# 142: Model Calibration and Uncertainty Estimation

**HOSHIN V GUPTA[1], KEITH J BEVEN[2] AND THORSTEN WAGENER[1]**

[1] *Department of Hydrology and Water Resources, The University of Arizona, Tucson, AZ, US*
[2] *Department of Environmental Sciences, Lancaster University, Lancaster, UK*

*All rainfall-runoff models are, by definition, simplifications of the real-world system under investigation. The model components are aggregated descriptions of real-world hydrologic processes. One consequence of this is that the model parameters often do not represent directly measurable entities, but must be estimated using measurements of the system response through a process known as model calibration. The objective of this calibration process is to obtain a model with the following characteristics: (i) the input-state-output behavior of the model is consistent with the measurements of catchment behavior, (ii) the model predictions are accurate (i.e. they have negligible bias) and precise (i.e. the prediction uncertainty is relatively small), and (iii) the model structure and behavior are consistent with current hydrologic understanding of reality. This article describes the historic development leading to current views on model calibration, and the algorithms and techniques that have been developed for estimating parameters, thereby enabling the model to mimic the behavior of the hydrologic system. Manual techniques as well as automatic algorithms are addressed. The automatic approaches range from purely random techniques, to local and global search algorithms. An overview of multiobjective and recursive algorithms is also presented. Although it would be desirable to reduce the total output prediction error to zero (i.e. the difference between observed and simulated system behavior) this is generally impossible owing to the unavoidable uncertainties inherent in any rainfall-runoff modeling procedure. These uncertainties stem mainly from the inability of calibration procedures to uniquely identify a single optimal parameter set, from measurement errors associated with the system input and output, and from model structural errors arising from the aggregation of real-world processes into a mathematical model. Some commonly used approaches to estimate these uncertainties and their impacts on the model predictions are discussed. The article ends with a brief discussion about the current status of calibration and how well we are able to represent the effects of uncertainty in the modeling process, and some potential directions.*



## THE NATURE OF RAINFALL-RUNOFF MODELS

The hydrology of any catchment involves complex interactions driven by a number of spatially distributed and highly interrelated water, energy, and vegetation processes. Any computer-based model intended to represent the behavior of a catchment must, therefore, conceptualize this reality using relatively simple mathematical equations that involve parameters to be specified for any particular application. Two characteristics of the modeling process are relevant to our discussion. First, all rainfall-runoff (RR) models, regardless of how spatially explicit, are to some degree lumped, so that their equations and parameters describe the processes as aggregated in space and time. As a consequence, the model parameters are typically not directly measurable, and have to be specified through an indirect process of parameter estimation. This process of parameter estimation is often called model *calibration* if values of parameters are adjusted to fit some observations made on the system of interest. Rainfall-runoff models usually contain several such "conceptual" parameters. While many of these parameters cannot be assumed to have direct physical (measurable) interpretations, they are often assumed

Encyclopedia of Hydrological Sciences. Edited by M. Anderson.

to have physical relevance, insofar as they are related to inherent and invariant properties of the hydrologic system. Second, the structure of the RR model is generally specified prior to any attempt to model the catchment being undertaken (Wheater *et al.*, 1993). While this specification is usually based on observed characteristics of the catchment, other factors that play a major role include the availability of data, modeling goal, and a variety of subjective considerations including personal preference and experience. This article will not address the issue of model structure specification, but will focus on the difficulties of model calibration, assuming that a suitably representative and acceptably accurate model structure has already been selected. The article reviews the historical development leading to current views of model calibration, and discusses the estimation and propagation of uncertainties in RR modeling.

## MODEL CALIBRATION

Calibration is a process in which parameter adjustments are made so as to match (as closely as possible) the dynamic behavior of the RR model to the observed behavior of the catchment (Figure 1). The process therefore requires measurements of catchment behavior, usually in terms of the inputs (rainfall) and the outputs (e.g. streamflow at the catchment outlet). Because, the outputs of RR models are usually related to the parameters in a nonlinear way, explicit linear-regression-type solutions are generally not possible, and some degree of directed iterative guesswork is required to arrive at a suitable solution (estimates for the parameters). Necessary conditions for an RR model to be "well-calibrated" are that it has (at least) the following three characteristics: (i) the input-state-output behavior of the model is consistent with the measurements of catchment behavior, (ii) the model predictions are accurate (i.e. they have negligible bias) and precise (i.e. the prediction uncertainty is relatively small), and (iii) the model structure and behavior are consistent with a current hydrologic understanding of reality. Note that, for the second requirement to be met, some method for estimating and reporting model prediction uncertainty must be included (Figure 2). Further, the third requirement is critical if the model is to be used to estimate the effects of perturbations to the structure of the real system (e.g. land-use changes).

It is important to stress that the process of model identification should not be understood as that of simply finding a model structure and parameter set that "fits the model to the data". It is actually a process of progressive model identification in which the initial (large) uncertainty in our knowledge of what constitutes a good model structure and good parameter estimates is sequentially reduced while constraining the model to be structurally and functionally (behaviorally) consistent with the available qualitative (descriptive) and quantitative (numerical) information about the catchment. Because, as mentioned before, any selected model will be (at best) a structural and functional approximation of the true (unknown) watershed structure and function, the calibrated estimates of the parameters and the resulting predictions will always contain some remaining uncertainty. These and other uncertainties will also lead to the result that our model will generally not be able to fit the data perfectly, that is, we will not be able to perfectly track the observed system behavior with our model.

Early methods for the calibration of RR models were based on manual, so-called "trial-and-error" procedures. The manual calibration process can be considered to have three levels (Boyle *et al.*, 2000). In Level Zero, the initial uncertainty of the estimates is defined by selecting feasible ranges for each parameter, using estimates from similar

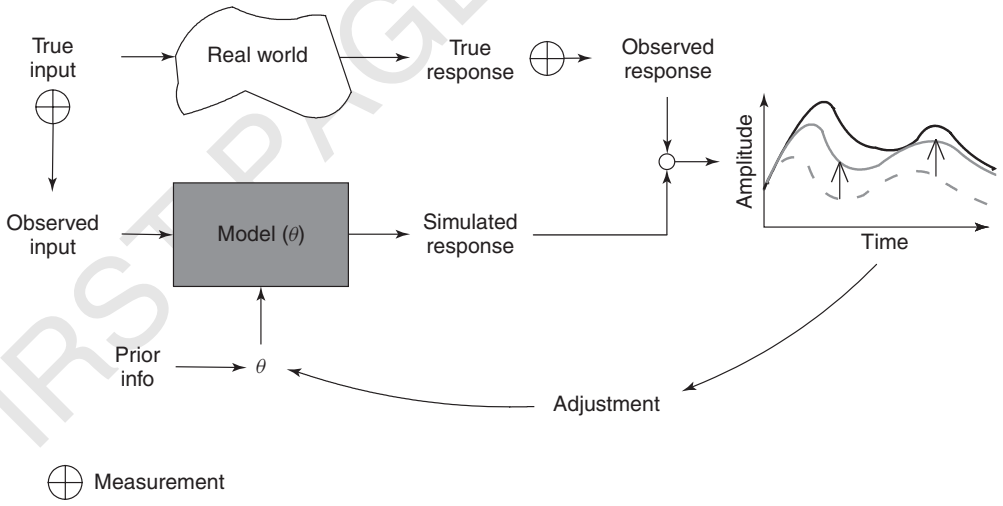


**Figure 1** Strategy for model calibration. The model parameter set is represented by $\theta$

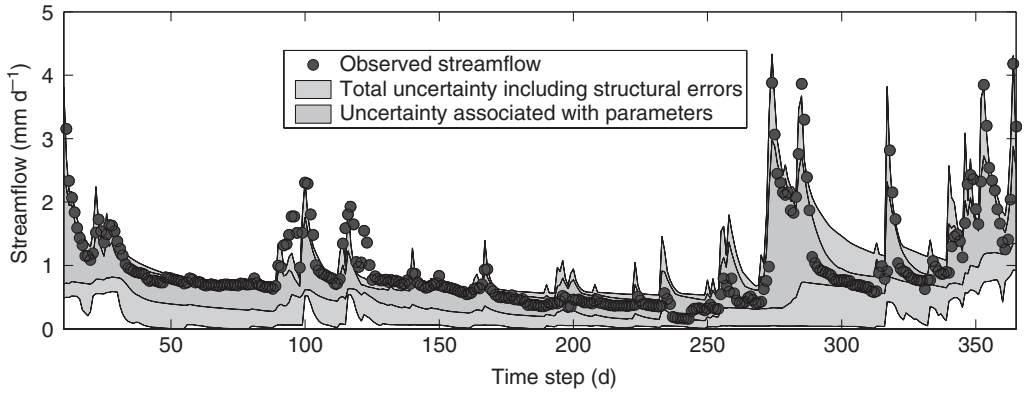


**Figure 2** Probabilistic streamflow prediction. Flow is shown in transformed space

catchments, look-up tables, maps, and databases. In Level One, the hydrologist attempts to reduce the size of the parameter uncertainty by adjusting one parameter at a time to try and match the particular segments of the catchment input-output (IO) response to which those parameters are most sensitive. Parameter interaction is usually ignored at this stage. Finally, in Level Two, the behavior of the entire hydrograph is examined, and the parameters further adjusted to allow for parameter interactions, while further reducing the distance between simulated response and the observed catchment behavior. This last stage is the most difficult, owing to the complex nature of the parameter interactions and the nonlinear nature of the model. A systematic approach to Level Two parameter estimation requires (i) a strategy to define (measure) the closeness between the observations and the model response, and (ii) a strategy to reduce the size of the feasible parameter space.

While the manual approach to model calibration (as described above) is based on subjective judgment and expertise, a trained and experienced hydrologist can often obtain excellent results, so that the model response generates a realistic simulation of the response of the catchment. However, the process can be very time consuming, and because it involves subjective decisions by the modeler, requires considerable training and practice. Further, the knowledge and skills so obtained are not easily transferred from one person to another. These limitations have led to interest in methods for model calibration that can be carried out automatically using the speed and power of a digital computer.

