter space to represent the catchment.[1]

E But this takes us back to how important it is to have a correct model. How far does this transformation depend on having a correct representation of the processes? A simple model could provide adequate predictions of the fluxes required after such a conditioning under one set of circumstances but be completely wrong under a different set of circumstances because it doesn't correctly represent the processes.

P That can be true of an optimised model as well, of course, or a model with measured parameter values as Keith Loague has demonstrated at Chickasha.[2] But if we have evidence that a model is invalid, then rejection of that model should be part of the process. It would represent an inappropriate mapping (if we can ever bear to reject the model that we have developed, of course). Remember we are discussing conditioning using the available data and certainly not ruling out refinement of that conditioning by carrying out experiments and collecting more data. In fact, that is one of the things I like about the idea – it focuses attention on the value of collecting data and not on theory alone. There does seem to be a tendency to think that the more physically based a model is the more accurate will be its predictions, and to ignore the fact that all models should depend on some field data to define the characteristics of a particular unique catchment or site.[3] I will accept that a process based model may be richer scientifically than, say, a transfer function model, but that does not mean it will be more accurate in reproducing the data with which it can be compared. Process based models may well have additional value in exploring the implications of different assumptions and representations of the physics and chemistry, but they also must ultimately face the same tests of being consistent with the available data, or else they should not be considered useful (except in demonstrating what is *not* a realistic description). The difference, perhaps, is that many of their predictions may well not be testable with the data currently available.

C I am sorry to belabour the point but where does all this conditioning and relativism leave the idea of validation? What use is more data if the model is based on the wrong theory?

P But will you not agree that any useful theory must be consistent with the data available, and that if it is not a critical rationalist like yourself would reject it.

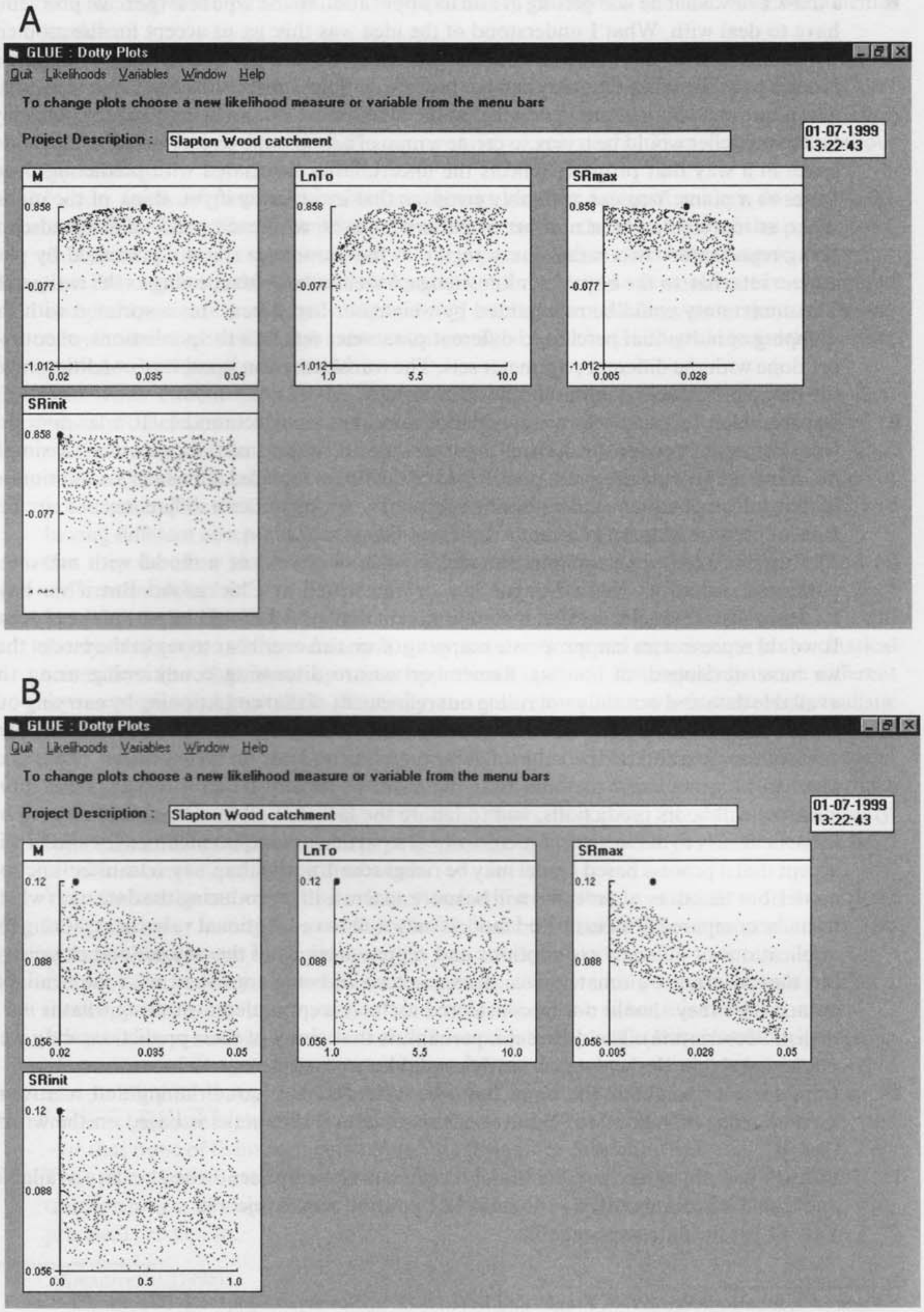
C Yes, if I felt the data were reliable.

