Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox

Thorsten Wagener*, Joshua Kollat

Department of Civil and Environmental Engineering, The Pennsylvania State University, 226b Sackett Building, University Park, PA 16802, USA

Abstract

The detailed evaluation of mathematical models and the consideration of uncertainty in the modeling of hydrological and environmental systems are of increasing importance, and are sometimes even demanded by decision makers. At the same time, the growing complexity of models to represent real-world systems makes it more and more difficult to understand model behavior, sensitivities and uncertainties. The Monte Carlo Analysis Toolbox (MCAT) is a Matlab library of visual and numerical analysis tools for the evaluation of hydrological and environmental models. Input to the MCAT is the result of a Monte Carlo or population evolution based sampling of the parameter space of the model structure under investigation. The MCAT can be used off-line, i.e. it does not have to be connected to the evaluated model, and can thus be used for any model for which an appropriate sampling can be performed. The MCAT contains tools for the evaluation of performance, identifiability, sensitivity, predictive uncertainty and also allows for the testing of hypotheses with respect to the model structure used. In addition to research applications, the MCAT can be used as a teaching tool in courses that include the use of mathematical models.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Evaluation; Uncertainty analysis; Sensitivity analysis; Identifiability; Hypothesis testing; Model diagnostics; Matlab; Visualization

Software Availability

Name of software: Monte Carlo Analysis Toolbox (MCAT).
Developers: Thorsten Wagener, Matthew Lees, Joshua Kollat.
Contact Address: Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA 16802, USA.
Telephone: 1 814 865 5673.
Fax: 1 814 863 7304.
E-mail: thorsten@engr.psu.edu.
Year first available: 2000.
Hardware required: PC.
Software required: MATLAB.
Program language: MATLAB.
Program size: 0.4 MB.

Availability and cost: Available free of charge. Details in text. Source code is available from the corresponding author on request.

1. Introduction and motivation

Hydrological and environmental models are irreplaceable components in studies relating for example to flood and drought prediction, water resource assessment, climate and land use change impacts, or non-point source pollution analysis (e.g. Singh and Woolhiser, 2002; Letcher et al., 2007). The complexity of problems hydrologists are asked to investigate has grown over the years. The increasing complexity of the problem has in turn lead to an increase of complexity in the underlying conceptual model representation of the hydrological or environmental system under investigation, particularly when multiple variables of interest have to be considered simultaneously (e.g., the flow of energy between

* Corresponding author. Tel.: +1 814 865 5673; fax: +1 814 863 7304.
E-mail address: thorsten@engr.psu.edu (T. Wagener).

landscape and atmosphere, the flow of solutes through the watershed, as well as the flow of water). This increasing complexity is subsequently translated into the mathematical models that are used as representations of these conceptualizations. This often includes a large number of parameters that have to be estimated from data with limited information content (see for example the discussion by Beven and Freer, 2001). The resulting models are often high-dimensional and non-linear multi-output models integrating several sub-models that had previously been applied separately. This has negative implication with respect to the uncertainty in the identified parameter values and the identified model structure, which can for example preclude the successful regionalization of model parameters for the prediction in ungauged basins (Wagener and Wheater, 2006). In addition to the complexity problem, there has been a recent surge in attention given to methods for the treatment of model uncertainty. It is the experience of the authors that decision makers have begun to push for better quantification of the accuracy and precision of hydrological model predictions. In addition, interest has grown in methods for properly merging data with models and for reducing predictive uncertainty and scientists have begun to search for better ways to represent what is, and is not, well understood about the hydrological systems they study (Wagener and Gupta, 2005). Discussions on the uncertainty aspect in hydrological and environmental modeling can for example be found in Beven (2002), Van Asselt and Rotmans (1996), and Funtowicz and Ravetz (1990).

Available hydrological and environmental models differ widely in terms of their theoretical basis, ranging from physically based models which directly derive their governing equations from physical laws, to empirical models which derive both model structure (equations) and parameters from observed system behavior. Hydrological models also vary with respect to their spatial representation, ranging from lumped models which typically represent a region of interest as a single point, to spatially distributed models (Dingman, 2002). While these models differ with respect to their physical realism, at least some of their parameters and states represent effective values at the particular scale of model application (Wagener and Gupta, 2005). Available measurement scales are usually different from this scale and some parameter adjustments based on comparing observed and measured model outputs is required to yield reliable model results. As a result, a major constraint is the lack of identifiability of those parameter sets within a selected model structure that represent the given watershed best. Lack of identifiability is present when different combinations of parameters (e.g., Johnston and Pilgrim, 1976; Beven and Binley, 1992), and sometimes even different model structures (e.g., Uhlenbrook et al., 1999) yield similar results in terms of a defined performance measure or objective function.

In general terms, the modeler combines information regarding the intended model use and the characteristics of the system under study to develop a conceptual system representation that then has to be implemented in a mathematical model (Fig. 1). Issues to be considered are required spatial and
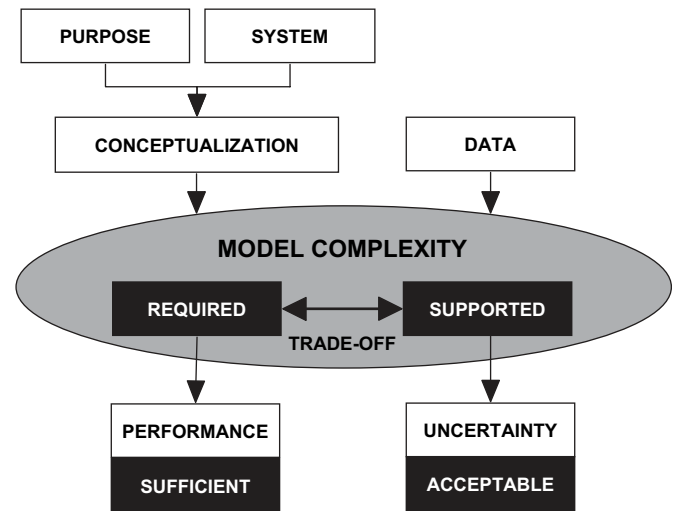

Fig. 1. General evaluation framework for model development.

temporal resolution of state and output variables for the specific model purpose, and dominant processes controlling the system behavior. Available observations (data) of the system input-response behavior can then be used to constrain the model behavior, estimate parameters, etc. Often, the modeler has to perform a complexity trade-off to achieve a balance between the required model performance while not introducing excessive uncertainty through parameters that cannot be estimated from limited information.

