Stoch Environ Res Risk Assess (2005) 19: 378–387
DOI 10.1007/s00477-005-0006-5

ORIGINAL PAPER

Thorsten Wagener · Hoshin V. Gupta

Model identification for hydrological forecasting under uncertainty

Published online: 13 September 2005

Abstract Methods for the identification of models for hydrological forecasting have to consider the specific nature of these models and the uncertainties present in the modeling process. Current approaches fail to fully incorporate these two aspects. In this paper we review the nature of hydrological models and the consequences of this nature for the task of model identification. We then continue to discuss the history ("The need for more POWER"), the current state ("Learning from other fields") and the future ("Towards a general framework") of model identification. The discussion closes with a list of desirable features for an identification framework under uncertainty and open research questions in need of answers before such a framework can be implemented.

Keywords Hydrological models · Model identification · Flood forecasting · Uncertainty · Data assimilation · Model realism · Predictions in ungauged basins

1 Introduction

Water resources and ecosystems around the globe are coming under increasing stress due to both natural and human-induced climate change/variability, and due to increasing population pressure leading to increased water use, land use changes including urbanization, etc. (Sivapalan et al. 2003). Recent research suggests that a warmer climate might increase the frequency of extreme events, i.e. heavy precipitation and droughts (Karl et al. 1995; Tsonis 1996). This change includes an increase of the amount of randomness in the system, which in turn leads to a decrease in the predictability of the system (Tsonis 2004). Both flood risk and the occurrence of drought therefore appear to be increasing, while regions where observational data are sparse tend to be the most vulnerable. This is particularly true for many less developed countries where floods and droughts consistently result in substantial loss of life. A devastating example is the 1991 flood in Bangladesh that killed 140,000 people in a period of just 2 days (Kundzewicz and Kaczmarek 2000). These areas also include many countries in which the decline of hydrological measurement networks is most severe, leading to a need for modeling approaches that can provide reliable predictions in ungauged watersheds (e.g. Sivapalan et al. 2003). But countries in the developed world are also increasingly affected; for example, the 1993 flood in the Midwest USA showed recurrence intervals between 100 and 500 years in some locations along the Mississippi and Missouri Rivers (Kundzewicz and Kaczmarek 2000).

Sustainable management policies are required to respond to these trends. Among the sources of information available to policy makers are predictive models capable of simulating the behavior of hydrological systems over a broad range of space and time scales, and (potential) climates. Many such models are in operational use, applied to assess future climate or land use scenarios, analyze current systems, forecast flood events in real-time, or in support of sustainable water resources management (e.g. Singh and Frevert 2002a, b). In general, models attempt to represent the complex, spatially distributed, interactions of water, energy and vegetation by means of mathematical equations.

T. Wagener (✉)
Department of Civil and Environmental Engineering,
Pennsylvania State University, 226B Sackett Bldg.,
University Park, PA 16802, USA
E-mail: thorsten@engr.psu.edu
Tel.: +1-814-86556733085
Fax: +1-814-8637304

H. V. Gupta
SAHRA & Department of Hydrology and Water Resources,
University of Arizona, Marshall Building, 845 E. Park Avenue,
Tucson, AZ 85721, USA

These predictive models conceptualize (and thus simplify) reality using systems of coupled expressions that can be solved using a computer. The models are typically driven by observations of forcing variables such as precipitation, temperature, etc. The conceptualization process and the need for data are just the first of many sources of uncertainty impacting the modeling process. Perhaps the most fundamental of these uncertainties is the conceptualization of reality, which reflects the hydrological modeler's (incomplete and/or biased) understanding and abstraction of the important processes and interactions in the real-world system and will therefore vary from person to person. At the same time, data used to drive the model will only be an approximation of the real-world forcing due to problems of detection and measurement (e.g. precipitation uncertainty resulting from inadequate spatio-temporal sampling densities). Good modeling practice requires that the forcing data, structural and other uncertainties be propagated into the model predictions and communicated in an appropriate manner to the decision maker or stakeholder, thereby allowing an appropriate degree of confidence to be attributed to the model results. Recently, there has been a surge in attention given to methods for the treatment of model uncertainty as (a) decision makers have begun to push for better quantification of the *accuracy and precision* of hydrological model predictions, (b) interest has grown in methods for properly merging data with models and for reducing predictive uncertainty and (c) scientists have begun to search for better ways to represent what is, and is not, well understood about the hydrological systems they study. Another aspect that makes the inclusion of uncertainty estimates important is the (obvious) fact that *extreme events are rare* (Kundzewicz 2002; Hall and Anderson 2002), which limits the data available to condition models for many relevant cases and often synthetic scenarios have to be used.

This discussion paper exemplifies the need for probabilistic hydrological predictions. This need has generally been acknowledged and the importance of reducing uncertainty has been made the focus of the International Association of Hydrological Science's (IAHS) Prediction in Ungauged Basins (PUB) initiative (Sivapalan et al. 2003). However, how to estimate this uncertainty is still an open question. In the subsequent sections, we review the nature of hydrological models and the consequence of this nature for any system identification approach in the presence of uncertainty. We then proceed to discuss the current state-of-the-art of model identification and how it has emerged from earlier thinking, including current bottlenecks for scientific progress and potential ways forward. We close by listing several open questions in need of attention. The intention of this paper is not to provide an extensive review, which would be difficult within the limited space available, but a discussion of the main historic and recent developments, directions and some open questions in the context of an increasing appreciation and inclusion of uncertainty in hydrological forecasting.