The goal of the automatic calibration approach is to use a computer to perform the difficult Level Two stage of the three-level strategy outlined above. Level zero is still performed manually to provide a crude description of the feasible parameter space, and the Level One stage is generally ignored (see Wagener *et al.*, 2003a, for a discussion of this problem). The potential advantages of a computer-based approach are not difficult to enumerate – properly designed computer algorithms can be fast and objective, while handling more complex problems (e.g. stronger nonlinearities and larger numbers of parameters). The closeness between the simulated and observed responses is typically measured by one (sometimes two or more) mathematical measures (called *objective functions*; OFs) and the parameters are adjusted by an iterative search procedure (called an *optimization algorithm*) towards the optimal value(s) of the OF(s).

An objective function (OF) is a numerical summary of the magnitude of the residuals, that is, the difference between the observed (measured) and the simulated system response (typically the streamflow hydrograph). The goal of calibration is usually to minimize (or maximize depending on definition) the value of this OF. The residuals are calculated as follows,

$$e(\theta) = y_t{}^{\text{obs}} - y_t(\theta) \tag{1}$$

where $y_t{}^{\text{obs}}$ is the observed response at time $t$, $y_t(\theta)$ is the simulated response, and $\theta$ is a vector of parameter values. The residuals are usually aggregated using a prespecified function,

$$F(\theta) = F\{e_t(\theta), \quad t = 1, 2, \ldots, n\} \tag{2}$$

The most commonly applied OFs are of the Weighted Least Squares (WLS) type (e.g. Sorooshian and Gupta, 1995), derived from regression theory,

$$F(\theta) = \sum_{t=1}^{n} w_t \cdot [y_t{}^{\text{obs}} - y_t(\theta)]^2 \tag{3}$$

where $w_t$ is the weight at time $t$, and $n$ is the number of data points analyzed. In absence of additional information, the weights are commonly set to the value 1.0 for all-time steps. This leads to the Simple Least Squares (SLS) OF that yields unbiased parameter estimates when the

following assumptions regarding the residual distribution are valid: (i) the residuals are independent and identically distributed (i.i.d.), (ii) the distribution shows a homogeneous variance, and (iii)] the residual distribution follows a normal distribution with zero mean (e.g. Gershenfeld, 1999). Because these assumptions are often violated in RR modeling, several researchers have tested the use of alternative error models, and therefore alternative OFs (e.g. Sorooshian and Dracup, 1980; Sorooshian *et al.*, 1983; Kavetski *et al.*, 2003). Examples include the Heteroscedastic Maximum Likelihood Estimator (HMLE, Sorooshian and Dracup, 1980; Sorooshian *et al.*, 1983), which considers a heteroscedastic variance in the system response measurements, and the Bayesian Total Error Analysis (BATEA, Kavetski *et al.*, 2003), which also considers errors in the input data. A more general form of OF, which can be adjusted to be consistent with different error models, was introduced to hydrologic modeling by •Thiemann *et al.* (2001), and is based on an exponential power density (Box and Tiao, 1973).

Research into automatic methods (by the hydrologic community) began in the 1960s and 1970s. At that time, it was assumed that the problem of model calibration is similar to that of estimating the parameters in a nonlinear statistical regression. As mentioned above, the OF was typically selected to be some kind of $l$-norm, such as the weighted sum of squared errors ($l = 2$) shown in equation (3). The minimization of the OF was typically carried out using a "Local Search" algorithm, beginning with an initial parameter guess and employing a preprogrammed iterative strategy to move the parameter search in a direction of local improvement.

In general, Local Search strategies belong to one of two classes – derivative-free (direct) methods and derivative-based (gradient) methods. Examples of popular direct methods include the Downhill Simplex (Nelder and Mead, 1965), the Pattern Search (Hooke and Jeeves, 1961), and the Rotating Directions (Rosenbrock, 1960) algorithms. Gradient methods are potentially more powerful than direct methods, because they use additional information – estimates of the local downhill direction based on the first and/or second derivative of the response surface with respect to the model parameters (Bard, 1974). Although Gupta and Sorooshian (1985) and Hendrickson *et al.* (1988) showed that analytical or numerical derivatives could be computed, even for complex conceptual RR models, Hendrickson *et al.* (1988) found that, in practice, gradient methods do not perform better than direct methods. The use of local search algorithms for model calibration has been tested extensively (Ibbitt, 1970; Johnston and Pilgrim, 1976; Pickup, 1977; among many others), with the general conclusion that such methods are unable to provide a reliable estimate of the globally optimal solution to the RR model minimization problem. Instead, the solution to a Local Search is typically strongly dependent on the accuracy of the initial guess.

Initial responses, during the 1980s, to the failure of automatic calibration methods based on Local Search, were to try and put the optimization problem onto a more rigorous statistical footing. Two (related) directions can be found in the literature, one based on the use of maximum likelihood theory (e.g. Sorooshian and Dracup, 1980) and the other based on the use of Bayesian theory (e.g. •Kuczera, 1983). However, neither of these directly addressed the causes of the inability to find the optimum for a selected OF.

Towards the end of the 1980s, with the advent of easier access to powerful digital computers, attention shifted to the testing of "Global Search" algorithms (e.g. Brazil and Krajewski, 1987). A characteristic of many Global Search algorithms is to begin with a number of initial guesses distributed throughout the feasible parameter space, and to evolve this population of guesses iteratively towards promising regions of the OF response surface. Global search algorithms that have been tested include Adaptive Random Sampling (Masri *et al.*, 1980; Brazil, 1988), Simulated Annealing (Kirkpatrick *et al.*, 1983; Thyer *et al.*, 1999), Controlled Random Search (Price, 1987; Klepper *et al.*, 1991) and the Genetic Algorithm (Holland, 1975; Goldberg, 1989; Wang, 1991). Duan *et al.* (1992) conducted a detailed analysis of the properties of the OF response surface associated with a typical RR model and found that:

- It contains more than one main region of attraction.
- It has many local optima within each region of attraction (Figure 3).
- It is rough with discontinuous derivatives.
- It is flat near the optimum with significantly different parameter sensitivities.
- Its shape includes long and curved ridges.

These insights were incorporated into the design of a novel Global Search procedure called the *Shuffled Complex Evolution* (SCE-UA) algorithm, which combines elements of a number of different strategies, including the Downhill Simplex, Controlled Random Search, and Competitive evolution with the newly proposed idea of "Complex Shuffling" (Duan *et al.*, 1992, 1993, 1994; Sorooshian *et al.*, 1993). Extensive testing of the SCE-UA method by numerous researchers has proven its effectiveness, ability to consistently find the global optimum), and efficiency (low probability of failure of any trial) in reliably finding the global solution, when a unique solution exists.

However, these studies have also demonstrated that numerous parameter sets usually exist, widely distributed throughout the feasible parameter space, which have very similar values for the selected OF. This poses difficulties for local or global optimization methods, referred

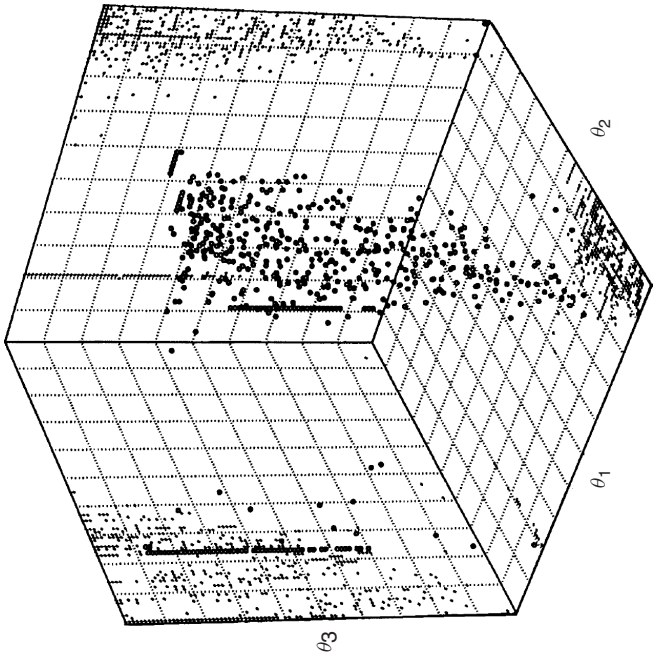


**Figure 3** Three-parameter ($\theta_i$) subspace of a simple conceptual catchment model (SIXPAR, Duan *et al.*, 1992), showing locations of multiple local optima

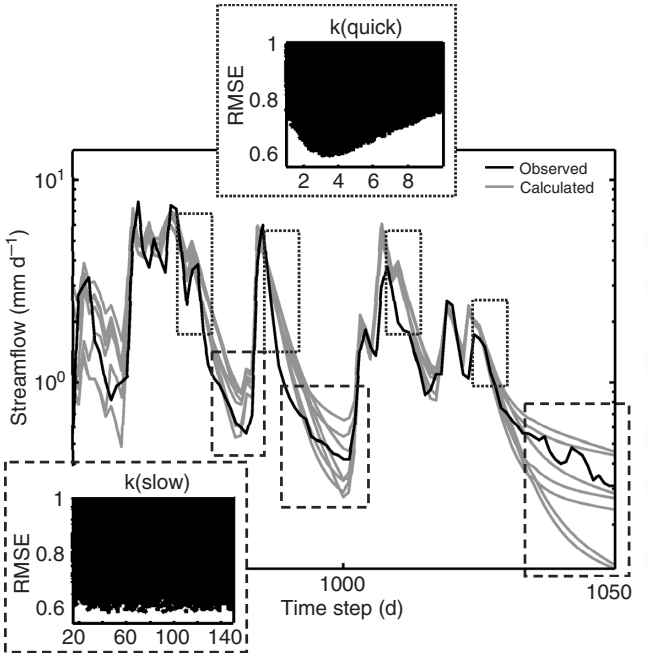


**Figure 4** Hundred days extract of six years of daily streamflow data. Observed flow in black, seven different model realizations in grey. Inlets show dotty plots for the linear reservoir residence times k(quick) and k(slow) versus their corresponding Root Mean Squared Error (RMSE) values. The model structure used consists of a Penman soil moisture accounting and a parallel routing component of linear reservoirs with fixed flow distribution

to in the optimization literature as problems of non-uniqueness, indeterminacy, or nonidentifiability. In reviewing this problem, Beven (1993) applied the term "equifinality" to describe the generic nature of finding multiple feasible models that make predictions consistent with any observations available for model calibration in any reasonably complex modeling problem.