A problem of the trend of increasing model complexity is that the task of understanding how these models work, identifying their uncertainties and sensitivities, etc. becomes more and more difficult. Powerful numerical and visual evaluation tools are needed to help analyze model (state, parameter) and output spaces. Ultimately, these tools should serve as diagnostic tools that help to improve the mathematical model, and hopefully help us to learn about the underlying perceptual model, or at least show us where gaps in our knowledge are most severe and are most strongly affecting prediction uncertainty, and therefore help to guide field measurement campaigns. As a result, there are at least three things that one would like to explore using such a tool (Wagener, 2003):

– Performance, generally in a multi-objective sense.
– Uncertainty, e.g. with respect to the model parameters and the model output.
– Underlying hypotheses/assumptions, regarding the model structure chosen.

A general tool for this purpose should integrate a variety of approaches to model evaluation and be applicable to as wide a range of models as possible. One example of such a tool is the Monte Carlo Analysis Toolbox (MCAT). The MCAT is a collection of evaluation tools integrated through a Graphical User Interface (GUI) in the MATLAB programming environment. It is developed for the off-line evaluation of model results derived through Monte Carlo or population evolution

sampling strategies. The MCAT has been applied to research in rainfall-runoff modeling (Wagener et al., 2001, 2003, 2004; Lee et al., 2005), in-stream solute transport (Wagener et al., 2002), water quality modeling (Sincock et al., 2003), and for watershed-scale phosphorous transfer modeling (Smith and Wheater, 2004; Smith et al., 2005). Other toolboxes for uncertainty/sensitivity analysis of hydrological and environmental models are for example the Generalized Likelihood Uncertainty Estimation approach (Beven and Freer, 2001) and the SIMLBA toolbox (Saltelli et al., 2004). The links to these toolboxes are provided at the end of the paper.

The objective of this paper is not to provide a comprehensive analysis of the model study presented, but to introduce the MCAT using a modeling procedure with typical steps and components. The subsequent sections describe such a modeling procedure, introduce a typical model/data combination used as a case study, and provide examples of the MCAT components in use.

2. Modeling procedures

A wide variety of suggested modeling procedures have been presented in the literature (examples can be found in Dooge, 1972; Singh, 1988; Beven, 2000; Young, 2001; Jakeman et al., 2006, etc.). Here we use a relatively general framework and only include the most common steps. Such a framework would consist of at least four-steps as shown in Fig. 2:

2.1. Model structure selection/development

There is no single structured way of approaching the model selection or development stage. Model structures should be chosen based on relatively objective criteria including: (a) the system characteristics; (b) available data; and (c) the modeling objective. Subjective criteria such as the modeler's expertise and experience, or a preference within a certain organization might also have a strong influence on the final decision regarding model selection. The question of appropriate levels of model complexity has been discussed at length elsewhere and is not included in detail here (Jakeman and Hornberger, 1993; Grayson et al., 1992). Throughout the remainder of this paper, we will assume that a particular model structure has already been selected.

2.2. Model identification

Here we define model identification as the task of identifying a single parameter or a group of parameter sets (models) within a specific model structure as potential (or behavioral) representations of the system under study.

The narrower the space that these chosen models cover within the feasible model space, the more identifiable the model. A model is by definition a simplified representation of reality, which means that its parameters necessarily aggregate more complex and heterogeneous real-world characteristics in simpler mathematical form. These parameters and their associated states are therefore often 'conceptual' or 'effective'
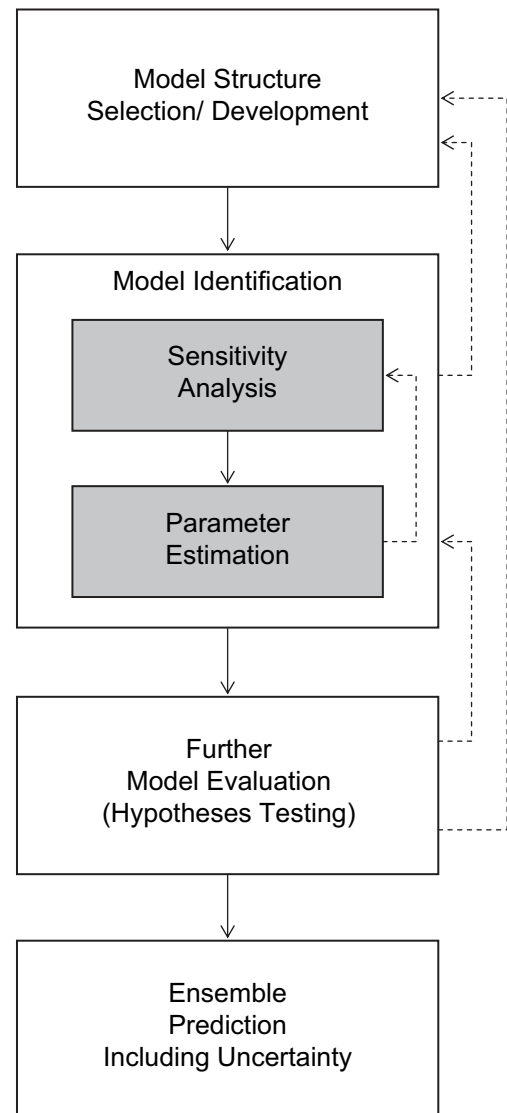

Fig. 2. Typical components of a generic modelling procedure.

