HYDROLOGICAL PROCESSES
Hydrol. Process. **15**, 1655–1657 (2001)
DOI: 10.1002/hyp.436

INVITED COMMENTARY

HP TODAY

On hypothesis testing in hydrology

Keith Beven*

Institute of Environmental and Natural Sciences, Lancaster University, Lancaster LA1 4YQ, UK

*Correspondence to:
K. Beven, Institute of Environmental and Natural Sciences, Lancaster University, Lancaster LA1 4YQ, UK. E-mail: K.Beven@lancaster.ac.uk

In a previous comment in *HP Today* (vol. 15, no. 2), I suggested that a worthwhile approach to hydrological modelling is provided by looking at the modelling process as an uncertain mapping of a catchment into a model space. One interesting aspect of such an approach is to investigate the possibility of constraining that mapping, and the consequent predictive uncertainty, using hypothesis testing, with the result that the range of feasible models might be refined. The question that then arises is: what would be suitable hypotheses to test? Over the last year or so, I have been thinking about this problem in relation to the field measurement programme being established under the UK NICHE programme. The aim of these measurements is, in part, understanding the lowland and upland systems to be investigated, but also to test model formulations. In this case therefore, the art of the possible with current field techniques is related very closely to a landscape to model space mapping. The list of possible hypotheses is, of course, very long, but the following have some interesting aspects worthy of discussion.

Hypothesis 1: The water balance can be closed

This hypothesis is assumed by all models, though some do not achieve it. Distributed modellers, for example, face a choice between lumping interpolation errors arising from the nonlinearity of the unsaturated flow conditions into a water balance error or into the calculated boundary fluxes. The errors should get smaller as the model discretization gets finer, but the time steps also then need to be shorter and run times increase rapidly. Modellers should at least be required to report on the magnitude of water balance errors in every report and paper. Should those that do not maintain a water balance be rejected? In principle, yes, except that we cannot currently close the water balance by measurement. Traditionally, there was no direct way of measuring actual evapotranspiration, so errors in long term measured water balances tended to be assigned to the evapotranspiration term, despite the fact that we *know* that rainfall inputs, discharge outputs and changes of storage are not always accurately measured. Now we have an increasing number of sites where continuous eddy correlation measurements of actual evapotranspiration are being achieved. Will this help? Not a lot, actually, since there will still be the problem of fetch

corrections and spatial extrapolation of the measurements, and still no way of checking whether the catchment is indeed watertight. The continuity equation is the most fundamental law in hydrology, but as a hypothesis it would appear that we cannot currently verify it at the catchment scale, except by theoretical argument. Perhaps the latter is, however, sufficient to reject models that do not have water balance closure, so the errors should be reported.

Hypothesis 2: The fast response of a catchment is dominated by surface runoff derived from rainfall

This hypothesis is also commonly assumed by modellers, although very often their models do not actually rely on such an assumption. TOPMODEL is an interesting example in this respect. TOPMODEL predicts a fast runoff component due to rainfall onto a dynamic saturated contributing area that expands and contracts according to the pattern of a topographic index. Tracer experiments suggest, however, that rainfalls can infiltrate even when the soil is saturated and there is a return flow of subsurface water back to the surface. Local circulations in the topsoil associated with preferential flows and variations in the microtopography of the surface appear to be associated with this process. Field studies of surface runoff in Spain by Bergkamp have suggested that the average length of overland flow generated by an infiltration excess mechanism is less than 1 m. Thus, it might be that any surface runoff arriving at a gauging site is a mixture of rainfall derived and subsurface return flows. In addition, when the soil is wet, the effective porosity is small, and the wave speeds associated with the response of the water table can be very fast, inducing a rapid displacement of subsurface water, while speeds of surface runoff flowing through a vegetation or litter layer are not necessarily very fast. Thus the speeds of surface and subsurface responses might be similar, and there may be a fast subsurface response that could be simulated by a model as a surface runoff. In the prediction of hydrographs, of course, it does not really matter to the model (such as TOPMODEL) what process is involved; the model only needs a fast response over (a dynamic) part of a catchment area. If there is an interest in the sources of different waters, however, then the differences may be important and testable using tracers (albeit with great difficulty at the catchment scale). Such a hypothesis could be used to differentiate between model responses, but would require the models to be tested to have components and parameters additional to those required for hydrograph simulation, in order to predict the sources contributing to the fast response.

Hypothesis 3: Knowledge of bedrock topography is needed to adequately predict downslope flows on hillslopes

Recent work funded by NSF at the Panola Catchment, Georgia, USA, has made a detailed examination of the pattern of downslope flows on a hillslope underlain by a shallow granitic bedrock surface. The pattern of flows was found to be controlled much more by the shape of the bedrock surface than the surface hillslope form. This hypothesis is not considered by most modellers and perhaps needs to be investigated for a variety of situations as part of model evaluation. This, of course, raises an interesting set of problems, not the least of which is evaluating the depth to bedrock and its effective permeability (including faults and fractures) in any given situation (at Panola the depths were mapped using a knocking pole on a grid of 2 m over the hillslope investigated, while the granite appeared to be relatively impermeable). There are no generally useful measurement techniques for mapping depth to bedrock (GPR requires careful local calibration, seismic techniques are labour intensive and require careful local interpretation). So, if the information cannot be made generally available, why should such a hypothesis be considered in model evaluation? Because, in the same way that other forms of knowledge are lacking in the application of models (patterns of hydraulic conductivity, macropore networks, etc.), the predictions will be (at least locally) wrong in a way that is not simply a random local effect but which reflects the integration of flow pathways on hillslopes. Similar arguments can be

made in more permeable geologies, such as demonstrated by the tracing experiments of Dale Huff and David Genereux at Walker Branch in Tennessee. There, the incremental discharges to the stream appear to be much more closely related to bedrock structure than to hillslope form. In this case, perhaps if the information about geology was available, distributed models could take it into account. In most catchments, however, it will not be, so models should be expected to be wrong, at least in detail. Does it matter if models are wrong locally. It is partly a matter of opinion. If your opinion, as a determinist, is yes, then the model should be rejected. If your opinion allows that not everything about the subsurface is knowable, then local errors will be accepted and the uncertainty in the predictions should be evaluated (and only some model representations will be rejected).

Hypothesis 4: Graduate students can always think of useful new hypotheses with which to test models (rather than simply calibrating existing models and learning little)

The literature does not give the impression that this hypothesis is being tested with any enthusiasm at the current time. It seems to be the case that graduates are being taught more about how to use models than about how to critically evaluate models. This is hardly surprising given the complexity and learning curves necessary for some of the models that are currently available, and being able to use models is a precursor to being able to criticize them. However, there is surely a need to confront existing models with carefully collected field data sets and new hypotheses that will lead to improvements in how we approach the modelling process.