The consequences of this progression of events have been some shifts in the underlying philosophy and perceived objectives of model calibration. Model calibration procedures were traditionally based on an attempt to find a "best" (most likely) estimate of the parameter values conditioned on the data, and a subsequent best (most likely) estimate/prediction of the catchment response. Efforts were concentrated on finding the most efficient techniques for doing so with a view to saving computer time in model calibration. The findings reported above have helped to make it clear that the inherent uncertainty (indeterminacy) in the estimated parameter values must be explicitly considered during both calibration and prediction. Approaches to do so are discussed in the section "Considering parameter uncertainty".

Another response has been to look for causes of the parameter indeterminacy and to design ways to address them. In particular, the traditional way to pose the model calibration problem relies on the specification of a single OF, which provides an aggregate measure of the mean distance between observed and calibrated hydrograph over the whole length of the data time-series, as the measure of performance. This action is now understood to result in considerable loss of important information that can be used to distinguish between competing parameter sets. For example, Figure 4 shows a number of model-simulated hydrographs (from different parameter sets) that produce identical OF values, but are clearly visually and behaviorally different (Wagener *et al.*, 2003c). Based on this, Gupta *et al.* (1998) suggest that a calibration approach with higher discriminative power is required and proposed use of a multiobjective methodology. This approach is described further in the section "Considering structural uncertainty".

A special case of parameter estimation arises in predicting the response of ungauged catchments. These are catchments for which observed records of the variable of interest, usually streamflow, are either too short or nonexistent. The main approaches to estimate the parameters of lumped rainfall-runoff models for ungauged catchments are through *physical reasoning* (e.g. Koren *et al.*, 2003), *statistical analysis* (e.g. Jakeman *et al.*, 1992; Wagener *et al.*, 2003b), or a mixture of both. Physical reasoning in this context means that parameters are derived from catchment properties, either directly or through empirical equations. Koren *et al.* (2003) suggest that reasonable initial estimates for the parameters of the Sacramento soil moisture

accounting (SAC-SMA) model can be derived from data such as average field capacity or wilting point of the soil in a catchment. The main approach to ungauged modeling however is the derivation of statistical relationships between model parameters and catchment characteristics for a large number of gauged catchments. Typically, a single model structure is selected that is assumed suitable to represent all the available catchments. The parameters of this structure are derived through a calibration process for all the gauged catchments, and an attempt is then made to develop regression relationships between model parameters and catchment characteristics. It is hoped that these statistical relationships can then be used to predict the parameters in ungauged catchments. See Wagener *et al.* hsa140 (2003b) and **Chapter 144, Rainfall-Runoff Modeling of Ungauged Catchments, Volume 1** for a discussion on this type of approach.

It should be noted that, regardless of the chosen approach, it will generally be impossible to reduce the total output prediction error, that is, the difference between observed and simulated system behavior to zero. Even if we could be sure that we had the correct model equations, errors in the input data and observations used in calibration will result in a residual prediction error. Indeed, experience suggests that reducing the error during one system response mode, often leads to an increase in the error during another mode (Gupta *et al.* 1998). In the following sections we will discuss the components and characteristics of the error.

It should also be noted that streamflow is not the only catchment response that can be used in model calibration. There is a wide range of studies where alternative hydrological variables are used for this purpose. Examples are Q3 •Lamb *et al.* (1997) and Blazkova *et al.* (2002b) using distributed groundwater levels; Kuczera and Mroczkowski (1998) using groundwater level observations and stream salinity data; Franks *et al.* (1998) and Blazkova *et al.* (2002a) using information on saturated areas; and Seibert and McDonnell (2002) using "soft" data on the nature of catchment responses.

## ON THE SOURCES AND THE NATURE OF THE TOTAL ERROR IN RAINFALL-RUNOFF MODELING

There is a problem, in any modeling application, of trying to understand the origins of the error between model predictions of a variable and any observational data of the same variable. The difficulty arises because there are a variety of sources for the error but (at any given time) only one measure of the deviation or residual between prediction and observation (i.e. the "total error"). Thus, disaggregation of the error into its source components is difficult, particularly in cases common to hydrology where the model is nonlinear and different sources of error may interact to produce the measured deviation.

Obvious sources of error in the modeling process include the errors associated with the model inputs and boundary conditions, the errors associated with the model approximation of the real processes, and the errors associated with the observed (measured) variables. A less obvious source of error is when the variable predicted by a model is not the same quantity as that measured (even though they might be referred to by the same name) because of scale effects, nonlinearities or measurement technique problems. A soil moisture variable, for example, might be predicted as an average over a model grid element several 100 m in spatial extent and over a certain time step; the same variable might be measured at a point in space and time by a small gravimetric sample, or by time domain reflectrometry integrating over a few tens of cm, or by a cross-borehole radar or resistivity technique, integrating over several tens or hundreds of meters. Only the latter might be considered to approach the same variable as described by the model (but will also be associated with its own interpretation errors of the geophysical signal in obtaining a soil water estimate). Fortunately, in rainfall-runoff modeling, the predictions are most usually compared with the measured discharges at the outlet from a catchment area. The measured discharges may be considered to be essentially the same variable as that predicted by the model, although subject to measurement errors.

In general, no satisfactory approach to separate the sources of error that contribute to the total error has yet been proposed. In line with traditional statistical estimation theory, the total error is often treated as a single lumped additive variable of the form:

$$Q(t) = M(\Theta, t) + \varepsilon(t) \tag{4}$$

where $Q(t)$ is a measured variable, such as discharge, at time $t$; $M(\Theta,t)$ is the model prediction of that variable using parameter set $\Theta$; and $\varepsilon(t)$ is the total remaining error. Note that a multiplicative error form could also be used if appropriate by taking logs of the observation and the predicted variables. The additive form allows application of the full range of statistical estimation techniques to model calibration and uncertainty analysis, subject to the limitations of dealing with the nonlinearity of the model.

Implicit in this approach, however, is the assumption that the model structure is at least adequate, if not correct. In catchment RR modeling we cannot generally be sure of this. Further, it is generally necessary to make assumptions about the structure of the total errors $\varepsilon(t)$ – typical assumptions include normality of the underlying error distribution, constancy of variance and simplicity of the correlation structure. While such assumptions are convenient to the application of statistical theory, they have usually not been supported by the actual series of model residuals, which

may show variations in bias (nonstationarity), variance (heteroscedasticity), and correlation structures under different hydrologic conditions.

In the following sections we analyze the main contributors to model error and to the uncertainty in predictions of our current models in general. Approaches to deal with these are outlined and open research questions are discussed.

## CONSIDERING PARAMETER UNCERTAINTY

It was mentioned above that a large number of widely different parameter sets could, in many cases, yield practically identical results with respect to a particular OF. Three main responses to this problem of perceived equifinality can be found in the literature.

First, the finding of parameter indeterminacy can be interpreted as an indication that the chosen model structure is overly complex given the information about hydrologic behavior actually observable in the data. Following this interpretation, various researchers have tested and successfully applied simpler model structures such that the number of associated model parameters is only so large as to allow confidence in the results of the calibration (Wheater *et al.*, 1993; •Hornberger and Jakeman, 1993; Young *et al.*, 1996; Young, 2001; Wagener *et al.*, 2002, 2003c).

Second, the finding can be interpreted as supporting the need for set theoretic approaches, which assume that all plausible models should be retained unless and until evidence to the contrary becomes apparent. Many of these set theoretic approaches are related to the Regional Sensitivity Analysis (RSA; also sometimes called the *Hornberger-Spear-Young approach*) concept advanced by Spear and Hornberger (1980) that evaluates the sensitivity of the model output to changes in parameters without referring to a specific point in the parameter space (such as a most likely value for a parameter). These techniques commonly apply random sampling procedures to explore the feasible parameter space in search for plausible (behavioral) models. Examples of the set theoretic approach applied to RR modeling include the Generalized Likelihood Uncertainty Estimation (GLUE) technique of Beven and Binley (1992), the Dynamic Identifiability Analysis (DYNIA) approach of Wagener *et al.* (2003a), the PIMLI approach of Vrugt *et al.* (2002), the Monte Carlo set membership (MCSM) approach of van Straten and Keesman (1991), the explicit Bayesian approach of Kuczera and Mroczkowski (1998), the Bayesian Recursive Estimation (BARE) technique of Thiemann *et al.* (2001), and the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm of Vrugt *et al.* (2003b).

Third, the finding can be attributed to a failure to properly specify the (automatic) calibration problem in such a way as to properly exploit the information contained in the data. There are two (related) responses to this. On the one hand, a multicriteria approach can result in better exploitation of the information in the data, while the resulting optimal population solution (optimal in a multicriteria sense) defines a kind of parameter uncertainty attributable to model structural errors. This is discussed further in the section "Considering structural uncertainty". In addition to the improved use of information, recursive processing of the data can provide better extraction of the information in the data, because the temporal aggregation associated with batch processing of data is reduced. Examples of recursive algorithms that can be applied to RR models for the estimation of parameter uncertainty include the Kalman Filter and its extensions (e.g. Kitanidis and Bras, 1980a,b; Beck, 1987), the PIMLI approach (Vrugt *et al.*, 2002), the DYNIA approach (Wagener *et al.*, 2003a), the BaRE approach (Thiemann *et al.*, 2001) and the application of Transfer Functions (TF) with time-varying parameters identified using Instrumental Variable techniques (Young, 2001). The GLUE methodology can also be applied recursively with appropriate choice of the Likelihood criterion. Recursive approaches can also provide a method for checking violations of the underlying assumption that the parameters are constant – for example, both Wagener *et al.* (2003a) and Misirli (2003, BaRE2) reported that certain parameters of the models they tested displayed significant temporal variations, and suggested that this may be an indication of model structural error.

We now take a closer look at four approaches representative of the second and third responses to dealing with parameter uncertainty, that is, the use of set theoretic methods, GLUE and SCEM-UA, and the application of recursive approaches, DYNIA and BaRE. Common to these approaches is the selection or identification of a *set* (population) of models (different combinations of model structures and parameter values), and assignment of some relative degree of believability to each member of the set. That degree of believability is translated into interval estimates of the uncertainty (confidence) in model simulations/predictions. The approaches differ in the suite of assumptions underlying each technique, based on which the methods used to compute the relative degree of believability are derived. The idea is that, in principle, the sensitivity of the predictions and associated uncertainty to the underlying assumptions of the methods are testable and can be evaluated.