values, rather than directly measurable entities. This means that the modeler has to compare the observed and simulated behavior of the system for different parameter sets and find those that reproduce the system behavior most closely. There has generally been a migration away from the identification of a best parameter set with respect to a particular objective function towards approaches to constrain the parameter (or model) space in a way that only 'behavioral' models are retained (Beven and Binley, 1992; Beven and Freer, 2001; Gupta et al., 2006; Wagener and Gupta, 2005). The definition of behavioral is dependent on the modeling objective, assumptions about statistical distributions of the errors, etc. (Beven and Freer, 2001; Beven, 2006), but usually relates to a certain behavior or a specific performance level in connection with the intended application (Wagener and McIntyre, 2005). The step of searching for suitable parameter sets for the commonly very non-linear models of hydrological and environmental systems is often performed using population

evolution or Monte Carlo (MC) type sampling approaches. Methods used for MC sampling range from those purely based on an a priori defined distribution or strategy to those making strong assumptions about the expected response surface shape. Some of the more commonly applied methods are: (1) uniform random sampling where all parameters are sampled from a uniform distribution, typically without consideration of parameter interaction (e.g. Beven and Freer, 2001; Wagener et al., 2003); (2) Latin Hypercube Sampling (LHS) or other stratified sampling approaches where the parameter space is somehow broken into hypercubes to ensure a better sampling spread (e.g. McKay et al., 1979; Press et al., 1992; Helton and Davis, 2002; 2003); (3) techniques based on Monte Carlo Markov Chains (MCMC) where assumptions (e.g. proposed distributions) are required to utilize the potential of the approach (examples include the Metropolis algorithm and the Shuffled Complex Evolution Metropolis algorithm) (e.g. Vrugt et al., 2003); or (4) other evolutionary type algorithms in which a population of initially randomly sampled parameter sets are systematically improved (examples include Genetic Algorithms (GA) and the Shuffled Complex Evolution (SCE-UA) algorithm) (e.g. Duan et al., 1993).

The parameter estimation stage is often combined with a sensitivity analysis. *Sensitivity analysis evaluates how sensitive the model response is to changes in the model parameters, state variables, or input data.* At this stage of the modeling process, insensitive parameters can then be fixed or otherwise eliminated to reduce the computational burden during any optimization or sampling step. Sensitivity analysis is of particular importance for high-dimensional models where computational constraints still limit the size of the parameter space that can be explored, particularly if estimates of parameter uncertainty are to be included. More advanced uses of sensitivity analysis for hypotheses testing are discussed below. A sensitivity analysis procedure consists of the two basic components: (i) a strategy to vary the model parameters, inputs, or states; and (ii) the definition of a (numerical) measure to estimate how the model response has changed based on varying one or more parameters, inputs, or states.

2.3. Further model evaluation

Further model evaluation in the context of this paper refers to secondary testing of additional aspects of model performance and general behavior beyond what is required for an initial identification of an appropriate model.

Model evaluation in the past has mainly focused on analyzing the performance of a model with respect to one or more numerical measures that aggregate the model residuals (i.e., the difference between observed and simulated hydrographs) in a single value (i.e., an objective function). A wide range of studies is available that evaluate the value of different numerical measures or objective functions (e.g. ASCE, 1993; Martinec and Rango, 1989; Diskin and Simon, 1977). More recently, multi-objective approaches have become popular to account for multi-output models and for structural limitations that result in different optimal parameter sets

depending on the chosen objective function (e.g. Gupta et al., 1998). This approach already is a hypothesis testing process since it evaluates the (common) underlying assumption that the chosen model structure can represent all response modes of the watershed behavior with a single (and usually time-invariant) parameter set. This approach can be taken further and the variation of parameter values in time can be estimated more generally to test this hypothesis (e.g. Beck, 1987; Wagener et al., 2003). Two questions that should be asked in this context include: (i) does a model parameter or a group of parameters represent the processes they are intended to represent (i.e., do they dominate the model response when this process dominates the system response); and (ii) are regions of behavioral parameters constant in time, or do they vary with different response modes of the system? The use of split-sample testing, i.e. a validation step in which the model is applied to a time-series not used for parameter estimation, could be part of this further model evaluation.

2.4. Ensemble prediction including uncertainty

The prediction step propagates the behavioral model population and potentially other uncertainties into the output space. The result is thus highly dependent on the definition of behavioral applied, e.g., in a multi-objective optimal sense (Gupta et al., 1998) or allowing for deviations from the optimal fit based on some subjective behavioral threshold (Beven and Binley, 1992). Research is ongoing to find approaches that explicitly allow for the consideration and propagation of multiple sources of uncertainty, e.g., uncertainty in the precipitation input (Kavetski et al., 2002).

3. Case study

The case study uses a typical data set and a typical lumped hydrologic model as they can be found in many published research studies as outlined below. This is in line with the paper's objective to present the utility of the tool, rather than an extensive case study with new scientific insight.

Ten years of daily streamflow, precipitation, and potential evapotranspiration data from the Leaf River Watershed located north of Fort Collins, Mississippi, USA (1950 km^2), have been used (Fig. 3). The Leaf River Watershed can be classified as humid and has been widely used for other hydrologic studies (e.g. Boyle et al., 2000; Wagener et al., 2001; Vrugt et al., 2003).