2 The nature of hydrological models

The hydrological behavior of a watershed 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 hydrological 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 (Fig. 1). 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. Further, in virtually all cases the scale at which effective parameters are defined (by the model) is typically 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

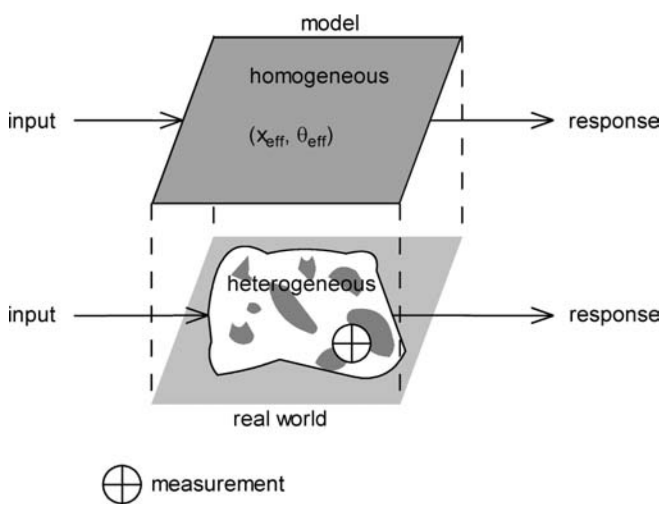


Fig. 1 Heterogeneous real world represented by homogeneous model element (though a sub-grid distribution might be included) using effective model parameters, *theta*, and states, *x*. Adapted from Grayson and Blöschl (2001)

indirect process of parameter estimation to select parameters values (sets) that provide appropriate model behavior (i.e. that emulates the behavior of the real-world system in relation to the modeler's needs and objectives) (e.g. Hornberger and Spear 1981; Young 1983; Van Straten and Keesman 1991; Beven 2005). 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 hydrological 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 hydrological 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, i.e. 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 fact must be taken into account when attempting to assimilate soil moisture content information (typically collected as point measurements or inferred from remote sensing measurements) into a hydrological model. The ability to assimilate information into the model depends on the degree of correlation between the model state variable(s) and the observed variable.

A second characteristic of hydrological models is the common practice of specifying/selecting the model structural equations *prior* to any modeling being undertaken (Wheater et al. 1993). 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 The nature of hydrological model identification under uncertainty

The process of identifying one or more suitable models for a specific application and then using them to derive model predictions requires that the following components be defined, measured or in some way estimated: (1) Model: structure, parameters and states; also, initial and boundary conditions, and (2) Data: measurements of forcing and response.

It is commonly accepted that these elements will all contain uncertainties that can affect the model predictions. These uncertainties stem from a variety of sources (e.g. Melching 1995; Gupta et al. 2005), and are related to our understanding and measurement capabilities regarding the real-world system under study:

- *Perceptual model uncertainty*, i.e. the conceptual representation of the watershed that is subsequently translated into mathematical (numerical) form in the model. The perceptual model (PM; Beven 2001) is based on our understanding of the real-world watershed system, i.e. flowpaths, number and location of state variables, runoff production mechanisms, etc. This understanding might be poor, particularly for aspects relating to subsurface system characteristics, and therefore our PM might be highly uncertain (Neuman 2002).

the data:

- *Data uncertainty*, i.e. uncertainty caused by errors in the measurement of input (including forcing) and output data, or by data processing. Additional uncertainty is introduced if long-term predictions are made, for instance in the case of climate change scenarios for which per definition no observations are available. A hydrological model might also be applied in integrated systems, e.g. connected to a socio-economic model, to assess for examples impacts of water resources changes on economic behavior. Data to constrain these integrated models is rarely available (e.g. Letcher et al. 2004). An element of data processing uncertainty is introduced when a model is required to interpret the actual measurement. A typical example is the use of radar rainfall measurements. These are measurements of reflectivity that have to be transformed to rainfall estimates using a (empirical) model with a chosen functional relationship and calibrated parameters, both of which can be highly uncertain.

and the mathematical/numerical model(s) and its (their) components:

- *Parameter estimation uncertainty*, i.e. the inability to uniquely locate a 'best' parameter set (model, i.e. a model structure parameter set combination) based on the available information. The lack of correlation often found between conceptual model parameters and physical watershed characteristics will commonly result in significant prediction uncertainty if the model is extrapolated to predict the system behavior under changed conditions (e.g. land use change or urbanization) or to simulate the behavior of a similar, but geographically different watersheds for which no observations of the variable of interest are available, i.e. the ungauged case. Changes in the represented system have to be considered through adjustments of the model parameters (or even the model structure), and the degree of adjustment has so far been difficult to determine without measurements of the changed system response.

- *Model structural uncertainty* introduced through simplifications and/or inadequacies and/or ambiguity in the description of real-world processes.

There will be some initial uncertainty in the model state(s) at the beginning of the modeled time period. This type of uncertainty can usually be taken care of through the use of a warm-up (spin-up) period or by optimizing the initial state(s) to fit the beginning of the observed time-series. Errors in the model (structure and parameters) and in the observations will also commonly cause the states to deviate from the actual state of the system in subsequent time periods. This problem is often reduced using data assimilation techniques as discussed later.

Another typical problem is a difference in scale and/or definition of observed and simulated variables that introduces uncertainty as already alluded to in the discussion of Fig. 1 (see also discussion by Beven 2001). Observations of soil moisture might for example be based on point measurements or (very shallow) remote sensing data whereas the model variable might represent the aggregate soil moisture content in the unsaturated zone over a particular grid box. In such a case, a perfect match between observations and simulation cannot be expected and is not even something the modeler should strive for (Gupta et al. 2005). This type of representation uncertainty (or error) could for example be attributed to the data uncertainty (e.g. a point estimate is used to represent a larger area), or it could be attributed to model structural error (e.g. an average value is predicted instead of a distribution for a certain area) (Sørensen et al. 2004; Fukomori et al. 1999).

