Consistency versus Optimality in Environmental Model Identification under Uncertainty

T. Wagener[a]*, P. Dornsife[a] and M. Clark[b]

[a]*Department of Civil and Environmental Engineering, The Pennsylvania State University, Sackett Building, University Park, PA16802, USA, *thorsten@engr.psu.edu*

[b]*University of Colorado/CIRES, 1333 Grandview Avenue, Campus Box 488, Boulder, CO 80309-0488*

Abstract: Current model identification strategies often have the objective of finding the model or model structure which provides the best performance in reproducing the observed response of a system at hand. Such a strategy typically favours more complex (bottom-up) models with a higher degree of freedom and thus larger flexibility. While this bias can be reduced through punishing models for being more complex, real advancements in our understanding with respect to appropriate system representations are made if we quantify the extent to which our model is consistent with the available data. In particular the idea of an optimal parameter set is very weak in the context of highly uncertain environmental modelling exercises using uncertain data and models. This paper discusses the problem of testing model consistency with the aim of falsifying models that are inconsistent with observations or underlying assumptions (e.g. stationary model parameters). Such a strategy can then be included in a general framework for evaluating performance, uncertainty and consistency for model identification.

Keywords: Hydrologic Modelling; Identification; Uncertainty; Consistency; Behavior

1. INTRODUCTION

Environmental models are widely used in research and operational settings. Applications range from predicting watershed response for hydrologic design or forecasting, to evaluate the feasibility of water resources management strategies under climate change, to predict the impact of land use changes on the water balance or ecology, or as load models for water-quality investigations. Available models vary widely in complexity, underlying process descriptions and assumptions, spatial resolution etc. A recurring problem is the identification of an appropriate model given the modelling objective, the available data and the characteristics of the hydrologic system to be modelled [Wagener et al., 2004]. This model identification problem has at least two main elements [Sorooshian and Gupta, 1985]: [1] the identification (development or selection) of one or more appropriate model structure(s), and [2] the identification (estimation) of one or more appropriate parameter set(s) with this (these) model structure(s). Woolhiser and Brakensiek [1982] concluded that objective methods of choosing the best model (structure) had not yet been developed and that this choice remains part of the art of hydrologic modelling. This statement is still valid.

In the past, the search for the appropriate model to represent a given system was largely driven by identifying the one model structure/parameter set combination that minimizes (or maximizes) some measure of performance. This measure of performance was typically one or more numerical objective functions that calculate the aggregated distance between the observed and simulated variable of interest. Such a strategy typically favours more complex (bottom-up) models with a higher degree of freedom and thus larger flexibility. We can punish models for being more complex through the use of information criteria [e.g. Jakeman and Hornberger, 1993], but these have generally not been developed for highly complex and non-linear models like the ones we are typically using. The traditional approach also does not properly exploit the information provided by the comparison of observed and simulated time-series since it aggregates the differences into one (or very few) numerical values. More sophisticated approaches are required to drive model development forward [Wagener, 2003].

In this paper we will discuss the problem of model (structure) identification. We will do so first by discussing the characteristics of hydrologic models and the consequence of hydrologic model characteristics for model identification. We will subsequently suggest characteristics that an identification procedure under uncertainty should contain. This discussion is based on the following premises:

1. The search for optimal models is of limited use in the face of major uncertainties and should be replaced by a search for consistent models (we discuss the term consistent below).
2. Model identification strategies need to expose where and when models fail and work as diagnostic (learning) tools, rather than optimization tools.
3. Model structure comparison needs to include a comparison of model components, not just overall model structures, to be of real value.

2. THE NATURE OF ENVIRONMENTAL MODELS

The response of a watershed to precipitation inputs can be conceptualized as a number of spatially distributed and highly interrelated water, energy and vegetation processes. Any computer-based model of watershed behavior must, therefore, implement this conceptualization using appropriately coupled systems of parametric mathematical expressions; with parameters allowing for flexibility in adapting the model to different (but conceptually similar) watersheds. These parameterizations can be of different levels of complexity, but are, by definition, much simpler than nature itself.