The model structure applied (Fig. 4) is a typical lumped parsimonious model structure often utilized in very similar form for research studies (e.g. Young, 1998; Jakeman and Hornberger, 1993; Croke and Jakeman, 2004; Anctil et al., 2004; Wagener et al., 2001; Ye et al., 1998). For this study, the model was implemented in the Rainfall-Runoff Modeling Toolbox (RRMT) in the Matlab environment (Wagener et al., 2002). The RRMT is a modular framework to implement lumped parsimonious model structures consisting of a soil moisture accounting and a routing component. The implemented model consists of a probability distribution of overflow
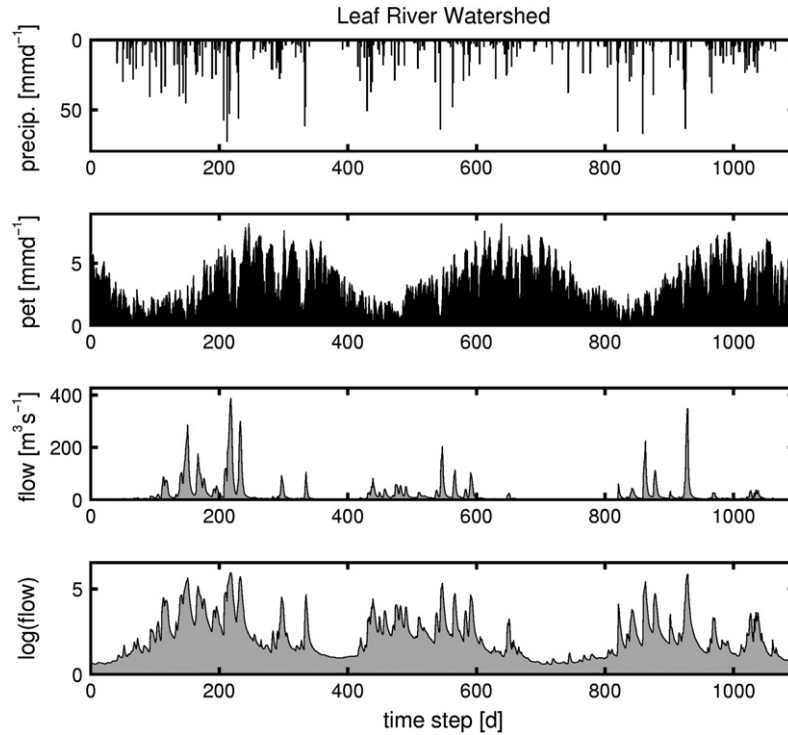
Fig. 3. Time-series extract from the perennial Leaf River watershed used in this study.

storage elements that serve as the soil moisture accounting component (Moore and Clark, 1981). The effective rainfall produced by the overflow mechanism is split into two parts of which one is routed through a quick-flow linear reservoir and the rest through a slow-flow linear reservoir. The model structure requires estimation of five parameters: C_{max}, the maximum storage size; b, a shape parameter describing the Pareto distribution of storage elements; α, the fraction of effective rainfall routed through the quick-flow reservoir; k_q, the quick-flow linear reservoir time constant; and k_s, the slow-flow linear reservoir time constant. Actual evapotranspiration is modeled as a linear correlation between the modeled soil moisture state and the potential evapotranspiration.

Uniform random sampling (URS) was used to explore the parameter space (10,000 samples) and three objective functions (F) were calculated. The three Fs are all based on the Root Mean Squared Error Measure (RMSE), but evaluate the model performance in fitting different parts of the streamflow hydrograph (Fig. 5),

$$F = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (o_i - p_i)^2} \qquad (1)$$

where F is the root mean squared error measure calculated for either of the three measures when the observed flow is in the high (FH), medium (FM), or low (FL) range, p is the predicted flow, o is the observed flow, N is the number of time-steps the observed flow is in the defined flow range, and i is an index running from 1 to N. For this case study, the low and high flow thresholds were set at 1 and 3 mm d^{-1} respectively.
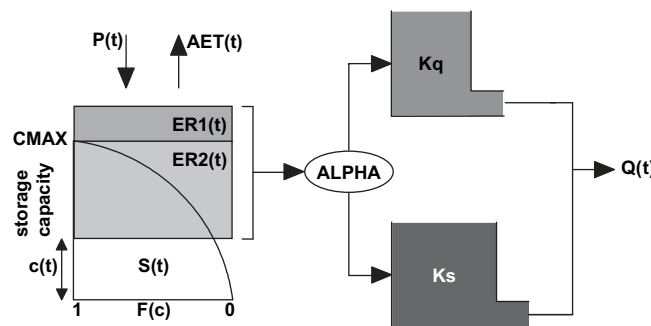

Fig. 4. Model structure used in this case study. Vertical processes are represented using a probability distribution of overflow storage elements. Effective rainfall produced by the overflow mechanism is split into two parts, one of which is routed through a quick-flow linear reservoir and the rest through a slow-flow linear reservoir. Five parameters are required for this model.
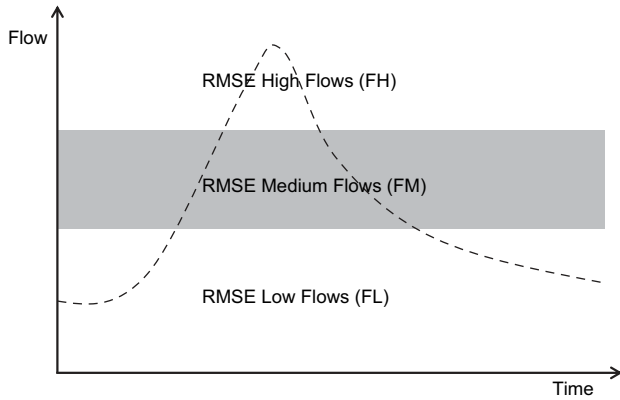
Fig. 5. Hydrograph segmentation used in case study. The dashed line is the observed flow. Low and high flow thresholds were set at 1 and 3 mm day^{-1} respectively.

4. Model identification

Model identification, in the context of this paper, refers to the identification of all models (parameter sets) within a given model structure, that can be considered feasible (behavioral) representations of the natural system under study. Following Fig. 2, we assume that this stage has two components: parameter estimation and sensitivity analysis.

4.1. Parameter estimation

Plots are available for this step to evaluate how narrowly a particular parameter can be identified using a specific performance measure.