3.1 Past model identification—the need for more "POWER"

Hydrological model structures of the continuous watershed response became feasible in the 1960s. They were usually relatively straightforward lumped, conceptual mathematical representations of the (perceived to be important) hydrological processes, with little (if any) consideration of issues such as identifiability of the parameters or information content of the watershed response observations. It became quickly apparent that the parameters of such models could not be directly estimated through measurements in the field and that some sort of adjustment (fine-tuning) of the parameters was required to match simulated system responses with observations (e.g. Dawdy and O'Donnell 1965). Adjustment approaches were initially based on manual perturbation of the parameter values. Over the years, a variety of manual calibration procedures have been developed, some having reached rather high levels of sophistication and able to provide very good (and hydrologically realistic) model parameters and predictions, i.e. a well-calibrated model (Burnash 1995). Necessary conditions for a RR model to be "well-calibrated" are that it has (at least) the following three characteristics (Gupta et al. 2005; Wagener et al. 2003b):

1. The input–state–output behavior of the model is consistent with the measurements of watershed behavior.
2. The model predictions are accurate (i.e. they have negligible bias) and precise (i.e. the prediction uncertainty is relatively small).
3. The model structure and behavior are consistent with a current hydrological understanding of reality. This last point is often ignored in operational settings, where the focus is generally on "useful" rather than realistic models. This will be an adequate approach in many cases, but will eventually lead to limitations of potential model uses. This problem is exemplified in the current attempts to modeling watershed residence times and flowpaths (McDonnell 2003). This aspect, often not crucial for reliable quantitative flow predictions, is however relevant for many of today's environmental problems, but cannot be simulated by many of the currently available models.

The high number of non-linearly interacting parameters present in most hydrological models makes manual calibration a very labor-intensive and difficult process requiring considerable experience. This experiential knowledge is difficult to acquire and cannot easily be transferred from one hydrologist to the next. In addition, manual calibration does not formally incorporate an analysis of uncertainty, as is required in a modern decision making context.

The obvious advantages of computer-based "automatic" calibration procedures began to spark interest in such approaches as soon as computers became more easily available for research. In automatic calibration, the ability of a parameter set to reproduce the observed system response is measured (summarized) by means of an "objective function" (also sometimes called loss or cost function). Typically, this objective function is an aggregated measure of the residuals, i.e. the differences between observed and simulated responses at each time-step. An important early example of automatic calibration is the dissertation work by Ibbitt (in the early 1970s) in which a variety of automated approaches were applied to several watershed models of varying complexity (Ibbitt 1970; see also Ibbitt and O'Donnell 1971). The approaches were mainly based on local-search optimization techniques, i.e. methods that start from a selected initial point in the parameter space and then "walk" through it, following some pre-defined rule system, to iteratively search for parameter sets that yield progressively better objective function values. Ibbitt (1970) found that it is difficult to conclude when the "best" parameter set has been found, because the result depends both on the chosen method and the initial starting parameter set.

The application of local-search calibration approaches to all but the most simple watershed models has been largely unsuccessful. In reflection of this, Johnston and Pilgrim (1976) reported the failure of their 2-year quest to find an optimal parameter set for a typical conceptual RR model. Their honesty in reporting this failure ultimately led to a paradigm shift as researchers started to look closely at the possible reasons for this lack of success.

The difficulty of the task at hand really only became clear in the early 1990s when Duan et al. (1992) conducted a detailed study of the characteristics of the response surface that any search algorithm has to explore. Their studies showed that the specific characteristics of the response surface, i.e. the $n+1$ dimensional space of n model parameters and an objective function, of hydrological models give rise to conditions that make it extremely difficult for local optimization strategies to be successful. They listed the following characteristics commonly associated with the response surface of a typical hydrological model:

- It contains more than one main region of attraction.
- Each region of attraction contains many local optima.
- It is rough with discontinuous derivatives.
- It is flat in many regions, particularly in the vicinity of the optimum, with significantly different parameter sensitivities.
- Its shape includes long and curved ridges.

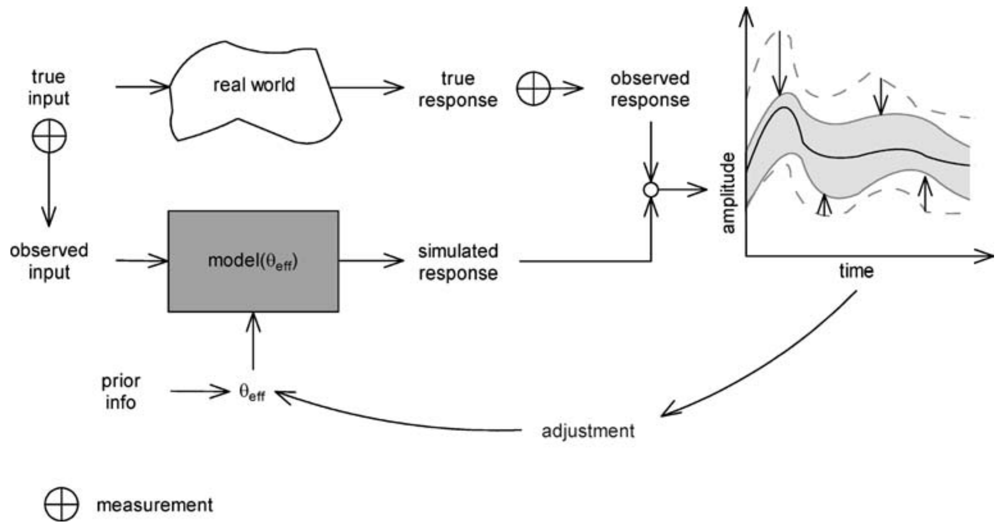
Concluding that optimization strategies need to be powerful enough to overcome the search difficulties presented by these response surface characteristics, Duan et al. (1992) developed the Shuffled Complex Evolution (SCE-UA) global optimization method. The SCE-UA algorithm has since been proven, via many studies, to be highly reliable in locating the optimum (where one exists) on the response surfaces of typical hydrological models. However, in a follow-up paper, Sorooshian et al. (1993) used SCE-UA to show that several different parameter combinations of the relatively complex Sacramento model (13 free parameters) could be found which produced essentially identical objective function values, thereby indicating that not all of the parameter uncertainty can be resolved through an efficient global optimizer. Similar observations of multiple parameter combinations producing similar performances have also been made by others (e.g. Binley and Beven 1991; Beven and Binley 1992; Spear 1995; Young et al. 1996). Part of this problem had been attributed to overly complex models for the information content of the system response data available, usually streamflow (e.g. Young 1992, 1998).