---

1 Figure 4.2, see also Beven (1995), Franks and Beven (1997b), Beven (2000).

2 See Loague and Kyriakidis (1997) and earlier studies of the R-5 catchment referenced therein.

3 See discussion in Beven (2000).


A
GLUE : Dotty Plots
Quit Likelihoods Variables Window Help
To change plots choose a new likelihood measure or variable from the menu bars
Project Description : Slapton Wood catchment
01-07-1999
13:22:43
M
LnTo
SRmax
SRinit
0.858
-0.077
-1.012
0.02 0.035 0.05
1.0 5.5 10.0
0.005 0.028
B
GLUE : Dotty Plots
Quit Likelihoods Variables Window Help
To change plots choose a new likelihood measure or variable from the menu bars
Project Description : Slapton Wood catchment
01-07-1999
13:22:43
M
LnTo
SRmax
SRinit
0.12
0.088
0.056
0.02 0.035 0.05
1.0 5.5 10.0
0.005 0.028 0.05
0.0 0.5 1.0

P And that if you have two models or theories that are both consistent with the data then you have no reason to reject either?

C Well, . . .

P And if there were many such models?

C Then surely you have either not yet collected the right data or have not defined consistency properly.

D But, as far as I can see, refining the assessment of the different models is not inconsistent with these ideas – whether that be by collecting more data or by changing your definition of acceptability but it would seem that you would need to have a good reason for redefining acceptability.

E I am still trying to think this through. Even if I accept these ideas, what am I supposed to do if I have to make a decision about whether to issue an abstraction or effluent discharge licence. If I understand correctly I should not just be running one model but a whole suite of models and different parameter sets within each model and then evaluating them all in terms of how well they predict what is happening at the site I may be investigating. I don't have the time to do that – I have to make decisions quickly and very often without any information available to evaluate the model predictions. I am also not sure that I want to have to worry about the uncertainty implied by all these different predictions in making a decision.

P But that doesn't change the principle involved, only what might be possible in practice. And the whole idea of discussing validation of models is surely one of establishing principles to work by, whether validation of a model is actually possible or not.

E I am sure, however, that even if we accept this multiple model framework, and make a whole range of predictions there is still plenty of scope for being wrong. It might only take an extreme flood, for example, to show that a prediction of sediment production might go way outside of any prediction limits, perhaps, due to a landslide changing the nature of the sediment supply. Allowing for multiple acceptable models and uncertainty in predictions does not mean that the predictions will be realistic.