Two important characteristics of this modeling process are relevant to our discussion. First, every environmental model, regardless of how spatially explicit, is to some degree a lumped approximation of a heterogeneous world, so that its parametric equations describe the real-world processes as being aggregated in space and time. Consequently, at least some (if not all) of the model parameters lose some degree of direct physical interpretation (or representativeness) and measurability, and should therefore be understood as being ‘‘conceptual’’ or ‘‘effective’’ parameters (Figure 1). Further, in virtually all cases the scale at which effective parameters are defined (by the model) is different (mostly larger) from the scale at which measurements can be made in the field. Therefore, the correspondence between the field measurements of ‘‘parameters’’ and the effective parameters defined in the model can become very weak, and it becomes necessary to resort to an indirect process of parameter estimation to select parameters values (sets) that produce simulations that closely resemble the observed behavior (i.e. that emulates the behavior of the real world system in relation to the modeler’s needs and objectives). In this process of parameter estimation (often called model calibration) the value of the parameter is adjusted so as to bring the model simulated input–output behavior into close correspondence with the system input–output behavior observed in the field. While environmental models usually contain several such parameters which cannot be assumed to have direct physical (measurable) interpretation, it is often assumed that their values should have physical relevance, insofar as they are believed to correspond to inherent and invariant properties of the environmental system. It is also important to note that the state variable within the model (or within the model element, i.e. a spatial sub-unit of the model) is an ‘‘effective’’ state, e.g. the distribution of moisture content within the model (element) domain is usually lumped into a single aggregate quantity (and less commonly represented as a statistical distribution of this variable within the particular element). This issue must be taken into account when attempting to assimilate data into an environmental model.

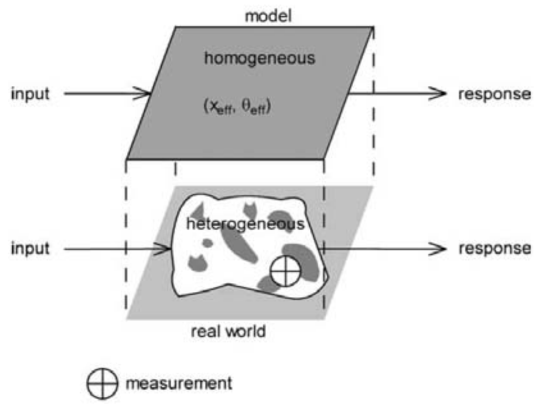

Figure 1. Heterogeneous real world represented by homogeneous model element (though a sub-grid distribution might be included) using effective model parameters, θ, and states, x.

A second characteristic of environmental models is the common practice of specifying/selecting the model structural equations prior to any modeling being undertaken. However, there appear to be no well-defined pathways or objective procedures that will lead to unambiguous selection of an appropriate model structure. Rather, this process is influenced by a combination of factors including observations about the characteristics of the watershed, available data, modeling objective and personal preference.

3. MODEL IDENTIFICATION UNDER UNCERTAINTY

Recent detailed reviews and discussions of hydrologic and environmental model identification have been published in Gupta et al. [2005], and Wagener and Gupta [2005]. These papers discuss the lack of an identification framework that considers all the main sources of uncertainty (data, model structure and parameters, states), the lack of diagnostic capabilities and the need for a shift in paradigm away from the search for optimal models. The approach taken here is that any model identification strategy should explore at least three dimensions [Wagener, 2003]: performance, uncertainty and assumptions.

3.1 Performance

In the past, the search for an optimal performing model was strongly present in the research literature. However, the presence of model structural errors, problems of overly complex models and data uncertainty, and our inability to develop a well-defined calibration problem should lead to the conclusion that a unique and optimal model cannot be robustly identified. An optimal model (parameter set) will very likely change with the chosen objective function, when multiple response variables are considered, with the inclusion of more sources of uncertainty etc. The search for optimal models is, however, necessary to answer for example the question whether a model structure is flexible enough to reproduce the behavior of a particular system. Searches for optimal models in high-dimensional parameter spaces will for a while continue to be done using intelligent optimization algorithms [e.g. Yang et al., in Press] since exhaustive searches will remain limited despite increasing computational power.

In general, *the notion of optimality of models should be replaced by a notion of consistency*. If we can identify all those models that are consistent with the observations of the environmental system at hand – while considering the uncertainties present – then we can more honestly start to analyse how much discriminative power our data contain. The question of what constitutes a consistent model has to be answered for each individual modelling study and will differ for different cases. The notion of consistency should also be extended to the modelling of ungauged sites, usually achieved through a process of model regionalization. Model regionalization will add even more uncertainties and will make the search for an optimal model at the ungauged site an illusion. Instead we can apply a consistency approach again, as for example shown by McIntyre et al. [2005].

3.2 Uncertainty

There has also been a gradual move from procedures that focus on the identification of a single best model towards procedures that seek to reduce the uncertainty in the predictions of all possible models in the presence of uncertainty using various types of ensemble methods (Figure 2). This notion is in line with a move from a philosophy of ‘‘optimization’’ towards a philosophy of ‘‘consistency’’ (i.e. finding models that are consistent with the behavior of the real world system). A variety of methods to create ensembles of simulations exist. Beven and Freer [2001] remind us that a good starting point is the realization that any model identification procedure consists of answering these three questions:

1. What constitutes a behavioral model?
2. How to identify the subset of behavioral models in the feasible model space?
3. How to propagate behavioral predictions into the output space, while considering the uncertainty in the input data, model states, boundary conditions, etc.?