Thus, at least three different schools of thought emerged from this period, each attempting to address the problem of parameter non-uniqueness in different ways:

- *Equifinality*. The philosophy that, in any modeling study of reasonably complex environmental systems, multiple models can be found that provide predictions that are consistent with available observations. This philosophy has been promoted by, for example, the Generalized Likelihood Uncertainty Estimation (GLUE) approach. See, for example, Beven and Binley (1992), Freer et al. (1996, 2004), Romanowicz et al. (1994), Beven (1993, 2005) and other papers.
- *Parsimony*. The philosophy that many hydrological models are too complex (in terms of the number of parameters) for the task at hand and in relation to the information content available in the observed time-series of the system response, and that simpler models can (and should) be used for practical modeling applications. This philosophy has been promoted by, for example Beck (1987), Young (1992), Jakeman and Hornberger (1993), Wheater et al. (1993), Young and Beven (1994), Young et al. (1996, 1997), Hooper et al. (1998) and others. Note, however, that Beck (1987; see also Reichert and Omlin 1999) already pointed out the dilemma that a simple model based on the modes found in the observations could miss key aspects of the system behavior because pertinent modes of behavior might not have been excited during the model identification process.
- *Power*. The philosophy that commonly used approaches to model (parameter) identification lack power, i.e. are poor in discriminative ability and are inefficient in the use of the available qualitative and quantitative information, including the information available in the observations. This philosophy has been promoted by, for example, Gupta et al. (1998), Willmott (1982), Boyle et al. (2000), Vrugt et al. (2003b), Wagener et al. (2003a).

It must be stressed that these three schools really differ only in emphasis and are not mutually exclusive. Clearly, the ideas of equifinality and of parsimony can only be fully explored if powerful identification procedures are applied. There has also been a gradual but general recognition of the need to 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 one or more models (Fig. 2), i.e. 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; e.g. Beven 2005; Seibert and McDonnell 2002; Gupta et al. 2005). The conclusion that *consistency is more important than optimality* seems generally accepted as a working paradigm, in particular since the notion of optimality might be fragile in the presence of the above mentioned uncertainties anyway. As mentioned earlier, the parameter sets identified in a typical single objective framework are often uncertain, and this uncertainty should be considered in a way (e.g. stochastic or set-theoretic framework) that allows for the identification of the posterior probabilistic distribution of parameter

Fig. 2 Constraining in contrast to single-model identification. Identification is seen as a process of uncertainty reduction



values. One simple way to achieve this is by propagating an estimate of the mean and the variance of each parameter into the output space (see Melching 1995, for a review of algorithms). Alternatively, Monte Carlo sampling strategies can be employed, either using simple approaches (e.g. uniform random sampling, Latin Hypercube sampling) or using advanced strategies that utilize the information obtained in initial steps (e.g. Markov Chain Monte Carlo sampling, Metropolis-Hastings) based on some assumptions regarding the expected posterior distribution. While the latter approaches do not require information regarding the expected posterior distribution, they nonetheless gain convergence speed when posterior and proposal distributions are similar.

While an improved consideration of parameter uncertainty is a significant step forward, Gupta et al. (1998) have pointed out that this by itself will not be sufficient because the identification problem itself is commonly ill-posed and uncertain. In particular, the use of an identification framework based on a single objective function is based on the erroneous assumption that all the available information can be summarized (in a recoverable form) using a single aggregate measure of model performance, leading unavoidably to the loss of information (see e.g. Wagener et al. 2001) and therefore poor discriminative power. Gupta et al. (1998) proposed the use of a multi-objective (MO) framework as a way towards addressing this problem (for MO approaches see also for example Yang and Haan 1991; Vrugt et al. 2003a). This and the following work along those lines has suggested that the use of a single objective function may only allow for the identification of around 3–5 parameters (see for example results in Jakeman and Hornberger 1993; Perrin et al. 2000), and that this limit is a consequence of the weakness of single objective identification procedures, rather than being a consequence of insufficient available information. It may therefore be dangerous (and premature) to conclude equifinality or that only a parsimonious low-order model can be identified given the available data. The work has also shown that the amount of structural uncertainty present had typically been underestimated in the past.