4.1.1. Dotty plots

Dotty plots map model parameter values and their corresponding objective function values to one-dimensional points and provide a means of assessing the identifiability of model parameters. Dotty plots of the objective values versus the *b* parameter values resulting from the URS of the parameter space are shown in Fig. 6a. Each row of the plot matrix corresponds with each of the three objective functions: the high (FH), medium (FM), and low (FL) flow RMSE. In the MCAT, the user is provided with a slider bar capable of changing the objective function threshold which is displayed on the dotty plots to aid in providing a visualization which best displays the identifiability of each parameter. In Fig. 6a, the top ranked 30-percent (in terms of the objective function values) of the 10,000 URS are shown in order to better illustrate the identifiability of the *b* parameter in terms of each objective function. The dotty plot of FL versus *b* shows a high level of identifiability indicating that this parameter is most important during low flow periods. A range of *b* values result in very similar objective function values for FH and FM, indicating low identifiability in terms of these objectives.
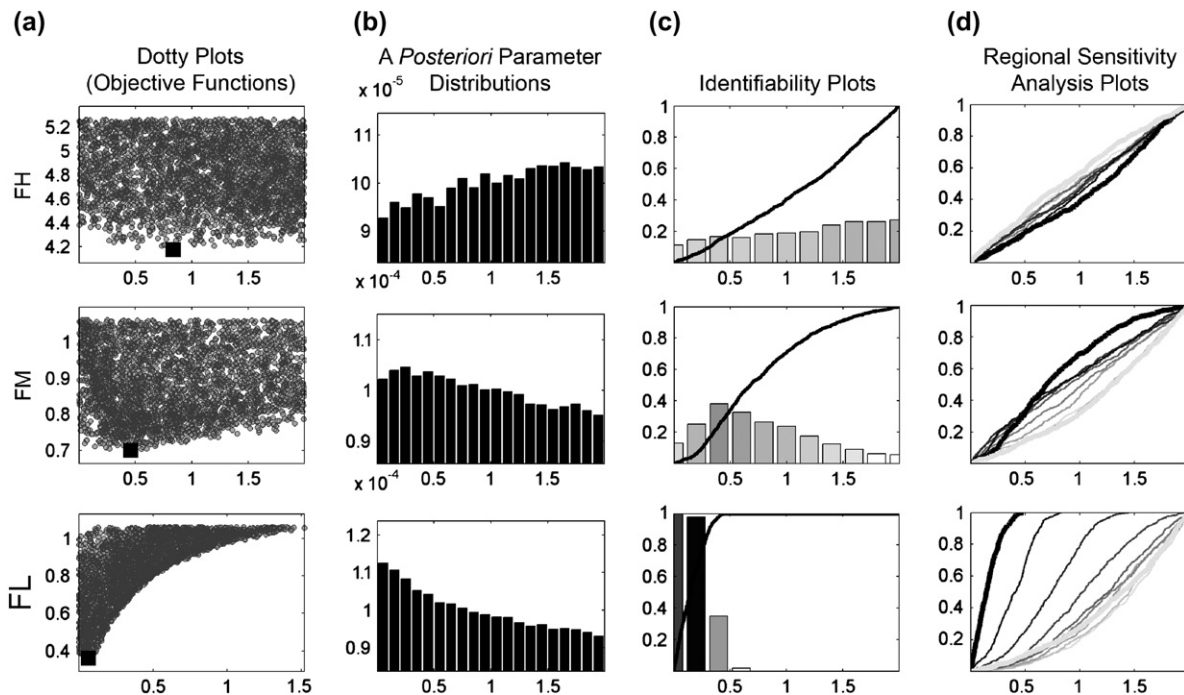

Fig. 6. Four types of MCAT visualizations intended for model identification and evaluation of the *b* parameter used in this case study. Dotty plots (a) provide a means of evaluating parameter identifiability. A posteriori parameter distributions (b) provide a means of visualizing the distributions of parameter values conditioned on an objective function. Identifiability plots (c) help visualize the identifiability of model parameters by plotting the cumulative distribution of the top 10-percent of the parameter population in terms of the objective function values. Regional Sensitivity Analysis Plots (d) rank the marginal parameter population from best to worst in terms of the chosen objective function. The ranked population is then divided into ten bins of equal size according to their objective values. The cumulative distribution of each group is then plotted as the likelihood value versus the parameter values.

4.1.2. A posteriori parameter distributions

A *posteriori* parameter distributions shown in Fig. 6b provide a means of visualizing the distributions of parameter values conditioned on an objective function. The range of each parameter is divided into 20 bins of equal size and the sum of all likelihoods (defined as one minus the objective function value normalized to between zero and one) is plotted for each bin. This visualization technique provides a tool for choosing high likelihood model parameter values based on their posterior distributions. For example, in Fig. 6b, the distribution of *b* conditioned on FL indicates that the lowest *b* values are associated with the highest likelihoods and likewise, the highest *b* values are associated with the lowest likelihoods.

4.1.3. Identifiability plots

Identifiability plots shown in Fig. 6c provide another means of visualizing the identifiability of model parameters by plotting the cumulative distribution of the top 10-percent of the parameter population in terms of objective function values. High gradients in the cumulative distribution indicate high identifiability in the top performing model parameters (see FL in Fig. 6c) whereas shallower gradients indicate low identifiability (see FH Fig. 6c). In addition to the cumulative distribution function, the top 10-percent of the parameter population is divided into ten bins of equal size and the gradient of the cumulative distribution is then calculated for each group. These gradients are plotted as bars on the identifiability plots with shading indicative of gradient. This provides additional visualization functionality as the height and shade of the bars indicate parameter identifiability within the range of each group. In Fig. 6c, the reader can see that *b* has the highest identifiability in terms of FL, which is determined from the high gradient of the cumulative distribution of the top performing parameter values as well as the high gradient at low *b* values indicated by the bar plot in the figure.

4.1.4. Response surface plots

This plot produces a visualization of 2-D parameter surfaces (no figure shown), i.e. interpolated plots of objective function values (*z*-axis) for two selected parameters (*x* and *y*-axis). In order to examine parameter interaction, a Monte-Carlo simulation should be run with two parameters free to vary and the other parameters fixed, otherwise the surface will be noisy since parameter pairs in close proximity will have large differences in the objective or variable value caused by variations in other parameters (scatter in dotty plots).