The GLUE methodology evolved out of early Monte Carlo studies of different realizations of parameter sets in rainfall-runoff modeling as a way of estimating the sensitivity of model predictions to different parameter values (see, e.g., Hornberger and Spear, 1981). The Hornberger-Spear-Young approach to sensitivity analysis involves the classification of many different parameter sets into those that are behavioral and those that are nonbehavioral in some predefined way. The parameter distributions for each

of these sets are then examined to check for significant differences between them.

The additional step in the GLUE methodology of Beven and Binley (1992) is to associate each of the behavioral simulations with a likelihood measure (as a way of quantifying model believability), to estimate the uncertainty associated with the model predictions as conditioned on some calibration data used in the likelihood value calculation. Models that perform well in the calibration period will be associated with a high likelihood weight in prediction, those that perform less well will carry a low likelihood weight. Those that are considered nonbehavioral will not be used in prediction. Such an approach allows for the equifinality of different parameter sets in fitting the available calibration data, it allows that different types of likelihood measure might be used in model evaluation, and it allows for the likelihood weights to be updated as new calibration data become available (see the review of applications by Beven and Freer, 2001). The steps in the GLUE methodology are as follows:

- Decide on a model structure or structures to be used.
- Sample multiple sets of values from prior ranges or distributions of the unknown parameters by Monte Carlo sampling, ensuring independence of each sample in the model space.
- Evaluate each model run by comparing observed and predicted variables.
- Calculate a likelihood measure or measures for those models considered behavioral.
- Rescale the cumulative likelihood weights over all behavioral models to unity.
- Use all the behavioral models in prediction, with the output variables being weighted by the associated rescaled likelihood weight to form a cumulative distribution of predictions, from which any desired prediction quantiles can be extracted.

The methodology depends on obtaining an adequate sample of behavioral models. In some model applications this may require many thousands or millions of simulations to adequately define the behavioral regions in the multidimensional model space. In most problems it remains quite difficult to define prior distributions for effective values of the different parameters, even in terms of simple means and covariances. Thus, most applications of GLUE define only a range for each parameter and sample uniformly within that range. For cases where the region of behavioral models is relatively small this will result in inefficient sampling and alternatively sampling strategies may be required. However, in many applications it has been found that behavioral models are scattered widely in the model space so that efficiency in sampling behavioral models is difficult to achieve. It is, of course, possible that no behavioral models will be found, particularly where models must satisfy multiple criteria to remain behavioral (see for example •Freer *et al.*, 2002) giving a good indication that there are problems either with the data set or with the model structure. A statistical estimation strategy might still find an optimal model in such a case, but would generally assign the deficiencies to a large "total error" component (unless other statistical inadequacy components had been added, as for example in Kennedy and O'Hagan, 2001). Q5

In fact, the GLUE methodology is general, in that statistical error assumptions and likelihood functions can be used where the assumptions are satisfied, but a much wider range of likelihood measures including fuzzy measures can also be used. The only requirements are that the relative likelihood value should increase monotonically with improving model performance and that nonbehavioral models should be given a likelihood of zero. These are much less stringent requirements than those of statistical theory and the GLUE approach does avoid the assumption that the model is correct (implicit where a statistical error model with zero bias is assumed in model calibration). As a result, unlike a statistical approach, GLUE does not attempt to estimate the probability of predicting an observation given the (optimal) model. Instead, it predicts the probability of a model prediction, conditioned on the ranges of model structures and parameter values considered, the period of calibration (or evaluation) data used and the likelihood measures used in model evaluation. The method therefore assumes that the error structures associated with a particular behavioral model parameter set will remain "similar" during any prediction period, so that the likelihood weights determined in calibration can be used to weight the predictions of any variable of interest. In this way, distributional assumptions are avoided, and nonlinear changes in predicted distributions of variables are allowed (as demonstrated in Freer *et al.*, 1996, where such changes between high and low discharges are shown to be consistent with expectations of system response). The emphasis is on the parameter set in obtaining a behavioral model, rather than on individual parameters and their covariation. The approach can be extended to including multiple model structures, provided that different models can be evaluated with respect to the same likelihood measures. Different applications of the GLUE methodology are described in Beven *et al.* (2000), Beven and Freer (2001), and Beven (2001).

Vrugt *et al.* (2003b) extended the SCE-UA algorithm (described above) to allow for both the estimation of the most likely parameter set, and also for its underlying posterior distribution. The authors replaced the Downhill Simplex method used for population evolution by Duan *et al.* (1992) with the Metropolis Hastings (MH) algorithm. By merging the strengths of the MH algorithm, controlled random search, competitive evolution, and complex shuffling, the SCEM-UA is designed to evolve to a stationary

posterior target distribution of the parameters. The stochastic nature of the MH annealing scheme avoids the tendency of the SCE-UA algorithm to collapse into a single region of attraction (i.e. the global minimum), while the information exchange (shuffling) between parallel sequences allows the search to be biased in favor of better regions of the solution space. Examination of the posterior parameter distribution allows the user to detect whether multiple and/or large regions of the parameter space continue to remain consistent with the current data, and this parameter uncertainty can be projected into the output space as uncertainties on the model predictions (in a manner similar to both BaRE and GLUE). A detailed description of this algorithm appears in hsa079 the article by Vrugt and Dane (**Chapter 84, Inverse Modeling of Soil Hydraulic Properties, Volume 1**).

The dynamic identifiability analysis (DYNIA) approach developed by Wagener *et al.* (2003a) is a recursive parameter identification approach, based on elements of the GLUE methodology, the popular Regional Sensitivity Analysis (RSA; Spear and Hornberger, 1980), aspects of wavelet analysis (e.g. Gershenfeld, 1999), and the use of Kalman filtering for hypothesis testing as applied by Beck (1987). Monte Carlo sampling based on a uniform prior distribution is used to examine the feasible parameter space. In contrast to the GLUE procedure, DYNIA uses only a window period rather than the full data period to calculate a measure of performance. This window is moved through the available time-series in a step-wise fashion, that is, at time step $t$ one considers the residuals between $t - n$ and $t + n$. The size of $n$ is selected depending on the length of time over which the parameter is influential, and on the quality of the data. The performance measure is then used to condition the marginal parameter distribution at that particular time step. A threshold is applied to consider only the best performing parameter sets (e.g. best 10%), which represent the peak of the parameter distributions. The shape of the resulting distribution is projected into the time-parameter space and variation of its shape in time can be visualized. This methodology can be applied to track the variation of parameter optima in time, to separate periods of information and noise, or to test whether model components (and therefore parameter values) represent those processes they are intended to represent (Wagener, 2003).

The Bayesian Recursive Estimation (BaRE) algorithm, developed by Thiemann *et al.* (2001) based on assumptions similar to those used in the SCEM-UA batch calibration algorithm, employs a recursive scheme for tracking the conditional probabilities associated with several competing parameter sets (models) in an on-line mode instead of searching for a single best solution in an off-line mode. The parameter probabilities are used to compute probabilistic predictions of the desired output variables (Figure 2). Probability updating, via Bayes theorem, facilitates the assimilation of new data as they become available. The BaRE algorithm belongs to a broad class of ensemble methods, which include the Ensemble Kalman Filter (EnKF, see e.g. Evensen, 1994; Madsen and Canizares, 1999; Reichle *et al.*, 2002) and which use multiple possible model realizations (possibly involving multiple parameter sets, model structures, error sequences, etc.). The main difference is that BaRE employs a full nonlinear updating procedure, while the EnKF uses a linear correlation updating rule. The initial BaRE algorithm suffered from several shortcomings (Beven and Young, 2003; Gupta *et al.*, 2003; Misirli, 2003), the most important being that the parameter distribution collapsed onto a single point owing to an insufficient sampling density. Misirli (2003) addressed this and other problems by developing BaRE2, an improved version of the original BaRE algorithm, introducing (among other things) a resampling procedure to ensure an appropriate sampling density in the high probability region. The development of BaRE has helped to stimulate some discussion about appropriate methods for handling various sources of uncertainty, including model structural uncertainty. Please see the comment and reply on this topic published recently in Water Resources Research (Beven and Young, 2003; Gupta *et al.*, 2003).

An alternative to Monte Carlo based approaches to the estimation and propagation of uncertainty are the "point" methods based on first-order analysis. Such techniques can be used to calculate the mean and variance of the predicted variable based on the mean and variance of uncertain inputs and parameters only. They do not require computer intensive Monte Carlo schemes to estimate the shape of the response surface that is then mapped into the output space. These techniques are thus particularly attractive for practical applications. An overview of such methods can be found in Melching (1995). They commonly apply a Taylor series expansion of the OF or the model output around a specific point in the parameter space, usually truncated after the first-order term, hence the term first-order analysis. In the Mean-value First-Order Second-Moment (MFOSM) method, the selected point is the mean value. Numerical and sometimes even analytical derivatives can be used (Melching, 1995) to calculate the expected value and the variance of the predicted variable (e.g. streamflow). An advantage of this approach is its relative simplicity and computational efficiency. The main weakness is *the* assumption that *a single linearization of the system performance function at the central values of the basic variables is representative of the statistical properties of system performance over the complete range of basic variables* (Melching, 1995). This is a difficult assumption to make in RR modeling where the system under investigation usually exhibits a strongly nonlinear behavior. The Advanced First-Order Second-Moment (AFOSM) improves on the MFOSM approach by using a "likely" point in the

parameter space, instead of the mean. Rosenblueth's point-estimation method (Rosenblueth, 1981) uses the mean and the covariance of the variables in a Taylor series expansion and does not require the calculation of derivatives, as do MFOSM and AFOSM. For applications of Rosenblueth's approach to hydrological models, see, for example, Rogers *et al.* (1985), Binley *et al.* (1991), Melching (1992), or McIntyre *et al.* (2002). Harr's point-estimation method (Harr, 1989) reduces the number of simulations required for Rosenblueth's method from $2^p$ to 2p, where p is the number of model parameters. See McIntyre *et al.* (2002) for details on this method and for an application in hydrology.

All of the point-estimation methods described above, are, however, limited by some assumptions made during their application, most importantly approximate linearity of the model. The first two moments of the predicted variable can sometimes be calculated accurately using these approaches if the nonlinearity of the model (and model structural error) and the uncertainty in the model parameters is not too large (Garen and Burges, 1981; Kuczera, 1988; Høybye, 1998).