B But there is no implication that they are necessarily realistic – only that they are consistent with the data available (including our perception of what is acceptable). In that respect, events that produce observations that go outside the prediction limits of the models may

---

Figure 4.1 A. Dotty plots of model goodness-of-fit (here, the coefficient of determination based on a simple sum of squared errors) versus parameter value for 1000 random sets of parameter values chosen assuming prior uniform and independent distributions for each parameter. The model is the rainfall–runoff model, TOPMODEL (see Beven et al. 1995; Beven 1997, 2001b). The application is the simulation of discharges for the 1 km^2 Slapton Wood catchment over a 950 hour winter period. Each dot represents one run of the model. The larger dot represents the maximum value over all the sample of realisations. The plots represent the projection of points on the goodness-of-fit response surface onto a single parameter axis, noting that the results of each model run are controlled by the whole set of parameter values. The input rainfall and potential evapotranspiration time series used to run the model are the same in each case. B. As Figure 4.1A but with the results of each run presented in terms of a predicted variable (here, the predicted discharge summed over all time steps) rather than a goodness-of-fit measure. The dots represent exactly the same set of Monte Carlo realisations as for Figure 4.1A. An option to produce a file of such Monte Carlo realisation results is included in the TOPMODEL demonstration Windows software package. The analyses have been made using the demonstration Windows GLUE software. Both software packages can be downloaded from *http://www.es.lancs.ac.uk/hfdg/hfdg.html*