5. Sensitivity analysis

Sensitivity analysis evaluates the impact of changes in the model parameters, inputs or (initial) states on the model output of interest. The reason for performing a sensitivity analysis is usually one of the following (e.g. Wagener et al., 2002, 2003; Gooseff et al., 2005; Sieber and Uhlenbrook, 2005; McIntyre and Wheater, 2004; Letcher et al., 2007): (1) testing which parameters dominate a certain response in order to eliminate insensitive parameter to reduce the calibration burden; (2) as part of an a priori uncertainty analysis to test how well parameters are defined, or to test where additional effort should be placed to reduce uncertainty (e.g., improve quality of input data); and (3) more recently, different variations of temporal analysis of uncertainty have emerged in the literature to test for example whether parameters are sensitive in periods where the processes they represent are assumed to dominate. Sensitivity analysis can therefore be used as a tool for model structure evaluation. This aspect is discussed in detail in Section 7. Sensitivity analysis is also used to evaluate the for evaluating different water resource management or scenario options (e.g. Merritt et al., 2005; Dessai et al., 2005).

A wide variety of approaches to sensitivity analysis exist, but they can generally be reduced to the following two basic components:

(1) a strategy to vary the parameters (or inputs or state variables),
(2) a numerical or visual measure to quantify the impacts of the variation on the model output of interest.

How these two components are implemented varies widely with the approach used and a wide range of techniques are available for this purpose (e.g. Hamby, 1994; Frey and Patil, 2002; Patil and Frey, 2004; Pappenberger et al., 2006; Vandeberghe et al., 2006). These techniques vary from the simplest "one parameter at a time" perturbation approach in which individual parameters are varied using a certain step size and the impact of this variation is measured based on a chosen objective function. This approach has the advantage of simplicity, but is usually unreliable for high-dimensional and non-linear models with correlated parameters that we commonly face in environmental and hydrological modeling. Global approaches are more commonly used today in which the local sensitivity around a specific point in the parameter space is not only tested, but an attempt is made to evaluate the entire parameter space. Many of these approaches used for global sensitivity analysis are related to the Regional Sensitivity Analysis (RSA) technique originally proposed by Hornberger and Spear (1981). The RSA method begins with a Monte Carlo sampling of *N* points in the feasible parameter space, drawn from a multivariate uniform distribution. The sampled parameter population is partitioned into a behavioral (*B*) and a non-behavioral (*NB*) group. Behavioral means parameter sets that produce a model response (behavior) that is preferred. The division into behavioral and non-behavioral can be based on the predicted state of the system (e.g., Spear and Hornberger, 1980) or on a measure of performance (e.g., Hornberger et al., 1985; Beven and Binley, 1992). The cumulative marginal parameter distributions for the two groups are computed. A separation between the distribution curves indicates a statistical difference between the characteristics of the two (behavioral and non-behavioral) subpopulations. This indicates that the tested parameter is sensitive, i.e., its value can be strongly correlated with model performance. The significance of the separation can be estimated using statistical tests such as the Kolmogorov–Smirnov (KS)

two-sample test (Kottegoda and Rosso, 1997), and a heuristic ranking scheme can be introduced based on the actual values of the KS measure (Spear and Hornberger, 1980). An unfortunate weakness of this approach is that a lack of separation between the cumulative distributions is only a necessary, and not a sufficient condition for insensitivity of the parameter (Spear, 1993). Insensitivity can also be caused by strong correlation with other parameters. Evaluation of the parameter covariance can be used to estimate whether this is the case (Hornberger and Spear, 1981; Hornberger et al., 1985). The interaction between two parameters can also be investigated in the MCAT by plotting their response surface with respect to a particular objective function. The RSA approach has also been used for the identification of model structures (Osidele and Beck, 2001) and for the evaluation of data requirements (Lence and Takyi, 1992). Other popular global approaches to sensitivity analysis are for example based on variance decomposition (Saltelli et al., 2004; Helton, 1997; Andres, 1997).

5.1. Regional sensitivity analysis

Parameter sensitivity can be evaluated in the MCAT using the Regional Sensitivity Analysis (RSA) plotting tool. The variation on the original RSA approach used here was originally introduced by Freer et al. (1996). In this approach, the marginal parameter population is ranked from best to worst in terms of the chosen objective function. The ranked population is then divided into ten bins of equal size according to their objective values. The objective values associated with each parameter set are then converted to likelihood measures (described in Section 5.1). The cumulative distribution of each group is then plotted as the likelihood value versus the parameter values. RSA plots for the b parameter in terms of each objective are shown in Fig. 6d with the highest likelihood parameter distributions indicated by bold black lines and the lowest likelihood distributions indicated by light gray lines. The colors of the lines representing the cumulative distributions of the remaining eight groups are scaled accordingly. Parameter sensitivity can be evaluated from Fig. 6d by assessing the "spread" of the ten lines of cumulative parameter distributions. For example, low values of b in terms of FL are associated with the highest likelihood which can be seen from the quick rising black line. This indicates that b is highly sensitive because there tends to be large differences in the parameter value for large differences in likelihood of the parameter producing good model output. Insensitive parameters are identified by noting similarity in the cumulative distributions across the ten ranked groups. The amount of "spread" associated with the cumulative distributions of the ranked distributions indicates the parameters sensitivity with high spread indicating highly sensitive parameters and low spread indicating relatively insensitive parameters.

6. Further model evaluation

As discussed in Section 2, two questions that should be asked in addition to the standard evaluation of parameter uncertainty are: (1) does a model parameter or a group of parameters represent the processes they are intended to represent (i.e., do they dominate the model response when this process dominates the system response); and (2) are regions of behavioral parameters constant in time, or do they vary with different response modes of the system?