From one point of view, *stochastic or set-theoretic single objective identification procedures allow for the consideration of uncertainty in the solution (answer), while multi-objective identification procedures allow for the consideration of uncertainty in the problem (question) formulation*. Of course, we can always infer something about the quality of our problem definition from the uncertainty in the solution, e.g. the problem could be too ill-defined to significantly constrain the model or too well-defined to maintain a sufficient range of model behaviors for a validation period. If we are uncertain about what the definition of the problem should be, then multiple-objectives help us to consider this. In addition, we of course need a multi-objective approach when we have multiple output variables (e.g. Hooper et al. 1988; Madsen 2003; Madsen et al. 2003).

3.2 Current model identification—learning from other fields

Currently available methods to consider model uncertainty mainly map the uncertainty entirely into the parameter space, then sample from the parameter space to propagate the uncertainty into the model predictions. The other sources of uncertainty listed in Sect. 2 have not yet become standard aspects of model uncertainty approaches (for exceptions and discussions see for example Kavetski et al. 2003; Freer et al. 2004; Gupta et al. 2005). Gupta et al. (2005) note that usually only a single value of the difference between model prediction and observation is available at any given time-step, most commonly streamflow. This single residual value may contain little (if any) information that would allow for it to be separated into the different contributions that gave rise to its existence.

As mentioned earlier, single-objective batch approaches lead to the loss of information due to the aggregation into an overall summary measure. This has led to an increased interest in recursive approaches that allow for more of the available information to be used for model (parameter) identification. Thiemann et al. (2001) demonstrated this using a simplified Bayesian framework, but reported that the posterior probability distributions eventually collapsed due to an insufficient sampling density in the high probability regions of the parameter space. Misirli (2003) later addressed this problem by introducing the use of parameter re-sampling, and by incorporating entropy corrections to account for model and data uncertainty. Wagener et al. (2003a, b) introduced a set-theoretic smoothing algorithm based on a modification of the GLUE approach to separate time periods of information and noise, and to track the variation of optimal parameter regions in time. Using this approach they showed how parameter variation can be used to analyze model structural error. Other researchers have introduced alternative types of recursive identification or sensitivity analysis approaches with similar objectives, but based on somewhat different underlying assumptions and sampling strategies (e.g. Beck 1987; Young 2001; Vrugt et al. 2002; Gooseff et al. 2005).

Of particular interest for the issue of flood forecasting (or other short-term predictions) are methods that incorporate state-updating (data assimilation) to optimize the short-term performance of models, even if the temporal persistence of this updating might not be long. Most approaches are based on filtering techniques ranging from the simple Kalman (KF) and extended Kalman (EKF) filters to the more recent Ensemble Kalman filters (EnKF) that are better able to cope with the non-linearities inherent to hydrological models (for different filter implementations see for example Kitanidis and Bras 1980a, b; Young 2002; Reichle et al. 2002; Vrugt et al. 2005). However, such filters typically propagate only second order characteristics and use linear correlation-based updating rules that are unable to summarize the higher order characteristics required to describe the highly skewed posterior distributions characteristic of hydrological models. In case of the EnKF, an ensemble approximation of the entire distribution of the state is propagated, while only the first and second order moments are used for updating. Issues regarding how the structures of errors and covariances may change in time, and what effects this may have when using data assimilation techniques that assume that those error characteristics are time-invariant are still in discussion.

A general conclusion that can be drawn from the studies mentioned above is that model structural error can be the most significant component of the overall predictive uncertainty. While research to date has focused mainly on the treatment of parameter and data uncertainty, it has recently become apparent that the impact of model structural error can often be more severe than that of uncertain parameters (Carrera and Neuman 1986). Several researchers have attempted to assess the level of structural uncertainty present in hydrological models using multi-objective approaches to demonstrate the inability of the model to simultaneously fit all system response modes (Gupta et al. 1998; Boyle et al. 2000; Madsen 2000; Wagener et al. 2001, 2003a; Vrugt et al 2003a). It is clear that structural error gives rise to time-varying bias in the model predictions, but there is little understanding of its magnitude in relation to other uncertainties present, and how the impact of this bias changes in time. If prediction is the main aim, then the best currently available approach may be the use of a data assimilation technique to bring about temporary (short-term) reductions in this and other biases.

The model structural space is infinite and it is therefore only possible to find a currently best model structure (rather than a 'true' structure) or currently acceptable set of model structures by comparing them to all available observations (Butts et al. 2004). Figure 3 illustrates a range of potential situations that one could encounter with respect to the knowledge of the underlying hydrological system. It ranges from a well-known system (with a high level of prior system knowledge), for which the model structure can be very accurately defined, to a poorly known system where very little may be known about how the system functions, i.e. the structure has to be based on a very uncertain perceptual model. When the model we have a chosen is a relatively good description of the system under study, it may be possible to evaluate the effects of system uncertainty by simply perturbing the model structure to account for any remaining uncertainty. Another approach, introduced by Kennedy and O'Hagan (2001), uses a model inadequacy function in a Bayesian context to account for structural uncertainty by developing a more complex error model. The underlying model assumptions can then be tested by examining the characteristics of the residual time-series. When the system is less certain, it may be necessary to incorporate several potential system representations, for example using a Bayesian averaging scheme (Hoeting et al. 1999; Neuman 2002). The GLUE approach mentioned earlier (and several others) also theoretically allows for the consideration of multiple model structures. However, if the underlying system is very poorly understood, but input–output response data are available, then a data-based approach might be a way to gain some understanding of the underlying model structure, though much work remains to be done to further develop this approach (Young 2001). In general, more research is needed to develop systematic approaches to handle model structural uncertainty. A less formal approach to model development might be needed to open up new pathways in developing new model structures beyond our current ideas. This could include soft modeling approaches that do not pre-impose structural constraints but let the data (response observations) constrain potential behavior (Fig. 3).

