INVITED COMMENTARY

HP TODAY

HYDROLOGICAL PROCESSES
Hydrol. Process. **17**, 3375–3378 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.5158

Evaluation of catchment models

Thorsten Wagener*

National Science Foundation (NSF), Science and Technology Center (STC) for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ, USA

*Correspondence to:
Thorsten Wagener, SAHRA, Harshbarger Building, University of Arizona, Tucson, AZ 85721-0011, USA.
E-mail: thorsten@sahra.arizona.edu

Introduction

Catchment models are by definition simplified representations of the real world system. This aggregation takes place in space and time and has several important consequences. First, there are no generally applicable rules to perform this aggregation, and the resulting model structure is usually a function of the modeller's hydrological understanding. Secondly, the model parameters cannot be measured directly in many cases, but have to be estimated. In this process, we usually assume that the parameters are constant in time and representative of inherent properties of the real system. This is particularly relevant when transferring parameters to ungauged catchments. The modeller's task is to find a model, i.e. a combination of model structure and parameter set(s), suitable for the anticipated modelling purpose, data and catchment characteristics. Traditionally, the modeller defines an objective function, i.e. some aggregated measure of the distance between simulated and observed system response, and minimizes (or maximizes, depending on definition) its value, a procedure usually called calibration. The aim is to match simulated and observed system behaviour. This is often followed by an application of the identified model to another part of the time series not used during calibration, to show that it can be applied generally. This is usually called the validation step.

Many studies have shown that this type of approach is insufficient to test adequately the suitability of a model and that the scientific conclusions that can be drawn from such a procedure are very limited. A commonly found result is that several, often very different, parameter sets and even model structures are equally acceptable system representations in this context (e.g. Beven and Freer, 2001).

Therefore, we seek to apply approaches to model evaluation that are more discriminative. There are at least three dimensions in which this evaluation should be performed (Figure 1): (1) performance; (2) uncertainty; and (3) 'realism'.

Evaluation of Performance

Evaluating performance generally means analysing how closely the model behaviour matches the behaviour of the real system. The traditional measure of performance is a single objective function, as mentioned above. The (unwanted) consequence of using such a measure is that one loses information that could otherwise be used to discriminate between competing models. An alternative approach

Received 6 June 2003
Accepted 25 June 2003

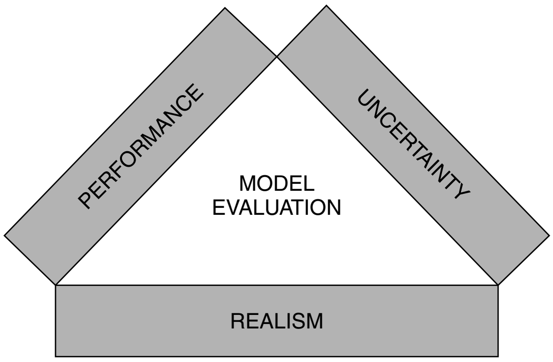

Figure 1. Dimensions of model evaluation

is to apply multiple measures of performance to increase the amount of information retrieved from the available data. These might, for example, be global (e.g. overall bias) and local (e.g. low flow performance) measures (Gupta *et al.*, 2003). This approach has been applied widely in recent years. One outcome of this work is that it has become clear that current catchment models cannot represent low and high flow behaviour of hydrologic systems with a single parameter set. This finding has highlighted some of the problems in current model structures. With respect to the comparison of different model structures, multi-objective approaches might enable us to allocate the performance to individual model components. However, currently, no objective way of defining the optimal number and type of objective function exists for this type of analysis.

Evaluation of Uncertainty

Uncertainty evaluation of models means analysing the range of parameter sets and sometimes even model structures that are viable for an anticipated study. Fortunately, such an analysis is becoming more and more common. The main sources of uncertainty are input and output data, and the model parameters and structure. Set theoretic approaches are often used to estimate these uncertainties. These techniques generally assume that all plausible models should be retained unless and until evidence to the contrary becomes apparent. Common to these approaches is the selection or identification of a *set* (population) of models (different combinations of model structures and parameter values), and the assignment of some relative degree of believability to each member of the set. This degree of believability is translated into uncertainty (confidence) interval estimates on the model output. The approaches differ in the suite of assumptions underlying each technique, based on which methods are used to compute the relative degree of believability (e.g. Beven and Freer, 2001; Vrugt *et al.*, 2003).

The use of multiple objectives plays an important role in this context. With respect to parameter uncertainty, it is likely that different parameters will be more or less identifiable for different objective functions, indicating that different model components are important in fitting various system response modes. A problem that may evolve, however, is the selection of a best parameter set, or a group of best parameter sets, since a different optimal set will often be identified for each objective function. Possible options to deal with this problem are the use of set theory rules or the identification of that group of models (i.e. parameter sets) that is optimal in some multi-objective sense (Beven and Freer, 2001; Gupta *et al.*, 2003). It is currently not obvious how such a model population should be derived. Alternatively, one can derive a compromise solution, for example by the sequential identification of different parameter subsets using a variety of objective functions (Hogue *et al.*, 2000).

An even more serious problem is the fact that we currently have no framework to separate out the amount of uncertainty stemming from the different sources (mainly parameters, data and model structure). The amount of uncertainty allocated to inadequacies of the model structure is particularly unclear and cannot be separated from the approach(es) chosen to represent the other uncertainties at the moment.

It may often be necessary to trade off performance and uncertainty to derive a suitable level of model complexity for the anticipated purpose (Wagener *et al.*, 2002; Figure 2). This is because the information content in the data is limited and potentially supports only a certain number of parameters to be identified (Young, 2001). However, the use of an overparameterized model might be justified in certain cases (Reichert and Omlin, 1997).