## CONSIDERING STRUCTURAL UNCERTAINTY

RR models are, by definition, simplifications of reality that aggregate catchment processes into simpler representations. The process of defining a perceptual model of the catchment and translating this model into mathematical code depends on the imagination and hydrologic understanding of the modeler. To corroborate or reject a model as suitable for the anticipated purpose, a procedure of model testing and evaluation must be applied. The imperfect model representation arising from the aggregation process introduces a degree of uncertainty into the model predictions, which is difficult to quantify. However, some of the consequences of this uncertainty can be detected and even used for improvements in the model structure.

A major consequence of model structural imperfection is that the model is incapable of reproducing all aspects and portions of the hydrograph equally well with a single parameter set. Because the classical manual modeling/calibration approach seeks a single "best" parameter set, the hydrologist is forced to select a trade-off between the errors in fit to different parts of the hydrograph, thereby arriving at some suitable compromise parameter set that meets the needs and objectives of the modeling exercise. The manual approach typically depends on visual examination of the "local" fit between various segments of the simulated and observed hydrographs, while also checking to see that some selected "global" OFs take on values that are within acceptable distance of their "optimal" values. The goal is to find a parameter set that produces a realistic hydrograph shape while giving an acceptable level of overall (statistical) performance. The classical single OF automatic calibration approaches result, in essence, in an implicit (difficult to specify or control) aggregate weighting of different aspects of hydrograph fit, which, in practice, tends to produce simulations that are biased towards specific aspects of the observed hydrograph (e.g. high or low flows). To date, it has not become clear if the complex thought processes that lead to successful manual calibration could be encapsulated into a single OF. This has fueled the recent research on multicriteria approaches.

Gupta *et al.* (1998) argued that the calibration of RR models is inherently a multiobjective problem. Their multiobjective approach offers a way forward by emulating the ability of Manual-Expert calibration to employ a number of complementary ways of evaluating model performance, thereby compensating for various kinds of model and data errors, and extracting greater amounts of information from the data. The outcome is a set of models that are constrained (by the data) to be structurally and functionally consistent with available qualitative and quantitative information and which simulate, in an uncertain way, the observed behavior of the watershed. By maintaining the independence of the various performance criteria, and by performing a full multicriteria optimization, the entire set of Pareto optimal solutions is identified. Chankong and Haimes (1993) define the concept of a Pareto optimum as follows: "A solution is said to be Pareto optimal (also synonymously known in the literature as efficient, noninferior, and nondominated) if the value of any OF cannot be improved without degrading at least one of the other Ofs." In simple language, it is possible to divide the parameter space into 'good' and 'bad' solutions, but one cannot objectively favor one of the good solutions, since there is always another one that is better in a certain aspect, that is, with respect to another OF.

Yapo *et al.* (1998), and later Vrugt *et al.* (2003a), presented algorithms capable of solving, in a single optimization run, the multiobjective problem posed by this approach. Yapo *et al.* (1998) developed the multiobjective complex evolution (MOCOM-UA) algorithm that uses a population evolution strategy (similar to the one employed by the SCE-UA algorithm) to converge to the Pareto set. In brief, the MOCOM-UA method involves the initial selection of a "population" of $p$ points distributed randomly throughout the $n$-dimensional feasible parameter space $\Theta$. In the absence of prior information about the location of the Pareto optimum, a uniform sampling distribution is used. For each point, the multiobjective vector $E(\theta)$ is computed, and the population is ranked and sorted using a Pareto-ranking procedure suggested by •Goldberg (1998). Simplexes of $n + 1$ points are then selected from the population according to a robust rank-based selection method. The MOSIM procedure, a multiobjective extension of the Downhill Simplex method (Nelder and Mead, 1965), is used to evolve each simplex in a multiobjective improvement direction. Iterative application of the ranking and evolution procedures causes the entire population to converge towards the Pareto

optimum. The procedure terminates automatically when all points in the population become nondominated.

Vrugt *et al.* (2003a) replaced the Downhill Simplex approach in the MOCOM-UA algorithm with an efficient Markov Chain Monte Carlo sampler (similar to the one employed in the SCEM-UA algorithm), which additionally allows for the estimation of the underlying probability distributions with respect to the different objective functions. They also improved on some of the weaknesses of the original approach such as the tendency of the MOCOM-UA algorithm to cluster the Pareto solutions in the most compromise region among the objectives, and premature convergence of the algorithm for cases involving a multitude of parameters and highly correlated performance measures. The multiobjective implementation of the SCEM-UA algorithm is termed *MultiObjective SCEM-UA* (MOSCEM-UA).

Several researchers used the fact that different parameter sets are required to represent different response modes of the hydrologic system in a more structured manner (e.g. •Beck, 1985; Young, 2001; Wagener *et al.*, 2003a). Beck (1987) applied the extended Kalman filter (EKF) to recursively estimate optimum parameter values, and used the variation of these optima in time to detect structural inadequacies. He reported that the EKF was inadequate for this purpose owing to a lack of robustness and owing to restrictions imposed by the filter assumptions. Wagener *et al.* (2003a) used the DYNIA methodology described earlier for the same purpose and showed structural problems in a RR model having typical structural elements by tracking parameter variations in time.

Kennedy and O'Hagan (2001) on the other hand, focused on deriving a more complex error model. They include a model inadequacy function in their Bayesian calibration framework. The goal, similar to the approach by Sorooshian and Dracup (1980; see next section), is to produce an error series having the desirable properties of constant variance and independence in time and space, so that unbiased estimates of the various parameters and correction terms can be more easily estimated. If the model structure is at least approximately correct, the statistical approaches lead to a concentration on finding a probabilistic description for the optimal values of the model and error parameters. Where there are model structural errors on the other hand, the statistical error model will, to some extent, be required to compensate for those errors. It is therefore important to carry out postcalibration diagnostic tests to ensure that the model and error assumptions are sound.

Another approach, arising from the equifinality concept (where more than one model or model structure appear to provide acceptable representations of the available observations; Beven, 1993), suggests that complications arising owing to the presence of nonlinear model structural error make it difficult to properly apply a rigorous statistical estimation procedure. This approach rejects the idea that an "optimal" model exists and concentrates instead on the task of finding a *set* of models that are *behavioral* in the sense of being acceptably consistent with the observations (however acceptable might be defined), or, more importantly, rejecting all those models that can be shown to be nonbehavioral. As discussed earlier, this is the basis for the Generalized Likelihood Uncertainty Estimation methodology and other set theoretic approaches to model calibration, which are easily extended to consider multiple objectives and multiple model structures, at the expense of significant additional computer run time. Within this framework, as noted earlier, it is possible that all the models tried will be rejected where such consistency with the observations cannot be demonstrated (Freer *et al.*, 2002), leading to the serious reconsideration of model structure, input data, or calibration data that would be justified in such a case.

Other researchers have also explored the same premise, that is, the need to consider multiple model structures. Neuman (2002), for example, suggests that not allowing for different system conceptualizations can lead to statistical bias and an underestimation of uncertainty. He introduces an approach based on Bayesian model averaging (Hoeting *et al.*, 1999) to account for this problem, where each model is treated as if it were the correct structure in trying to maximize its contribution to the averaging process.

## CONSIDERING DATA UNCERTAINTY

Data used for RR modeling are measurements of the input and output fluxes of the hydrologic system, and sometimes of its states. The input is precipitation, usually as rainfall or snow, while output data are streamflow and potential evapotranspiration. The latter is sometimes replaced by measurements of temperature. State variables that are of potential use in rainfall-runoff modeling are, for example, measurements of groundwater levels or soil moisture content.

Measurement errors with respect to streamflow occur owing to underflow and bypassing of gauging structures, and rating curve inaccuracies, especially at very high and very low flows. Sorooshian and Dracup (1980) addressed the problem of heteroscedastic errors in streamflow measurements by deriving a likelihood estimator, which uses a power transformation to stabilize the variance. However, research with respect to data uncertainty and its effect on the predictions of RR models has focused mainly on errors in the precipitation. This focus is based on the assumption that the dominant source of error stems from poor knowledge of the rainfall input. One of the earliest examples in this respect is the work of •Crawford and Linsley (1966), who used a rain gauge scaling factor as a calibration parameter in the Stanford Watershed Model, with values up to 1.1 to account for wind or orographic influences. The use of such

a factor can still be found in some of today's modeling exercises. Kavetski *et al.* (2003), for example, included an adjustment coefficient for each rainstorm within a statistical estimation framework.

Much of the error in precipitation measurements is related to the inability of available gauging networks to properly capture the amount and variability of precipitation in space and time. Not surprisingly, Beven and Hornberger (1982) and Obled *et al.* (1994) found that a correct assessment of the global volume of rainfall input in a variable pattern is more important than a rainfall pattern (by itself) for simulating streamflow hydrographs. However, the variability of the rainfall pattern can exert a strong influence on the timing of the hydrograph peak at a downstream gauging station. Initial studies on precipitation error focused on how well networks of rain gages were capable of estimating the Q9 actual total rainfall, first using synthetic data (•Wei and Lar- Q10 son, 1971; Troutman, 1983; •Watts and Calver, 1993) and then real data (Krajewski *et al.*, 1991; Ogden and Julien, Q11 1993, •1994; Shah *et al.*, •1996a). Such studies commonly Q12 assume that the highest resolution data available are approximately representative of the real pattern of precipitation behavior. The use of radar rainfall estimates has now become increasingly common and has been tested either individually or in combination with gauged data (e.g. Smith *et al.*, 1996; Moore and Hall, 2000; Morin *et al.*, 2001; Carpenter *et al.*, 2001). Also under investigation is the use of satellite-based remotely sensed information for deriving precipitation estimates (Hsu *et al.*, 1997; Sorooshian *et al.*, Q13 2000, •2003), which has the potential of providing global estimates of precipitation. Satellite-based precipitation estimates enable some knowledge of rainfall over the relatively extensive ungaged portions of the world (including locations that are difficult to access or are blocked from radar coverage owing to topography), and are particularly useful for large (regional and continental) scale hydrologic studies. Both radar and satellite estimates of rainfall, however, are dependent on interpretative models of the recorded signals that also have parameters subject to calibration. It is also important to consider that errors in the precipitation data will also introduce a bias on the parameter estimates that in turn impacts the model predictions (e.g. Troutman, 1983; Andréassian *et al.*, 2001; Kavetski *et al.*, 2003).