A wide variety of definitions and methods are currently available that attempt to answer these three questions, but there is little guidance regarding which approach to apply under specific circumstances. Progress is likely to come both from research by individual groups and by comparison studies involving larger scale participation and including as many different techniques as possible.

3.3 Assumptions

One approach to model diagnostics is the test of assumptions underlying the developed model structure. Testable assumptions include evaluating whether parameter sensitivity is highest during those periods (response modes) in which parameters (model components) are expected to dominate the model response. As a simple example, a baseflow recession component should be sensitive during long dry spells over the summer. If insensitivity of the parameter is found during such a period, assuming that this is not due to interaction with other parameters describing the same period, then the model component described by the parameter needs to be revisited and probably modified.

Another assumption that is testable is the time-invariance of the model parameters. If the posterior probability distributions for different model parameters are conditioned on those periods for which the parameter shows sensitivity, then a tightening of the probability distribution function (pdf) is to be expected. The area of highest probability during different response modes should be in the same region of the parameter space. If this region varies, e.g. sometimes high parameter values perform better and sometimes low values, than a violation of the time-invariant assumption has been found. The reasons for this violation need to be investigated to suggest model structural improvements. Reasons could include different processes being lumped into a single model component or an apparent change of this physical characteristic of the watershed with time (e.g. vegetation change).

These two assumption tests can be implemented using Kalman filter based [e.g. Beck, 1987] or Monte Carlo based approaches [e.g. Wagener et al., 2003]. Wagener et al. [2003] developed a modification of the generalized Likelihood Uncertainty Estimation [GLUE, Beven and Freer, 2001] algorithm, in which a randomly sampled population of parameter sets is conditioned on different periods of the response variable time-series (e.g. streamflow) using a smoothing approach.

3.4 Consequence for Comparison Studies

A range of comparison studies have been performed in the past. These were either more or less global initiatives (MOPEX, DMIP, HEPEX, PILPS etc.) or studies performed by individuals or small groups. These studies generally compare a wide range of hydrologic models with respect to their performance in reproducing streamflow at the watershed outlet. Typical conclusions are that the difference in performance often decreases once a certain level of model complexity is reached (about 5 parameters), unless there are considerable differences in process descriptions that render certain models unsuitable to represent a particular system [e.g. Jakeman and Hornberger, 1993].

However, many studies have shown that it is difficult to draw conclusions above the ones just stated. This has improved somewhat through the consideration of multiple objectives in the evaluation process [Gupta et al., 2005]. It is likely though that any identification study that includes multiple possible model structures will be inconclusive if the comparison simply investigates the performance of the 'complete' model structure. What is meant by this is that it is very difficult to separate out how far model structures are different if only the overall output of the model is compared to that of other models. Put differently, there are too many degrees of freedom in the models (and too few data points in the experiments) to make profound statements about the functional behavior of different models.

A better picture would be obtained if outputs of the individual model components would be compared, e.g. how different is the description of interception or evapotranspiration? While we generally do not have measured data available to compare these outputs against, we can at least compare them against each other and decide whether one process description is closer to another one with respect to our understanding of what should happen with respect to the system under investigation. Such an approach might also lead us to question our understanding, particularly if a variety of possible models cannot be falsified. A result of this type might lead us to collect new data to test model components separately and find out which system representation is more likely. This general approach requires that a synthetic testing stage be included in the comparison study. A step not usually considered in model identification. Hence we advocate shifting away from a paradigm of simple model intercomparison, towards a paradigm of model deconstruction and controlled comparison. A more detailed analysis might also help us in identifying whether we are dealing simply with different mathematical implementations of the same process understanding, or whether the differences go beyond that.