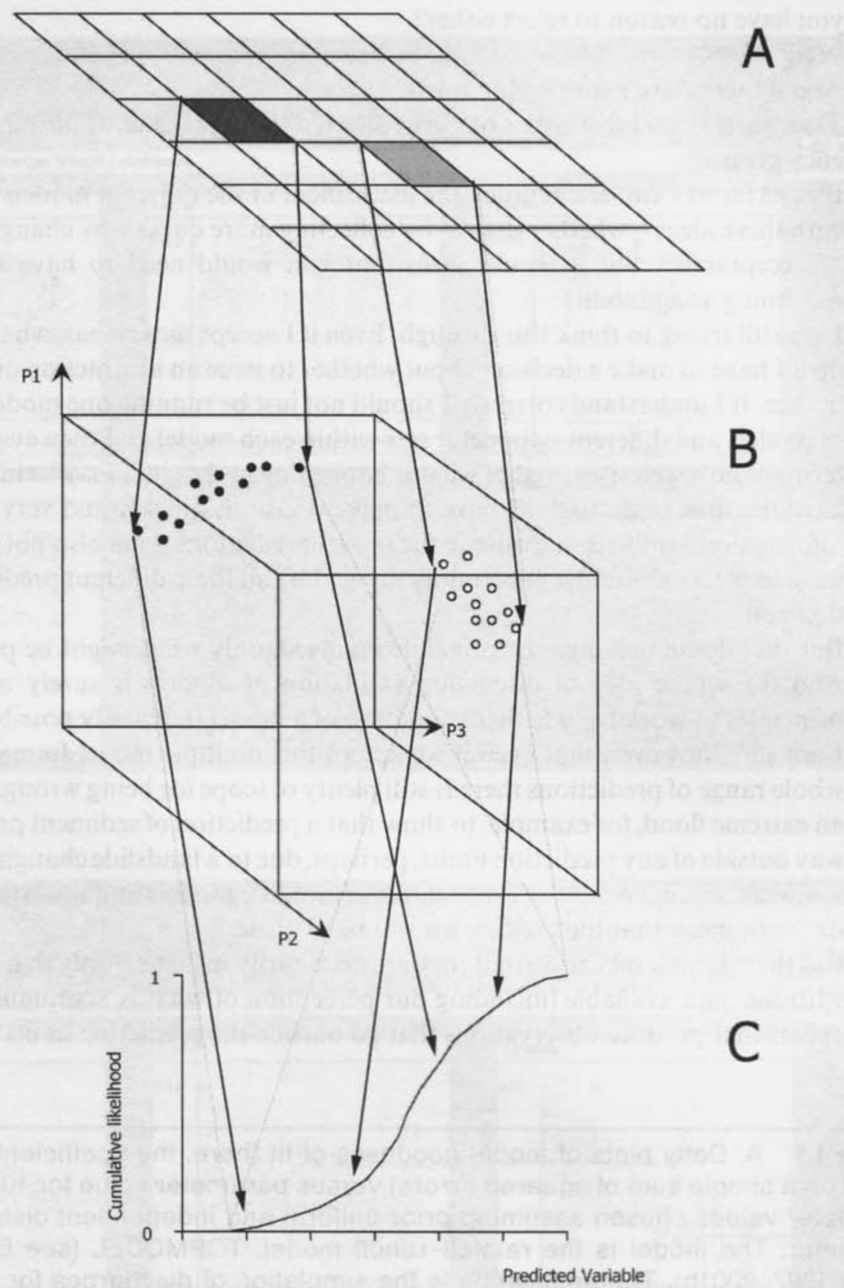


Figure 4.2 A schematic representation of mapping a 'landscape space' into a 'model space', using the same type of Monte Carlo realisations as in Figure 4.1. For any particular point or parcel of the landscape (Figure 4.2A) there may be many acceptable models in the model space (Figure 4.2B). The model space is here represented in terms of different parameter axes but an extension to multiple model structures is also feasible within this framework. The mapping from landscape space to model space, indicated by the dotted lines, is achieved using a weighting function for each model that reflects both the number of points in the landscape that model is representing and also, where appropriate, the goodness-of-fit of that model to any observational data available. These weighting functions can be used in the prediction of either local or landscape scale responses (Figure 4.2C). See Beven (2000) for further discussion of this approach

be the most valuable in refining the model space description in future, in that such failures are a driving force to extend the models to cover a wider range of conditions.

G So do we conclude that we might never be able to validate model predictions of hydrological systems in the sense of showing that a model is a true representation of reality, but can only continue to monitor the performance of models that have been acceptable up to now?

D That process would be much easier if we had measurement techniques that gave us better data about the process variables and fluxes – especially if they gave spatially distributed data like remote sensing of water table depths or latent heat fluxes.

E What do you mean by easier? It might turn out that the user would get so overwhelmed by data that it would make practical prediction and decision-making more difficult. Our experience of using spatially distributed models is that it becomes more and more difficult to set the model up with all the inputs and parameters it requires. Then when you do have some observations to compare the predictions with, which parameters do you start to change to improve the fit or even to evaluate which multiple models are consistent with the spatial data?

P And you have to remember that remote sensing data itself requires a model with parameters to interpret the digital signal into a variable that we are interested in. Such interpretative model parameters are often treated as constants even though they are known to be time variable. This must introduce additional uncertainty even for a variable like surface temperature from thermal imaging, let alone for derived variables like soil moisture values estimated from microwave sensing. So we might be trying to evaluate uncertain models with uncertain data – or even data associated with an uncertain level of uncertainty.

C Hang on. I am not sure my brain can cope with all these layers and layers of uncertainty – how are we ever going to take them all into account? Is it necessary to take them all into account? Is it not true that even for the simplest prediction, all we need to do is to estimate the risk of being wrong, or rather the potential cost of being wrong, however we do it? Routine decisions will surely require simple and easy to understand ways of doing that, not layer upon layer of uncertainties. We don't have the money or manpower to continue monitoring to check and revise our decisions, but there is a cost if the initial decision does not prove to be adequate. And failures almost invariably involve some 'unforeseen circumstances' that are not easily assessed since they are, by definition, unforeseen.

E But how are the levels of risk to be set without someone making some assessment in terms of the value of different types of data in constraining all these layers of uncertainty for different models – at least for some example cases?

C But is that not equivalent to saying that if we could validate a model, say by a split record test, we might be able to have more faith in the predictions?