6.1. Dynamic Identifiability Analysis (DYNIA)

A Monte Carlo based approach to address the above questions has been introduced by Wagener et al. (2003). The approach is called Dynamic Identifiability Analysis (DYNIA, refer to Wagener et al., 2002, 2003 for a detailed description) which presents an algorithm to improve the amount of information that can be retrieved from observations for model evaluation. The algorithm is based on ideas presented by Beck (1985; see also Beck, 2005) and extends components of the GLUE algorithm (Beven and Binley, 1992; Freer et al., 1996). DYNIA can be used to find informative regions with respect to model parameters, to test model structures (assuming that varying parameter optima indicate structural problems), and to analyze experimental design. It uses the above introduced identifiability measure (Fig. 6c) and applies it in a dynamic fashion using a Monte Carlo based smoothing algorithm. The user must choose a window size to calculate a running mean of the model's performance (using the mean absolute error criterion). A different identifiability plot is thus produced for every time step and a gray color scheme is used to show the variation in the marginal posterior distributions for each parameter. The window size has to be selected with respect to the function of the parameter (temporal length of the region of influence) and the quality of the data (better data allows the use of a smaller window). Fig. 7 is an example output where the top plot shows the identifiability plot with darker gray regions indicating peaks in the distributions. The dotted line is the 90-percent confidence limit. The bottom plot shows (in black) the width of the 90-percent confidence limits as an indicator of regions of high information content with respect to this parameter. The continuous line in both plots is the observed time-series. The algorithm only utilizes the top 10-percent of all data sets to calculate the distribution at every time-step in this case.

For example from Fig. 7, the user can see that the information content of the data with respect to parameter b is highest at small runoff events during low flow periods (bottom plot). The top plot shows that during these periods, good values of the parameter tend to be towards the lower end of its feasible range (dark pixels). The posterior distribution is peaked then, while it is flat (and the pixels are light gray) during other periods. The value of b has to be low in order not to overpredict runoff during low flow.

6.2. Multi-objective analysis

Several plots in the MCAT allow for the analysis of multiple objectives and their impact on model identification and performance. One important plot in this context is the
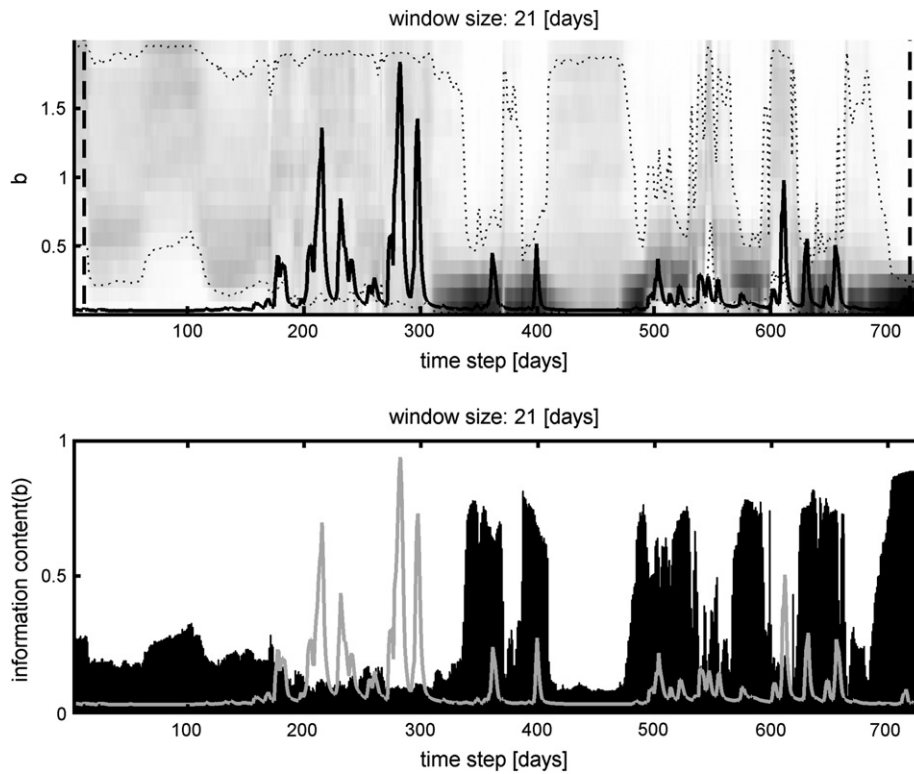
Fig. 7. Dynamic identifiability analysis (DYNIA) plots of parameter b, showing: (Top) Temporal change of the marginal posterior distribution, (black line: observed streamflow, dotted lines: 90% confidence limits); (Bottom) Temporal variation of data information content with respect to b (i.e. defined as normalized width of 90% confidence limits).

multi-objective plot which allows us to evaluate the correlation and trade-offs between different objective functions. These objective functions often provide conflicting optimization targets (Chankong and Haimes, 1993). In most cases, no solution will be available that is best in terms of all objective functions. As a result, a population of solutions is identified which are non-dominated (or incomparable) with respect to one another, meaning that an improvement in performance in one objective results in a degradation of performance in one or more of the remaining objectives. This concept was introduced by the economist Pareto at the end of the eighteenth century (Pareto, 1906), and as a result, the optimal population in a multi-objective sense is referred to as a Pareto front. Fig. 8 shows multi-objective plots and linear correlations between the FL, FM, and FH objective functions. This figure is a scatter plot where all objective functions are plotted against each other and their linear correlation coefficients are calculated. A large scatter indicates that objectives are uncorrelated and a trade-off front facing the bottom left corner of the scatter plots is usually apparent. On the other hand, a convergence of the objective functions towards the bottom left corner indicates that these measures of performance retrieve similar information from the investigated time-series. Trade-offs between criteria can for example be used to analyze model structural problems (e.g. Gupta et al., 1998; Boyle et al., 2000; Wagener et al., 2001; Lee et al., 2005). For example, Fig. 8 shows that FL and FM are highly correlated, while a clear trade-off between the two and FH is visible.