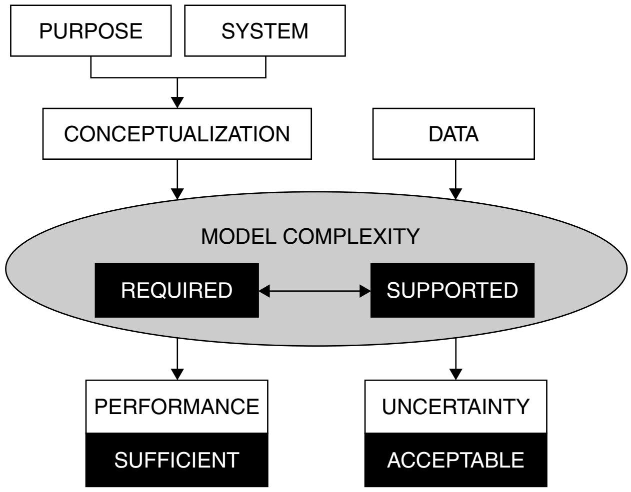

Figure 2. Model complexity evaluation

Evaluation of Realism

A third aspect has recently become part of evaluation studies for catchment models: testing the 'realism' of the model, or how far a model is consistent with our understanding of reality. For example: Do the model components really represent the hydrological processes we want them to represent? Are the parameters really constant in time?

Separating data periods of noise and information for particular parameters can test the first aspect. Information should be found in periods where model components (and therefore parameters) are expected to dominate. From the above discussion on performance evaluation we know that optimum parameters often vary between high- and low-flow periods. The question then is. Can we find techniques to track this variation explicitly and use the result to improve our model structure? Several approaches to do this have been published in recent years, and they seem to give at least some indication about potential problem areas (e.g. Misirli Baysal, 2003; Wagener *et al.*, 2003). In contrast to a batch calibration approach, these techniques estimate the parameter distributions at every time step (or possibly over a window period). A parameter that varies in optimum value between certain response periods suggests that some dynamic aspect of the catchment is not properly represented in the model. This can be used as a guideline for model structural improvements, but one must be careful to ensure that any variation detected is not due to other aspects, e.g. errors in the data.

A potentially more powerful approach is comparing model states to the states of the real system, e.g. groundwater. However, it is not always clear how model variables relate to real-world variables, and what influence model structural error has in this context (e.g. Lamb *et al.*, 1998). Additional information on a catchment's response to constrain the group of consistent models further might be available in experimental catchments (e.g. Seibert and McDonnell, 2002).

Conclusions

The evaluation of hydrological models has evolved considerably in recent years. Multi-objective, set theoretic, and recursive approaches are just three techniques that can help us to better understand the nature of current catchment models. In particular: How do (usually conceptual) parameters relate to real-world characteristics? Answering this question would be one step to modelling ungauged catchments successfully.

One problem of the current state of the art is that no single approach provides us with all the information that we can use in evaluation studies while explicitly considering all aspects of uncertainty. The future will tell whether such an approach exists. For the time being, it might be advisable to combine different techniques in a single study. A collection of currently available tools is presented by Duan *et al.* (2003). New and innovative approaches and tools are still needed to help us to identify hydrologically realistic catchment models.

Acknowledgements

Thanks to the German Academic Exchange Service (DAAD) for support under its postdoctoral fellowship program and to SAHRA (EAR-9876800). Thanks to Hoshin Gupta, Neil McIntyre and Mary Black for comments on this short piece.

References

Beven KJ, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *Journal of Hydrology* 249: 11–29.

Duan Q, Sorooshian S, Gupta HV, Rousseau A, Turcotte R (eds). 2003. *Advances in Calibration of Watershed Models*. AGU Monograph Series. Water Science and Application 6, American Geophysical Union, Washington, DC.

Gupta HV, Sorooshian S, Hogue TS, Boyle DP. 2003. Advances in automatic calibration of watershed models. In *Advances in Calibration of Watershed Models*, Duan Q, Sorooshian S, Gupta HV, Rousseau A, Turcotte R (eds). AGU Monograph Series: Water Science and Application 6, American Geophysical Union, Washington, DC 9–28.

Hogue TS, Sorooshian S, Gupta HV, Holz A, Braatz D. 2000. A multi-step automatic calibration scheme for river forecasting models. *Journal of Hydrometeorology* 1: 524–542.

Lamb R, Beven KJ, Myrabo S. 1998. Use of spatially distributed water table informations to constrain uncertainty in a rainfall-runoff model. *Advances in Water Resources* 22(4): 305–317.

Misirli Baysal F. 2003. *Improving efficiency and effectiveness of Bayesian recursive parameter estimation for hydrologic models*. PhD dissertation, Department of Hydrology and Water Resources, The University of Arizona, Tucson, USA.

Reichert P, Omlin M. 1997. On the usefulness of overparameterized ecological models. *Ecological Modelling* 95: 289–299.

Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research* 38(11): 1241.

Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003. A shuffled complex evolution Metropolis algorithm for optimisation and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39(8): 1201, doi: 10.1029/2002 WR 001 642.

Wagener T, Lees MJ, Wheater HS. 2002. A toolkit for the development and application of parsimonious hydrological models. In *Mathematical Models of large Watershed Hydrology* vol. 1, Singh VP, Frevert D (eds). Water Resources Publishers: 87–136.

Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes* 17: 455–476.

Young PC. 2001. Data-based mechanistic modelling and validation of rainfall–flow processes. In *Model Validation*, Anderson MG, Bates PD (eds). Wiley: 117–161.