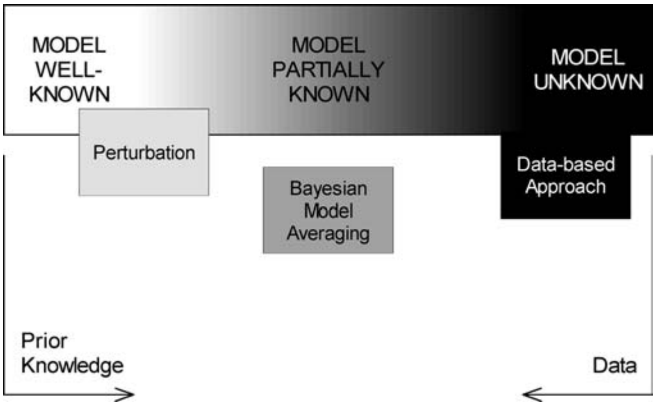


Fig. 3 Approaches to consider model structural uncertainty based on knowledge of underlying system

At present, it appears that much of the published work is focusing largely on developing solutions to the technical problems of sampling and propagating multivariate distributions to yield uncertainty propagation schemes more appropriate to the specific characteristics of hydrological models. Much might be gained in this context from adoption or adaptation of methods developed in other fields where such investigations are more advanced. For example, data assimilation has been extensively studied in meteorology and oceanography. This will for example lead to techniques less restricted by current assumptions as mentioned above.

3.3 Future model identification—towards a general framework

As discussed, a wide variety of methods to considering uncertainty in hydrological modeling is currently available. These methods differ in underlying philosophy, assumptions made; sampling strategies applied, etc. (see for example Beven 2005). However, there is remarkably poor understanding of the effect of these differences, and woefully little guidance on what approach should be used under what circumstances. An important driver to increase our understanding in this regard (and for the development of uncertainty analysis techniques) will be a search for a unifying uncertainty framework, as is currently being attempted under the umbrella of the PUB initiative of the IAHS. Such a framework does not currently exist; with the result that model estimated predictive uncertainties vary considerably depending on the underlying assumptions of any given technique. It is hoped that this initiative will encourage studies that compare uncertainty analysis techniques, thereby helping to provide better understanding of the effects of underlying assumptions, and ultimately resulting in guidance as to what uncertainty approach to use under what circumstances. Beven and Freer (2001) remind us that a good starting point is the realization that any model identification procedure should consider the following necessary steps:

1. Define how to measure the level of consistency between modeled and observed system behavior.
2. Locate all (or a representative set of) models that comply with this definition in the feasible model space.
3. Propagate the predictions of these models into the output space while considering other uncertainties.

Approaches vary widely in the manner in which these steps are implemented. Comparing uncertainty analysis techniques based on these three steps will increase our understanding about the consequences of differences in assumptions and technical implementation.

Gupta et al. (2005) compiled a list of desirable features for an identification framework under uncertainty that might, or might not, be achievable, presented here with some example science questions that need addressing if such a technique should be achieved. The model identification strategy should:

1. Explicitly incorporate all sources of uncertainty. Question: How can we estimate the level of contribution of the different sources of uncertainty to the overall uncertainty?
2. Incorporate multiple sources of information. Question: How much of this information can be assimilated into models, considering model structural error and differences in predicted and observed variables?
3. Permit recursive processing of data. Question: What approach can be used to sample and propagate from multivariate distributions, including higher order moments?
4. Provide probabilistic estimates of model outputs. Question: The level of predictive uncertainty differs widely if different approaches are tested, where do the differences originate and what do they mean for any decision making context? How can we define experiments that help in the development of guidance about what uncertainty analysis approach to use under what circumstances (i.e. watershed, data and modeling objective)?
5. Allow for the description of a priori uncertainty. Question: How can we define prior uncertainties regarding the data, model parameters and structure, etc.? What is their structure?

4 Conclusions and open questions

Years of research into appropriate methods for the identification of hydrological models under uncertainty have led from a process of attempting to identify some 'best' model, towards attempting to identify all models (or model structures) that are consistent with the observed system behavior and rank the retained models with respect to their performance. Any model identification procedure can thus be reduced to steps that address the following three basic 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.?

There are currently a wide variety of definitions and methods available that attempt to answer these three questions and 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.

A last aspect not yet receiving sufficient attention is the issue of realism in hydrological models (see, for example, discussions in Seibert and McDonnell 2002; Wagener 2003; Beven 2005). It is not clear how this issue can be properly handled in the context of regression and systems theory based approaches, but some way forward will have to be found if reliable predictions in ungauged basins are to be our ultimate objective.

Acknowledgements We particularly like to thank the editors for their invitation to contribute this discussion paper and the three reviewers for very constructive criticism that helped to improve the quality of the paper and clarify important points. Partial support for the authors 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.