G How do you mean validate? In such a framework the 'validation' period of a split record test (and any other available data) can effectively be used to condition the model space in terms of the level of acceptability of different parameter sets. Is that not the best confirmation we can have unless we go to collect more, and perhaps different types of, data?

P If I understood correctly, that was also the conclusion of the last paper this morning . . .

ACKNOWLEDGEMENTS

Over the years, many parts of this dialogue have been rehearsed with a variety of friends over a variety of beers and in other seminar and discussion sessions and I am indebted to all those who have, perchance, participated. The origins of the chapter go back before the invitation to contribute to this volume. The first draft was scribbled on a plane to Brazil resulting from an invitation from Robin Clarke to visit Porto Allegre. I am particularly grateful to Tom Dunne, not only for his support in making a sabbatical and seminar series in Santa Barbara possible a few years back, but also for his insightful and valuable review of this manuscript. I am also grateful to Jan Feyen and the Francqui Foundation for support for a stay and extended seminar series in Leuven and other Belgian universities that also helped clarify the presentation. I know that some people still feel these ideas are counterproductive and will undermine hydrology as a science. I hope this chapter will help to show that this need not be the case, and that a continuing dialogue between modellers and measurers is essential to the future of the science. What is still needed is an appropriate framework within which such a dialogue can take place. That is what this contribution is really about!

REFERENCES

Bair, E.S. 1994. Model (in)validation – a view from the courtroom. *Ground Water*, **32**, 530–531.

Bassett, K. 1994. Comments on Richards: the problems of real geomorphology. *Earth Surface Processes and Landforms*, **19**, 273–276.

Beck, M.B. 1987. Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research*, **23**, 1393–1442.

Beck, M.B. 1994. Understanding uncertain environmental systems. In: Grasman, J. and van Straten, G. (eds), *Predictability and Nonlinear Modelling in Natural Sciences and Economics*. Kluwer, Dordrecht, 294–311.

Beck, M.B., Kleissen, F.M. and Wheater, H.S. 1990. Identifying flow paths in models of surface water acidification. *Reviews of Geophysics*, **28**, 207–230.

Beven, K.J. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, **16**, 41–51.

Beven, K.J. 1995. Linking parameters across scales: subgrid parameterisations and scale dependent hydrological models. *Hydrological Processes*, **9**, 507–525.

Beven, K.J. 1996a. Equifinality and uncertainty in geomorphological modelling. In: Rhoads, B.L. and Thorn, C.E. (eds), *The Scientific Nature of Geomorphology*. John Wiley, Chichester, 289–313.

Beven, K.J. 1996b. A discussion of distributed modelling. In: Refsgaard, J.C. and Abbott, M.B. (eds), *Distributed Hydrological Modelling*. Kluwer, Dordrecht, 255–278.

Beven, K.J. 1997. TOPMODEL; a critique. *Hydrological Processes*, **11**, 1069–1086.

Beven, K.J. 2000. Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Science*, **4**(2) 203–213.

Beven, K.J. 2001a. Calibration, validation and equifinality in hydrological modelling: a continuing discussion, this volume.

Beven, K.J. 2001b. *Rainfall–Runoff Modelling: The Primer*. John Wiley, Chichester.

Beven, K.J. Lamb, R., Quinn, P., Romanowicz, R. and Freer, J. 1995, TOPMODEL, in Singh, V.P. (ed.), *Computer Models of Watershed Hydrology*, Water Resource Pubns. Colorado, 627–668.

Binley, A.M., Henry-Poulter, S. and Shaw, B. 1996. Examination of solute transport in an undisturbed soil column using electrical resistance tomography. *Water Resources Research*, **32**, 763–769.

Brooks, R.J., Lerner, D.N. and Tobias, A.M. 1994. Determining the range of predictions of a groundwater model which arises from alternative calibrations. *Water Resources Research*, **30**, 2993–3000.

Chalmers, A. 1990. *Science and its Fabrication*. Open University Press, Milton Keynes.

Davis, P.A., Olague, N.E. and Goodrich, M.T. 1992. Application of a validation strategy to Darcy's experiment. *Advances in Water Resources*, **15**, 175–180.