5. CONCLUSIONS

The process of model identification has long been dominated by the search for more powerful optimization algorithms or better objective functions. We think that this approach is unlikely to yield significant improvements since the identification problem is ill-defined in the presence of model structural and data errors. A paradigm shift needs to occur (and is already occurring) in which we move away from the

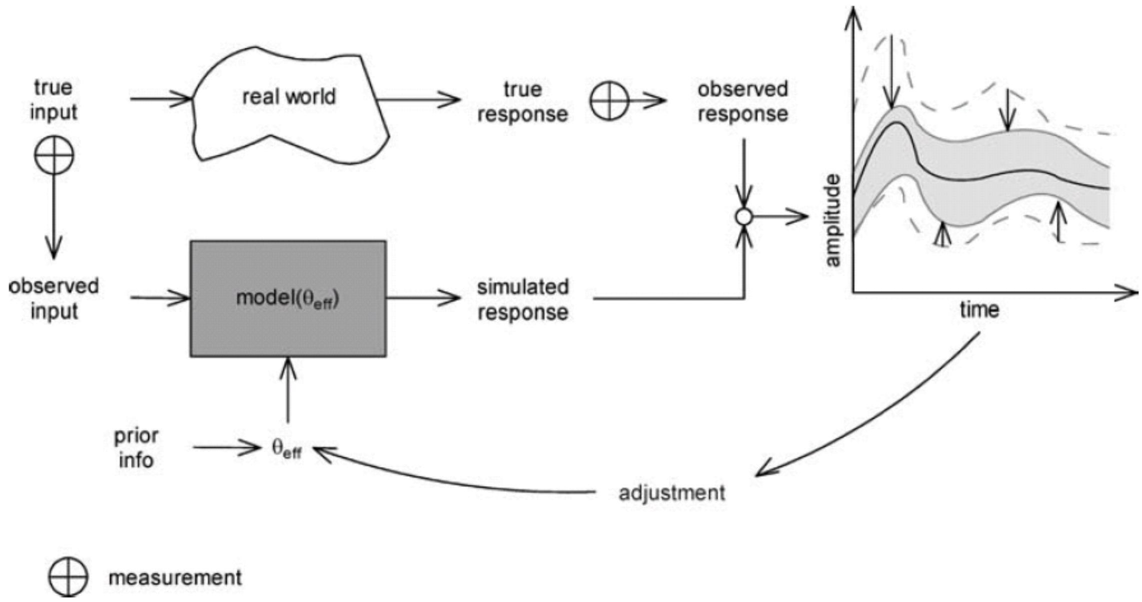

Figure 2. This figure shows a modelling procedure in which a model is constrained using observations of the input-response behavior of the real-world system.

notion of an optimal model towards an ensemble of models that are consistent with the observations of the environmental system at hand. This move needs to include the search for approaches that work as diagnostic or learning tools which help us improve the model structure while extracting as much information from the observations as possible with respect to how model behaviors differ.

6. ACKNOWLEDGEMENTS

Partial support for this work was provided by SAHRA under NSF- STC grant EAR-9876800, and the National Weather Service Office of Hydrology under grant numbers NOAA/NA04NWS4620012,UCAR/NOAA/COMET/S0344674, NOAA/DG 133W-03-SE-0916. We acknowledge the constructive criticism of the three anonymous reviewers that helped improve this paper.

7. REFERENCES

Beck M.B., Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research* 23(8):1393–1442, 1987.

Beven K.J., and J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249(1–4):11–29, 2001.

Gupta H.V., K.J. Beven, and T. Wagener, Model calibration and uncertainty estimation. In: Anderson, M.G. (ed) Encyclopedia of hydrologic sciences. Wiley, Chichester, 2005.

Jakeman, A.J., and G.M. Hornberger, How much complexity is warranted in a rainfall-runoff model? *Water Resources Research* 29(8): Pages 2637-2649, 1993.

McIntyre N., H. Lee, H. Wheater, A. Young, and T. Wagener, Ensemble predictions of runoff in ungauged catchments, *Water Resources Research* 41, W12434, doi:10.1029/2005WR004289, 2005.

Sorooshian, S., and V.K. Gupta, The Analysis of Structural Identifiability - Theory and Application to Conceptual Rainfall-Runoff Models. *Water Resources Research*, 21(4): 487-495, 1985.

Wagener, T. and H.V. Gupta, Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment* DOI 10.1007/s00477-005-0006-5, 2005.

Wagener T. Evaluation of catchment models. *Hydrological Processes* 17:3375–3378, 2003.

Wagener T., H.S. Wheater, and H.V. Gupta, Rainfall-runoff modelling in gauged and ungauged catchments. Imperial College Press, London, 2004.

Wagener T., N. McIntyre, M.J. Lees, H.S. Wheater, and H.V. Gupta, Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes* 17(2):455–476, 2003.

Woolhiser, D. A. and D. L. Brakensiek, Hydrologic System Synthesis. In Hydrologic Modeling of Small Watersheds. In: Haan, C.T., H.P. Johnson, and D.L. Brakensiek. ASAE Monograph No. 5. St. Joseph, MI: 3-16, 1982.

Yang, T., P. Reed, and T. Wagener, How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydroly and Earth System Sciences* In Press, 2006.