References

Beck MB (1987) Water quality modelling: a review of the analysis of uncertainty. Water Resour Res 23(8):1393–1442

Beven KJ (1993) Prophecy, reality and uncertainty in distributed hydrologic modelling. Adv Water Resour 16:41–51

Beven KJ (2001) Rainfall-runoff modelling—the primer. Wiley, Chichester

Beven KJ (2005) A manifesto for the equifinality thesis. J Hydrol (in press)

Beven KJ, Binley AM (1992) The future of distributed models: model calibration and uncertainty in prediction. Hydrol Process 6:279–298

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

Binley AM, Beven KJ (1991) Physically-based modelling of catchment hydrology: a likelihood approach to reducing predictive uncertainty. In: Farmer DR (eds) Computer modelling in the environmental sciences. Clarendon Press, Oxford, pp 75–88

Boyle DP, Gupta HV, Sorooshian S (2000) Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods. Water Resour Res 36(12):3663–3674

Burnash RJC (1995) The NWS river forecast system—catchment modelling. In: Singh VP (eds) Computer models of watershed hydrology. Water Resources Publications, Colorado, pp 311–366

Butts MB, Payne JT, Kristensen M, Madsen M (2004) An evaluation of the impact of model structure and complexity on hydrological modelling uncertainty for streamflow prediction. J Hydrol 298:242–266

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. Water Resour Res 22:199–210

Dawdy DR, O'Donnell T (1965) Mathematical models of catchment behaviour. J Hydraul Div. ASCE 91(HY4):123–137

Duan Q, Gupta VK, Sorooshian S (1992) Effective and efficient global optimisation for conceptual rainfall-runoff models. Water Resour Res 28:1015–1031

Freer J, Beven KJ, Ambroise B (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: an example of the GLUE approach. Water Resour Res 32(7):2161–2173

Freer J, McMillan H, McDonnell JJ, Beven KJ (2004) Constraining Dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. J Hydrol 291:254–277

Fukomori I, Raghunath R, Fy L-L, Chao Y (1999) Assimilation of TOPEX/Poseidon altimeter data in a global ocean circulation model: how good are the results? J Geophys Res 104(25):647–665

Gooseff MN, Bencala KE, Scott DT, Runkel R, McKnight DM (2005) Sensitivity analysis of conservative and reactive stream transient storage models applied to field data from multiple-reach experiments. Adv Water Resour 28:479–492

Grayson R, Blöschl G (eds) (2001) Spatial patterns in catchment hydrology. Cambridge University Press, Cambridge

Gupta HV, Sorooshian S, Yapo PO (1998) Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour Res 34:751–763

Gupta HV, Beven KJ, Wagener T (2005) Model calibration and uncertainty estimation. In: Anderson, MG (ed) Encyclopedia of hydrologic sciences. Wiley, Chichester (in press)

Hall J, Anderson M (2002) Handling uncertainty in extreme or unrepeatable hydrological processes—the need for an alternative paradigm. Hydrol Process 16:1867–1870

Hoeting JA, Madigan D, Raftery AE, Volinskiy CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14(4):382–417

Hooper RP, Stone A, Christophersen N, De Grosbois E, Seip HM (1988) Assessing the Birkenes model of stream acidification using a multisignal calibration methodology. Water Resour Res 24(8):1308–1316

Hornberger GM, Spear RC (1981) An approach to the preliminary analysis of environmental systems. J Environ Manage 12:7–18

Ibbitt RP (1970) Systematic parameter fitting for conceptual models of catchment hydrology. Unpublished Ph.D. Thesis, Imperial College of Science, Technology and Medicine, London

Ibbitt RP, O'Donnell T (1971) Fitting methods for conceptual catchment models. J Hydraul Div ASCE 97:1331–1342

Jakeman AJ, Hornberger GM (1993) How much complexity is warranted in a rainfall-runoff model?. Water Resour Res 29:2637–2649

Johnston PR, Pilgrim DH (1976) Parameter optimisation for watershed models. Water Resour Res 12(3):477–486

Karl TR, Knight RW, Plummer N (1995) Trends in high-frequency variability in the twentieth-century. Nature 377:217–220

Kavetski D, Franks SW, Kuczera G (2003) Confronting input uncertainty in environmental modelling. In: Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds) Calibration of watershed models. Water science and application series 6. American Geophysical Union, Washington, pp 49–68

Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. J R Stat Soc 63(3):425–464

Kitanidis PK, Bras RL (1980a) Real-time forecasting with a conceptual hydrologic model—Part I: analysis of uncertainty. Water Resour Res 16(6):1025–1033

Kitanidis PK, Bras RL (1980b) Real-time forecasting with a conceptual hydrologic model—Part II: applications and results. Water Resour Res 16(6):1034–1044

Kundzewicz ZW, Kaczmarek AW (2000) Coping with hydrological extremes. Water Intern 25(1):66–75

Kundzewicz ZW (2002) Non-structural flood protection and sustainability. Water Intern 27:3–13

Letcher RA, Jakeman AJ, Crok BFW (2004) Model development for integrated assessment of water allocation options. Water Resour Res 40:W05502. DOI 10.1029/2003WR002933

Madsen H (2000) Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. J Hydrol 235:276–288

Madsen H (2003) Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv Water Resour 26:205–216

Madsen H, Rosbjerg D, Damgård J, Hansen FS (2003) Data assimilation in the MIKE 11 flood forecasting system using Kalman filtering. In: Bloeschl G, Franks S, Kumagai M, Musiake K, Rosbjerg D (eds) Water resources systems—hydrological risk, management and development. IAHS Publ. no. 281, pp 75–81

McDonnell JJ (2003) Where does water go when it rains? Moving beyond the the variable source area concept of rainfall-runoff response. Hydrol Process 17:1869–1875

Melching CS (1995) Reliability estimation. In: Singh VP (ed) Computer models of watershed hydrology. Water Resources Publications, Colorado, pp 69–118

Misirli F (2003) Improving efficiency and effectiveness of Bayesian recursive parameter estimation for hydrologic models. Ph.D. Dissertation, Department of Hydrology and Water Resources, University of Arizona, USA

Neuman SP (2002) Accounting for conceptual model uncertainty via maximum likelihood Bayesian model averaging. Acta Universitatis Carolinae Geologica 46(2/3):529–534

Perrin C, Michel C, Andreassian V (2000) Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J Hydrol 242:275–301