The results of the above mentioned studies suggest that the importance of capturing the spatial variability of rainfall depends significantly on whether the catchments are Q14 infiltration- or saturation-excess dominated (e.g. •Ogden *et al.*, 1994, 2002; Koren *et al.*, 1999). Spatial variability of rainfall seems to be of particular importance for infiltration-excess dominated catchments. In such catchments, the location of runoff production typically shows a stronger correlation with the location of high rainfall intensity (Michaud and Sorooshian, 1994; Winchell *et al.*, 1998). On the other hand, other factors such as the topography of the catchment can have a stronger influence on the location of runoff production in saturation-excess dominated catchments. The spatial distribution of rainfall can also be of higher importance in cases where the catchment is dry (Shah *et al.*, 1996b). Smith *et al.* (1993) suggest that results related to the importance of estimating precipitation variability should be treated with some degree of caution at this early stage of research, and that it needs to be demonstrated that the sensitivity of the models used in the aforementioned studies is actually representative of the sensitivity of the real catchments.

## DISCUSSION AND CONCLUSIONS

This article began with the premise that all rainfall-runoff models are (at some level) lumped and conceptual representations of a real-world system. The main consequences of this premise are that the model structure is defined, prior to any modeling being undertaken, by the modeler's understanding of the natural system, and that estimates of the model parameters must be provided. Such parameter estimates are typically derived through a process of model calibration using observed system behavior. Three necessary conditions for a rainfall-runoff model to be considered as being properly calibrated are that: (i) the input-state-output behavior of the model is consistent with the measurements of catchment behavior, (ii) the model predictions are accurate (i.e. they have negligible bias) and precise (i.e. the prediction uncertainty is relatively small), and (iii) the model structure and behavior are consistent with the hydrologists understanding of reality. With respect to all three aspects, problems have been encountered that are still not satisfactorily solved to this day. Regarding the first point, it has been found that different parameter sets are required to simulate different behaviors of the natural system. This is usually taken to be an indication of model structural problems. Multiobjective approaches can be used to consider this problem, and recursive methods can be applied to more objectively track parameter variation in time. At this time, however, structured approaches to improve model representations are only available for certain simple types of models (e.g. linear). With respect to the second aspect, it is often found that the estimate of prediction uncertainty is relatively large and depends on the type of approach chosen to analyze it. To be blunt, there is currently no unifying framework that properly addresses uncertainty in hydrological modeling. Statistical approaches (such as SCEM and BaRE) require assumptions that are often difficult to justify; alternative approaches (such as GLUE and DYNIA) require subjective decisions about model evaluations that can also be difficult to justify. And finally, the issue of the realism of hydrologic models in current practice is receiving increasing attention. Many modeling approaches

have been based on the powerful mathematics of regression and systems theory with insufficient consideration for the conceptual nature of the model structure and parameters. Recent developments suggest that proper consideration of this issue is required if successful prediction of ungauged catchments or those undergoing land-use changes is to be achieved (Beven, 2002, 2004).

We remain confident, of course, that other strategies to understanding and dealing with the various sources of error, including those arising from model structural deficiencies, will emerge as increasing numbers of intelligent and energetic minds are brought to bear on the problem. If nothing else, history teaches us that the progress of science is inexorable, and that today's "truths" are all too often tomorrow's "mistakes"! However, the nature of hydrological systems is such that even if new measurement techniques become available in the future, uncertainty in hydrological prediction will not be eliminated. Thus, hydrology as science must learn to be realistic about the uncertainties that arise in the modeling process.

### Acknowledgments

Partial support for the first and third authors was provided by SAHRA under NSF-STC grant EAR-9876800, and the National Weather Service Office of Hydrology under grant number NOAA/NA06WH0144. The third author also received support from the German Academic Exchange Service (DAAD) under the auspices of its postdoctoral fellowship program. We thank Mary Black for her editing work. We also greatly appreciate the constructive criticism provided by George Kuczera, which resulted in considerable improvements to the manuscript.

## SOFTWARE LINKS TEXTBOX

Hydrologic software, including many of the above-described algorithms (including SCE-UA, SCEM-UA, BaRE, etc.), is available for noncommercial use at **http://www.sahra.arizona.edu/software.html** (Wagener *et al.*, 2004). Demonstration GLUE software can be downloaded from **http://www.es.lancs.ac.uk/glue.html.** The Monte Carlo Analysis Toolbox (MCAT) is a compilation of several techniques to analyze the parameter and output space including the corresponding uncertainties. Any dynamic mathematical model can be analyzed for which a Monte Carlo sampling or a population evolution procedure can be run. The Toolbox includes the DYNIA methodology, elements of the GLUE procedure and multiobjective plots as explained later. Copies of the MCAT can be obtained from **http://ewre-www.cv.ic.ac.uk/software.** Other popular optimization packages, not explicitly mentioned in the text, are NLFIT (developed by George Kuczera, **http://www.eng.newcastle.edu.au/~cegak/**) and PEST (developed by John Doherty, **http://www.sspa.com/pest/**).

## FURTHER READING

•Anderson M.G. and Burt T.P. (1985) Modelling Strategies. In *Hydrologic Forecasting*, Anderson M.G. and Burt T.P. (Eds.), John Wiley & Sons: Chichester, pp. 1–13. Q15

Beven K.J. (1989) Changing ideas in hydrology - the case of physically-based models. *Journal of Hydrology*, **105**, 157–172.

Beven K.J. (2001) *Rainfall-Runoff Modelling – The Primer*, John Wiley & Sons: Chichester. (This book provides both a primer for the novice and a detailed and practical description of techniques and difficulties demanded by more advanced users and developers.).

Boyle D.P., Gupta H.V., Sorooshian S., Koren V., Zhang Z. and Smith M. (2001) Towards improved streamflow forecasts: the value of semi-distributed modelling. *Water Resources Research*, **37**, 2739–2759.

Duan Q., Gupta H.V., Sorooshian S., Rousseau A.N. and Turcotte R. (2003) *Calibration of Watershed Models*, Water Science and Application Series, 6, American Geophysical Union: Washington. (This edited monograph is a collection of papers representing a state-of-the-art analysis of mathematical methods used in the identification of models for hydrologic forecasting, design, and water resources management. This book is suitable for scientists, researchers and students of watershed hydrology, practicing hydrologists, civil and environmental engineers, and water resource managers.).

Ehrgott M. (2000) *Multicriteria Optimization*, Springer-Verlag: Berlin.

Finnerty B.D., Smith M.B., Seo D.-J., Koren V. and Moglen G.E. (1997) Space-time sensitivity of the Sacramento model to radar-gage precipitation inputs. *Journal of Hydrology*, **203**, 21–38.

Gupta H.V., Sorooshian S. and Boyle D.P. (2001) Assimilation of data for the construction and calibration of watershed models. *Keynote Paper Presented at the International Workshop on Catchment Scale Hydrologic Modelling and Data Assimilation*, Wageningen.

Harlin J. (1991) Development of a process oriented calibration scheme for the HBV hydrologic model. *Nordic Hydrology*, **22**, 15–36.

Hornberger G.M., Beven K.J., Cosby B.J. and Sappington D.E. (1985) Shenandoah watershed study: calibration of the topography-based, variable contributing area hydrologic model to a small forested catchment. *Water Resources Research*, **21**, 1841–1850.

Johnston D., Smith M., Koren V. and Finnerty B. (1999) Comparing mean areal precipitation estimates from NEXRAD and rain gauge networks. *ASCE Journal of Hydrologic Engineering*, **4**(2), 117–124.

Kuczera G. and Williams B.J. (1992) Effect of rainfall errors on accuracy of design flood estimates. *Water Resources Research*, **28**(4), 1145–1153.

Legates D.R. and McCabe G.J. Jr (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, **35**, 233–241.

Madsen H. (2000) Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, **235**(3–4), 276–288.

Madsen H., Wilson G. and Ammentorp H.C. (2002) Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology*, **261**, 48–59.

Nash J.E. and Sutcliffe J.V. (1970) River flow forecasting through conceptual models, I, A discussion of principles. *Journal of Hydrology*, **10**, 282–290.

Pilgrim D.H. (1983) Some problems in transferring hydrologic relationships between small and large drainage basins and between regions. *Journal of Hydrology*, **65**, 49–72.

Piñol J., Beven K.J. and Freer J. (1997) Modelling the hydrologic response of Mediterranean catchments, Prades, Catalonia. The use of distributed models as aid to hypothesis formulation. *Hydrological Processes*, **11**, 1287–1306.

Popper K. (2000) *The Logic of Scientific Discovery*. First published 1959 by Hutchinson Education, Routledge.

Seibert J. (2000) Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, **4**, 215–224.

Singh V.P. (1997) Effect of spatial and temporal variability in rainfall and •watershed characteristics on stream flow hydrograph. *Hydrological Processes*, **1**, 1649–1669.

Smith M.B., Seo D.-J., Koren V.I., Reed S., Zhang Z., Duan Q.-Y., Cong S., Moreda F. and Anderson R. (2003) The distributed model intercomparison project (DMIP): an overview. *Journal of Hydrology*, DMIP Special Issue, •in press.

Sorooshian S. and Gupta V.K. (1985) The analysis of structural identifiability: theory and application to conceptual rainfall-runoff model's. *Water Resources Research*, **21**, 487–495.

Sorooshian S., Gao X., Hsu K., Maddox R.A., Hong Y., Imam B. and Gupta H.V. (2002) A diurnal variability of tropical rainfall retrieved from combined GOES and TRMM satellite information. *Journal of Climate*, **15**, 983–1001.

Uhlenbrock S., Seibert J., Leibundgut C. and Rohde A. (1999) Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal*, **44**, 779–797.

Wagener T., Wheater H.S. and Gupta H.V. (2003) *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*, Imperial College Press: London. (This monograph is an extensive treatment of the theory, development and application of conceptual lumped rainfall-runoff models to gauged and ungauged catchments. This book will be valuable for hydrologic researchers, graduate students, and everybody who applies lumped catchment-scale models in operational or research settings.).

Watts L.G. and Calver A. (1991) Effects of spatially-distributed rainfall on runoff for a conceptual catchment. *Nordic Hydrology*, **22**, 1–14.