Emmett, W. 1978. Overland flow. In: Kirkby, M.J. (ed.), *Hillslope Hydrology*. John Wiley, Chichester, 145–176.

Feyerabend, P. 1978. *Against Method*. Verso, London.

Feyerabend, P. 1991. *Three Dialogues on Knowledge*. Blackwell, Oxford.

Flury, M., Flühler, H., Jury, W.A. and Leuenberger, J. 1994. Susceptibility of soils to preferential flow of water: a field study. *Water Resources Research*, **30**, 1945–1954.

Franks, S.W. and Beven, K.J. 1997a. Bayesian estimation of uncertainty in land surface–atmosphere flux predictions. *Journal of Geophysical Research*, **102**(D20), 23991–23999.

Franks, S.W. and Beven, K.J. 1997b. Estimation of evapotranspiration at the landscape scale: a fuzzy disaggregation approach. *Water Resources Research*, **33**, 2929–2938.

Franks, S.W., Gineste, Ph., Beven, K.J. and Merot, Ph. 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Research*, **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. *Water Resources Research*, **32**(7), 2161–2173.

Gupta, H.V., Sorooshian, S. and Yapo, P.O. 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research*, **34**(4), 751–763.

Harvey, J.W., Wagner, B.J. and Bencala, K.E. 1996. Evaluating the reliability of the stream tracer approach to characterise stream–subsurface water exchange. *Water Resources Research*, **32**, 2441–2451.

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. *Hydrological Processes*, **10**(11), 1451–1466.

Howson, C. and Urbach, P. 1993. *Scientific Reasoning, The Bayesian Approach*, 2nd edition. Open Court, Peru, IL.

Loague, K. and Kyriakidis, P.C. 1997. Spatial and temporal variability in the R-5 infiltration data set: déjà vu and rainfall–runoff simulations. *Water Resources Research*, **33**, 2883–2895.

Lohmann, D. and 27 co-authors. 1998. The project for intercomparison of land-surface parameterization schemes (PILPS) Phase 2(c) Red-Arkansas basin experiment: 3. spatial and temporal analysis of water fluxes. *Global and Planetary Change*, **19**, 161–179.

McLaughlin, D. and Townley, L.R. 1996. A reassessment of the groundwater inverse problem. *Water Resources Research*, **32**, 1131–1161.

Morton, A. 1993. Mathematical models: questions of trustworthiness. *British Journal of Philosophy of Science*, **44**, 659–674.

Oreskes, N., Schrader-Frechette, K. and Belitz, K. 1994. Verification, validation and confirmation of numerical models in the earth sciences. *Science*, **263**, 641–646.

Rhoads, B.L. 1994. On being a real geomorphologist. *Earth Surface Processes and Landforms*, **19**, 269–272.

Richards, K.S. 1990. Real geomorphology. *Earth Surface Processes and Landforms*, **15**, 195–197.

Richards, K.S. 1994. Real geomorphology revisited. *Earth Surface Processes and Landforms*, **19**, 277–281.

Romanowicz, R. and Beven, K.J. 1998. Dynamic real-time prediction of flood inundation probabilities. *Hydrological Science Journal*, **43**, 181–196.

Romanowicz, R., Beven, K.J. and Tawn, J. 1996. Bayesian calibration of flood inundation models. In: Anderson, M.G., Walling, D.E. and Bates, P.D. (eds), *Floodplain Processes*, John Wiley, Chichester, 333–360.

Schrader-Frechette, K.S. 1989. Idealised laws, antirealism and applied science: a case in hydrogeology. *Synthese*, **81**, 329–352.

Sen, M.K. and Stoffa, P.L. 1995. *Global Optimisation Methods in Geophysical Inversion.* Elsevier, Amsterdam.

Zak, S., Beven, K.J. and Reynolds, B. 1997. Uncertainty in the estimation of critical loads: a practical methodology. *Soil, Water and Air Pollution*, **98**, 297–316.