Reichert P, Omlin M (1999) On the usefulness of overparameterized ecological models. Ecol. Model 95:289–299

Reichle RH, McLaughlin DB, Entekhabi D (2002) Hydrologic data assimilation with the ensemble kalman filter. Monthly Weather Rev 130(1):103–114. DOI 10.1175/1520-0493(2002)

Romanowicz R, Beven KJ, Tawn JA (1994) Evaluation of predictive uncertainty in nonlinear hydraulic models using a Bayesian approach. In: Barnett V, Turkman KF (eds). Statistics for the environment 2, water related issues. Wiley, New York

Seibert J, McDonnell JJ (2002) On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resour Res 38(11):1241. DOI 10.1029/2001WR000978

Singh VP, Frevert DK (eds) (2002a) Mathematical models of large watershed hydrology. Water Resources Publications, Colorado

Singh VP, Frevert DK (eds) (2002b) Mathematical models of small watershed hydrology. Water Resources Publications, Colorado

Sivapalan M, Takeuchi K, Franks SW, Gupta VK, Karambiri H, Lakshmi V, Liang X, McDonnell JJ, Mendiondo EM, O'Connell PE, Oki T, Pomeroy JW, Schertzer D, Uhlenbrook S, Zehe E (2003). IAHS decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrol Sci J 48(6):857–880

Sørensen JVT, Madsen H, Madsen H (2004) Efficient sequential techniques for the assimilation of tide gauge data in three dimensional modeling of the North Sea and Baltic Sea system. J Geophys Res 109. DOI 10.1029/2003JC002144

Sorooshian S, Duan Q, Gupta VK (1993) Calibration of rainfall-runoff models: application of global optimisation to the Sacramento soil moisture accounting model. Water Resour Res 29:1185–1194

Spear RC (1995) Large simulation models: calibration, uniqueness and goodness of fit. In: MODSIM'95: international congress on modelling and simulation, Modelling and Simulation Society of Australia, Newcastle

Thiemann M, Trosset MW, Gupta HV, Sorooshian S (2001) Recursive Bayesian inference for hydrologic models. Water Resour Res 37(10):2521–2535

Tsonis AA (1996) Widespread increases in low-frequency variability of precipitation over the last century. Nature 382:700–702

Tsonis AA (2004) Isglobal warming injecting randomness into the climate system? EOS Trans Am Geophys Union 85(38):361–364

Van Straten G, Keesman J (1991) Uncertainty propagation and speculation in projective forecasts of environmental change: a lake eutrophication example. J. Forecasting 10:163–190

Vrugt JA, Bouten W, Gupta HV, Sorooshian S (2002) Toward improved identifiability of hydrologic model parameters: the information content of experimental data. Water Resour Res 38(12), art. no. 1312

Vrugt JA, Gupta HV, Bastidas LA, Bouten W, Sorooshian S (2003a) Effective and efficient algorithm for multi-objective optimization of hydrologic models. Water Resour Res 39(8):1214. DOI 10.1029/2002WR001746

Vrugt JA, Gupta HV, Bouten W, Sorooshian S (2003b) A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour Res 39(8):1201. DOI 10.1029/2002WR001642

Vrugt JA, Diks CGH, Gupta HV, Bouten W, Verstraten JM (2005) Improved treatment of uncertainty in hydrologic modelling: Combining the strengths of global optimization and data assimilation. Water Resour Res (41). W01017 DOI 10.1029/2004WR003059

Wagener T (2003). Evaluation of catchment models. Hydrol Process 17:3375–3378

Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S (2001) A framework for development and application of hydrological models. Hydrol Earth Syst Sci 5(1):13–26

Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV (2003a) Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol Process 17(2):455–476

Wagener T, Wheater HS, Gupta HV (2003b) Identification and evaluation of watershed models. In: Duan Q, Sorooshian S, Gupta HV, Rousseau A, Turcotte R (eds) Advances in calibration of watershed models. AGU Monograph. American Geophysical Union, Washington, pp 29–47

Wheater HS, Jakeman AJ, Beven KJ (1993) Progress and directions in rainfall-runoff modelling. In: Jakeman AJ, Beck MB, McAleer MJ (eds) Modelling change in environmental systems. Wiley, Chichester, pp 101–132

Willmott C (1982) Some comments on the evaluation of model performance. Bull Am Meterol Soc 63:1309–1313

Yang J, Haan C (1991) Multiobjective parameter estimation for hydrologic models—weighting of errors. Trans ASAE 34(1):135–141

Young PC (1983) The validity and credibility of models for badly-defined systems. In: Beck MB, Van Straten G (eds). Uncertainty and forecasting of water quality. Springer, Berlin Heidelberg New York, pp 69–98

Young PC (1992) Parallel processes in hydrology and water quality: a unified time-series approach. J Inst Water Eng Sci 6(5):598–612

Young PC (1998) Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. Env Mod Software 13(2):105–122

Young PC (2001) Data-based mechanistic modeling and validation of rainfall-flow processes. In: Anderson MG, Bates PG (eds) Model validation. Wiley, Chichester, pp 117–161

Young PC (2002) Advances in real-time flood forecasting. Philos Trans R Soc Phys Eng Sci 360:1433–1450

Young PC, Beven KJ (1994) Data-based mechanistic (DBM) modelling and the rainfall-flow nonlinearity. Environmetrics 5:335–363

Young PC, Parkinson SD, Lees MJ (1996) Simplicity out of complexity in environmental systems: Occam's razor revisited. J Appl Stat 23(2/3):165–210

Young PC, Jakeman AJ, Post D (1997) Recent advances in the data-based modelling and analysis of hydrological systems. Water Sci Technol 36:99–116