Wheater H.S., Bishop K.H. and Beck M.B. (1986) The identification of conceptual hydrologic models for surface water acidification. *Hydrological Processes*, **1**, 89–109.

Yan J. and Haan C.T. (1991) Multiobjective parameter estimation for hydrologic models – weighting of errors. *Transaction of the American Society of Agricultural Engineers*, **34**(1), 135–141.

Young C.B., Bradley A.A., Krajewski W.F. and Kruger A. (2000) Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. *Journal of Hydrometeorology*, **1**, 241–254.

## REFERENCES

Andréassian V., Perrin C., Michel C., Usart-Sanchez I. and Lavabre J. (2001) Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology*, **250**(1–4), 206–223.

Bard Y. (1974) *Nonlinear Parameter Estimation*, Academic Press.

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

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

Beven K.J. (2001) *Rainfall-Runoff Modelling – The Primer*, John Wiley & Sons: Chichester.

Beven K.J. (2002) Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system. *Hydrological Processes*, **16**(2), 189–206.

Beven K.J. (2004) A manifesto for the equifinality thesis. *Journal of Hydrology*, •submitted.

Beven K.J. and Binley A.M. (1992) The future of distributed models: model calibration and uncertainty in prediction. *Hydrological Processes*, **6**, 279–298.

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

Beven K.J., Freer J., Hankin B. and Schulz K. (2000) The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems. In *Nonlinear and Nonstationary Signal Processing*, Fitzgerald W.J., Smith R.L., Walden A.T. and Young P.C. (Eds.), CUP: pp. 115–151.

Beven K.J. and Hornberger G.M. (1982) Assessing the effect of spatial pattern of precipitation in modelling stream flow hydrographs. *Water Resources Bulletin*, **18**, 823–829.

Beven K.J. and Young P.C. (2003) Comment on "Bayesian recursive parameter estimation for hydrologic models". By M. Thiemann, M. Trosset, H. Gupta, and S. Sorooshian. *Water Resources Research*, **39**(•5),.

Binley A.M., Beven K.J., Calver A. and Watts L.G. (1991) Changing responses in hydrology: assessing the uncertainty in physically based model predictions. *Water Resources Research*, **27**(6), 1253–1261.

Blazkova S., Beven K.J. and Kulasova A. (2002a) On constraining TOPMODEL hydrograph and simulations using partial saturated area information. *Hydrological Processes*, **16**, 441–458.

Blazkova S., Beven K., Tacheci P. and Kulasova A. (2002b) Testing the distributed water table predictions of TOPMODEL (allowing for uncertainty in model calibration): the death of TOPMODEL? *Water Resources Research*, **38**(•11), 10.1029/2001WR000912.

Box G.E.P. and Tiao G.C. (1973) *Bayesian Inference in Statistical Analysis*, Addison-Wesley-Longman: Reading.

Boyle D.P., Gupta H.V. and Sorooshian S. (2000) Towards improved calibration of hydrologic models: combining the

strengths of manual and automatic methods. *Water Resources Research*, **36**, 3663–3674.

Brazil L.E. (1988) *Multilevel Calibration Strategy for Complex Hydrologic Simulation Models*, Unpublished Ph.D. Thesis, Colorado State University, Fort Collins.

Brazil L.E. and Krajewski W.F. (1987) Optimisation of complex hydrologic models using random search methods. *Engineering Hydrology Proceedings, Hydraulics Division/ASCE/Williamsburg*, Virginia, pp. 726–731.

Carpenter T.M., Georgakakos K.P. and Spersflage J.A. (2001) On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use. *Journal of Hydrology*, **253**, 169–193.

Chankong V. and Haimes Y.Y. (1993) Multi-objective optimization: pareto optimality. In *Concise Encyclopaedia of Environmental Systems*, Young P.C. (Ed.) Pergamon Press: pp. 387–396.

Duan Q., Gupta V.K. and Sorooshian S. (1992) Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Resources Research*, **28**, 1015–1031.

Duan Q., Sorooshian S. and Gupta V.K. (1993) Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, **76**(3), 501–521.

Duan Q., Sorooshian S. and Gupta V.K. (1994) Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, **158**, 265–284.

Evensen G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10143–10162.

Franks S., Gineste P., Beven K.J. and Merot P. (1998) On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Research*, **34**, 787–797.

Freer J., Beven K.J. and Ambroise B. (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research*, **32**, 2161–2173.

Garen D.C. and Burges S.J. (1981) Approximate error bounds for simulated hydrographs. *Journal of the Hydraulics Division, American Society of Civil Engineers*, **107**(HY11), 1519–1534.

Gershenfeld N. (1999) *The Nature of Mathematical Modelling*, Cambridge University Press: Cambridge.

Goldberg D.E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.

Gupta V.K. and Sorooshian S. (1985) The relationship between data and the precision of parameter estimates of hydrologic models. *Journal of Hydrology*, **81**, 57–77.

Gupta H.V., Sorooshian S., Hogue T.S. and Boyle D.P. (2003) Advances in the automatic calibration of watershed models. In *Calibration of Watershed Models*, Water Science and Application Series, 6, Duan Q., Gupta H.V., Sorooshian S., Rousseau A.N. and Turcotte R. (Eds.), American Geophysical Union: Washington, pp. 9–28.

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

Gupta H.V., Thiemann M., Trosset M. and Sorooshian S. (2003) Reply to comment by K. Beven and P. Young on "Bayesian recursive parameter estimation for hydrologic models". *Water Resources Research*, **39**(●5),. Q21

Harr M.E. (1989) Probability estimates for multivariate analyses. *Applied Mathematical Modelling*, **13**, 313–318.

Hendrickson J.D., Sorooshian S. and Brazil L. (1988) Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. *Water Resources Research*, **24**(5), 691–700.

Hoeting J.A., Madigan D., Raftery A.E. and Volinskiy C.T. (1999) Bayesian model averaging: a tutorial. *Statistical Science*, **14**(4), 382–417.

Holland J.H. (1975) *Adaption in Natural and Artificial Systems*, University of Michigan Press: Ann Arbor.

Hooke R. and Jeeves T.A. (1961) Direct search solutions of numerical and statistical problems. *Journal of the Association for Computing Machinery*, **8**(2), 212–229.

Hornberger G.M. and Spear R.C. (1981) An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, **12**, 7–18.

Høybye J.A. (1998) *Uncertainty Modelling in Hydrological Applications*, Unpublished Ph.D. Dissertation, Department of Water Resources Engineering, University of Lund, Sweden.

Hsu K., Gao X., Sorooshian S. and Gupta H.V. (1997) Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology*, **36**(9), 1176–1190.

Ibbitt R.P. (1970) *Systematic Parameter Fitting for Conceptual Models of Catchment Hydrology*, Unpublished Ph.D. Thesis, Imperial College of Science, Technology and Medicine, London.

●Jakeman A.J. and Hornberger G.M. (1993) How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, **29**, 2637–2649. Q22

Jakeman A.J., Hornberger G.M., Littlewood I.G., Whitehead P.G., Harvey J.W. and Bencala K.E. (1992) A systematic approach to modelling the dynamic linkage of climate, physical catchment descriptors and hydrologic response components. *Mathematics and Computers in Simulation*, **33**, 359–366.

Johnston P.R. and Pilgrim D.H. (1976) Parameter optimisation for watershed models. *Water Resources Research*, **12**(3), 477–486.

Kavetski D., Franks S.W. and Kuczera G. (2003) Confronting input uncertainty in environmental modelling. In *Calibration of Watershed Models*, Water Science and Application Series, 6, Duan Q., Gupta H.V., Sorooshian S., Rousseau A.N. and Turcotte R. (Eds.), American Geophysical Union: Washington, pp. 49–68.

Kennedy M.C. and O'Hagan A. (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society*, **63**(3), 425–464.

Kirkpatrick S., Gelatt C.D. Jr and Vecchi M.P. (1983) Optimisation by simulated annealing. *Science*, **220**, 671–680.

Kitanidis P.K. and Bras R.L. (1980a) Real-time forecasting with a conceptual hydrologic model - part I: analysis of uncertainty. *Water Resources Research*, **16**(6), 1025–1033.

Kitanidis P.K. and Bras R.L. (1980b) Real-time forecasting with a conceptual hydrologic model - part II: applications and results. *Water Resources Research*, **16**(6), 1034–1044.

Klepper O., Scholten H. and van de Kamer J.P.G. (1991) Prediction uncertainty in an ecological model of the Oosterschelde Estuary. *Journal of Forecasting*, **10**, 191–209.

Koren V.I., Finnerty B.D., Schaake J.C., Smith M.B., Seo D.-J. and Duan Q.Y. (1999) Scale dependencies of hydrologic models to spatial variability of precipitation. *Journal of Hydrology*, **217**, 285–302.

Koren V.I., Smith M. and Duan Q. (2003) Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. In *Calibration of Watershed Models*, Water Science and Application Series, 6, Duan Q., Gupta H.V., Sorooshian S., Rousseau A.N. and Turcotte R. (Eds.), American Geophysical Union: Washington, pp. 239–254.

Krajewski W.F., Lakshim V., Georgakakos K.P. and Jain S.C. (1991) A Monte Carlo study of rainfall-sampling effect on a distributed catchment model. *Water Resources Research*, **27**(1), 119–128.

Kuczera G. (1988) On the validity of first-order prediction limits for conceptual hydrologic models. *Journal of Hydrology*, **103**, 229–247.

Kuczera G. and Mroczkowski M. (1998) Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resources Research*, **34**, 1481–1489.

Q23 •Lamb R., Beven K.J. and Myrabo S. (1998) Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Advances in Water Resources*, **22**, 305–317.

Madsen H. and Canizares R. (1999) Comparison of extended and ensemble Kalman filter for data assimilation in coastal area modelling. *International Journal of Numerical Methods Fluids*, **31**, 961–981.

Masri S.F., Bekey G.A. and Safford F.B. (1980) A global optimization algorithm using adaptive random search. *Applied Mathematics and Computation*, **7**, 353–375.

McIntyre N., Wheater H.S. and Lees M.J. (2002) Estimation and propagation of parametric uncertainty in environmental models. *Journal of Hydroinformatics*, **4**(3), 177–198.

Melching C.S. (1992) An improved first-order reliability approach for assessing uncertainties in hydrologic modelling. *Journal of Hydrology*, **132**, 157–177.

Melching C.S. (1995) Reliability estimation. In *Computer Models of Watershed Hydrology*, Singh V.P. (Ed.) Water Resources Publishers, pp. 69–118.

Michaud J.D. and Sorooshian S. (1994) Effect of rainfall-sampling errors on simulation of desert flash floods. *Water Resources Research*, **30**(10), 2765–2775.

Misirli F. (2003) *Improving Efficiency and Effectiveness of Bayesian Recursive Parameter Estimation for Hydrologic Models*, Ph.D. Dissertation, Department of Hydrology and Water Resources, The University of Arizona, Tucson.

Moore R.J. and Hall M.J. (Eds.) (2000) HYREX: the hydrology radar experiment. Special Issue of *Hydrology and Earth System Sciences*, **4**(4), 681.

Morin E., Enzel Y., Shamir U. and Garti R. (2001) The characteristic time scale for basin hydrologic response using radar data. *Journal of Hydrology*, **252**, 85–99.

Nelder J.A. and Mead R. (1965) A simplex method for function minimisation. *Computer Journal*, **7**, 308–313.

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

Obled C.H., Wendling J. and Beven K. (1994) The sensitivity of hydrologic models to spatial rainfall patterns: an evaluation using observed data. *Journal of Hydrology*, **159**, 305–333.

Ogden F.L. and Julien P.Y. (1993) Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales. *Water Resources Research*, **29**(8), 2589–2597.

Pickup G. (1977) Testing the efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models. *Hydrological Sciences Bulletin*, **12**(2), 257–274.

Price W.L. (1987) Global optimisation algorithms for a CAD workstation. *Journal of Optimization Theory and Applications*, **55**(1), 133–146.

Reichle R.H., McLaughlin D.B. and Entekhabi D. (2002) Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, **130**, 103–114.

Rogers C.C.M., Beven K.J., Morris E.M. and Anderson M.G. (1985) Sensitivity analysis, calibration and predictive uncertainty of the Institute of Hydrology Distributed Model. *Journal of Hydrology*, **81**, 179–191.

Rosenblueth E. (1981) Two-point estimates in probabilities. *Applied Mathematical Modelling*, **5**, 329–335.

Rosenbrock H.H. (1960) An automatic method for finding the greatest or least value of a function. *Computer Journal*, **3**, 175–184.

Seibert J. and McDonnell J.J. (2002) Multi-criteria calibration of conceptual runoff models – the quest for an improved dialog between modeler and experimentalist. In *Advances in Calibration of Watershed Models*, AGU Monograph Series, USA, Duan Q., Sorooshian S., Gupta H.V., Rousseau A. and Turcotte R. (•Eds.), pp. 301–315. Q24

Shah S.M.S., O'Connell P.E. and Hosking J.R.M. (1996a) Modelling the effects of spatial variability in rainfall on catchment response. 1. Formulation and calibration of a stochastic rainfall field model. *Journal of Hydrology*, **175**, 66–88.

Shah S.M.S., O'Connell P.E. and Hosking J.R.M. (1996b) Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *Journal of Hydrology*, **175**, 89–111.

Smith J.A., Seo D.-J., Baeck M.L. and Hudlow M.D. (1996) An intercomparison study of NEXRAD precipitation estimates. *Water Resources Research*, **19**(3), 791–810.

Sorooshian S. and Dracup J.A. (1980) Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resources Research*, **16**(2), 430–442.

Sorooshian S., Duan Q. and Gupta V.K. (1993) Calibration of rainfall-runoff models: application of global optimisation to the Sacramento soil moisture accounting model. *Water Resources Research*, **29**, 1185–1194.

Sorooshian S. and Gupta H.V. (1995) Model calibration. In *Computer Models of Watershed Hydrology*, Singh V.P. (Ed.) Water Resources Publishers: pp. 23–68.

Sorooshian S., Gupta V.K. and Fulton J.L. (1983) Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models – influence of calibration data

variability and length on model credibility. *Water Resources Research*, **19**(1), 251–259.

Sorooshian S., Hsu K.L., Gao X., Gupta H.V., Imam B. and Braithwaite D. (2000) Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, **81**(9), 2035–2046.

Spear R.C. and Hornberger G.M. (1980) Eutrophication in peel inlet, II, identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, **14**, 43–49.

Thyer M., Kuczera G. and Bates B.C. (1999) Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms. *Water Resources Research*, **35**(3), 767–773.

Troutman B.M. (1983) Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resources Research*, **19**(3), 947–964.

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

Vrugt J.A., Bouten W., Gupta H.V. and Sorooshian S. (2002) Toward improved identifiability of hydrologic model parameters: the information content of experimental data. *Water Resources Research*, **38**(12), 48-1–48-13 Art. No. 1312.

Vrugt J.A., Gupta H.V., Bastidas L.A., Bouten W. and Sorooshian S. (2003a) Effective and efficient algorithm for multi-objective optimization of hydrologic models. *Water Resources Research*, **39**(8), 1214, 10.1029/2002WR001746.

Vrugt J.A., Gupta H.V., Bouten W. and Sorooshian S. (2003b) A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, **39**(8), 1201, 10.1029/2002WR001642.

Wagener T. (2003) Evaluation of catchment models. *Hydrological Processes*, **17**(16), 3375–3378.

Wagener T., Gupta H.V., Carpenter C., James B., Vázquez R., Sorooshian S. and Shuttleworth J. (2004) A hydroarchive for the free exchange of hydrological software. *Hydrological Processes*, **18**(2), 389–391.

Wagener T., Lees M.J. and Wheater H.S. (2002) A toolkit for the development and application of parsimonious hydrologic models. In *Mathematical Models of Large Watershed Hydrology*, Vol. 1 Singh V.P. and Frevert D. (Eds.), Water Resources Publishers: pp. 87–136.

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

Wagener T., Wheater H.S. and Gupta H.V. (2003b) *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*, Imperial College Press: London.

Wagener T., Wheater H.S. and Gupta H.V. (2003c) Identification and evaluation of watershed models. In *Advances in Calibration of Watershed Models*, Duan Q., Sorooshian S., Gupta H.V., Rouseau A. and Turcotte R. (Eds.), AGU: Monograph, pp. 29–47.

Wang Q.J. (1991) The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resources Research*, **27**(9), 2467–2471.

Wheater H.S., Jakeman A.J. and Beven K.J. (1993) Progress and directions in rainfall-runoff modelling. In *Modelling Change in Environmental Systems*, Jakeman A.J., Beck M.B. and McAleer M.J. (Eds.), John Wiley & Sons: Chichester, pp. 101–132.

Winchell M., Gupta H.V. and Sorooshian S. (1998) On the simulation of infiltration- and saturation-excess runoff using radar-based rainfall estimates: effects of algorithm uncertainty and pixel aggregation. *Water Resources Research*, **34**(10), 2655–2670.

Yapo P.O., Gupta H.V. and Sorooshian S. (1998) Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, **204**, 83–97.

Young P.C. (2001) Data-based mechanistic modelling and validation of rainfall-flow processes. In *Model Validation*, Anderson M.G. and Bates P.G. (Eds.), John Wiley & Sons: Chichester, pp. 117–161.

Young P.C., Parkinson S. and Lees M.J. (1996) Simplicity out of complexity in environmental modelling: occam's razor revisited. *Journal of Applied Statistics*, **23**, 165–210.

**SOFTWARE TEMPLATE**

*SOFTWARE TITLE*

PEST

## SOFTWARE DESCRIPTION

: nonlinear parameter estimation package based on the Levenberg-Marquardt algorithm

## TYPICAL APPLICATIONS

inverse estimation of soil hydraulic parameters using a hydrologic model and observed data

## HYPERLINKS TO MORE DETAILED INFORMATION

**http://www.parameter-estimation.com/html/pest_overview.html**

PEST Software (1998). PEST: Model-Independent Parameter Estimation. Watermark Computing.

**Keywords:** rainfall-runoff modeling; model calibration; model identification; optimization; uncertainty estimation; parameter uncertainty; structural uncertainty; data uncertainty

**QUERIES TO BE ANSWERED BY AUTHOR (SEE MARGINAL MARKS Q..)**

**IMPORTANT NOTE: You may answer these queries by email. If you prefer, you may print out the PDF, and mark your corrections and answers directly on the proof at the relevant place. Do NOT mark your corrections on this query sheet. Please see the proofing instructions for information about how to return your corrections and query answers.**

Q1. This reference has not been listed in the reference list. Please provide the reference details.
Q2. This reference has not been listed in the reference list. Please provide the reference details.
Q3. 'Lamb *et al.* 1997' is not mentioned in the Reference list but the same is cited in the text, whereas 'Lamb *et al.* 1998' is mentioned in the Reference list but not cited in the text. Please confirm if the year should be either '1997' or '1998'.
Q4. Please clarify whether we can change "Hornberger and Jakeman, 1993" to "Jakeman and Hornberger, 1993" to match with the reference given in the reference list. Please confirm if this is fine.
Q5. This reference has not been listed in the reference list. Please provide the reference details.
Q6. This reference has not been listed in the reference list. Please provide the reference details.
Q7. This reference has not been listed in the reference list. Please provide the reference details.
Q8. This reference has not been listed in the reference list. Please provide the reference details.
Q9. This reference has not been listed in the reference list. Please provide the reference details.
Q10. This reference has not been listed in the reference list. Please provide the reference details.
Q11. This reference has not been listed in the reference list. Please provide the reference details.
Q12. We have changed "Shan *et al.*, 1996" to 1996a for the first occurence and 1996b for the second one. Please confirm if this is fine.
Q13. This reference has not been listed in the reference list. Please provide the reference details.
Q14. This reference has not been listed in the reference list. Please provide the reference details.
Q15. These references have been moved to Further Reading as they have not been cited in text. Please confirm if this is fine.
Q16. We have changed "waterghed" to "watershed" in this reference title. Please clarify if this is fine.
Q17. Please provide the volume number and page range for this reference.
Q18. Please provide the volume number and page range for this reference.
Q19. Please provide the page range for this reference.
Q20. Please provide the page range for this reference.
Q21. Please provide the page range for this reference.
Q22. 'Jakeman and Hornberger, 1993' has not been cited in the text please confirm whether we can move this reference to Further Reading
Q23. 'Lamb *et al.* 1998' has not been cited in the text please confirm whether we can move this reference to Further Reading.
Q24. Please provide the publisher's name for this reference.