7. Ensemble prediction including uncertainty

In the final modeling step, the predictions of all models (parameter sets) have to be propagated into the model output. The options currently available in MCAT to accomplish this are explained briefly below.
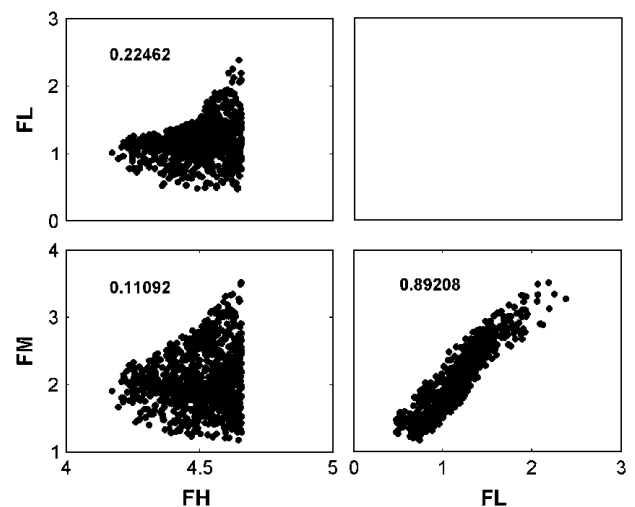

Fig. 8. Multi-objective plot showing linear correlation and trade-off between the different objective functions. High correlation indicates that the set of objectives are providing similar information whereas low correlation, (i.e., a larger trade-off) indicates that the objectives are providing information for multiple aspects of model performance.

7.1. Ensemble prediction using GLUE or multiple-objectives

Two options are available to consider uncertainty in the model predictions or output. These are adopted from Beven and Binley (1992) and from Gupta et al. (1998) respectively. The first uses the weighting procedure of the Generalized Likelihood Uncertainty Estimation (GLUE) procedure in which, for each time step, the output calculated with a specific parameter set is weighted by the likelihood of the same parameter set. For each point in time, a cumulative frequency distribution is generated using the selected objective function (converted to likelihood by normalizing it) and the confidence intervals are calculated using linear interpolation. The confidence intervals can be specified from the MCAT menu. Details of this approach can be found in Beven and Binley (1992) and in Beven and Freer (2001). The second option, a plot that propagates the simulated variable calculated with the Pareto solution into the output space — based on the paper by Gupta et al. (1998), can be used in cases where multiple objective functions are utilized.

7.2. Pixel simulation plots

Pixel simulation plots assist in: (1) guiding the search for optimal model parameters based on objective functions; and (2) subsequently fine tuning the search using large scale visualizations. A pixel-oriented visualization technique (Keim, 1996) provides the modeler with the ability to view large amounts of model simulation data on a single screen. Difficulty in viewing large time-series' can be remedied by representing the time series as a one-dimensional row of pixels which are colored based on observation data, model output, etc.

In the MCAT, the pixel simulation plot ranks each URS in terms of its objective value and plots the simulation as a row of pixels (i.e., one-dimensional time series) where the color of the pixel represents the error between the observed system and the model output (shown in Fig. 9). The traditional method of plotting the time-series is displayed at the top of the figure followed by the observed streamflow indicated by a row of colored pixels immediately below to provide a reference for the user. The model output error associated with the top 50 URS ranked in terms of FH are shown in the lower portion of Fig. 9. This reveals that assessing a particular combination of model parameters strictly in terms of an objective function results in a loss of information. Dark blue bands in the pixel simulation plot, indicating simulations which result in generally lower model output error make this apparent. Instead of basing model parameter selection strictly on objective function values, we can instead choose and compare simulations from the pixel plot which appear to be producing better model output than the objective function value may indicate. In the MCAT, the user is provided with the ability to pick out individual simulations which appear to result in low model output error, and subsequently compare these simulations side by side to make a final determination as to which simulation is

optimal. The user is also provided with a zooming capability in order to focus on interesting areas of the full pixel simulation plot, ultimately allowing them to choose ranked simulations which they would like to examine and compare more closely. These simulations are then entered into a list box and the ability to produce subsequent comparison plots is provided. This interactive capability provides the user with an effective parameter selection framework which utilizes objective functions to initially guide parameter search, and ultimately human cognition through visualization to fine tune parameter search.

7.3. Time series pattern plots

The time series pattern plotting capability, also based on the pixel-oriented visualization techniques of Keim (1996), is designed to show temporal trends in large observation data sets. An example time series pattern plot is shown in Fig. 10. This visualization technique plots a large time series separated into years on the x-axis and the time series data within each year on the y-axis. The color scale of each plot is used to represent the observed system response, the model output, and the model output error associated with the best simulation, which can be chosen through the pixel simulation plot methodology proposed above, or by choosing the simulation in terms of a specific objective. In Fig. 10, the seasonality associated with each water year is apparent along the y-axis as well as the temporal trend across water years as shown by looking along the x-axis. For the ten years of daily data available for the Leaf River, it appears that the streamflow may be increasing over time as is indicated by the higher flows associated with years 9 and 10.

8. Conclusions and outlook

Recent papers have discussed the importance of good modeling practice in hydrological and environmental modeling (e.g. Jakeman et al., 2006). This practice should include quality control (Refsgaard and Henriksen, 2004), the estimation of uncertainty and the evaluation of underlying assumptions (Beven, 2006; Wagener, 2003; Jolma and Norton, 2005). Addressing these aspects in detail and communicating the analysis results to the end-user of the model results is an important aspect of obtaining credible model predictions. The detailed consideration and estimation of uncertainty in the modeling of hydrological and environmental systems is increasingly required in modeling studies. If done properly, this requires a detailed evaluation of the mathematical models used. These requirements come at a time where models increase in complexity due to the inclusion of an increasing number of state and output variables. The Monte Carlo Analysis Toolbox (MCAT) is a library of visual and numerical analysis tools for the evaluation of hydrological and environmental models. Input to the MCAT is the result of a Monte Carlo or population evolution based sampling of the parameter space of the model structure under investigation. This loose (off-line) connection therefore enables every model to be
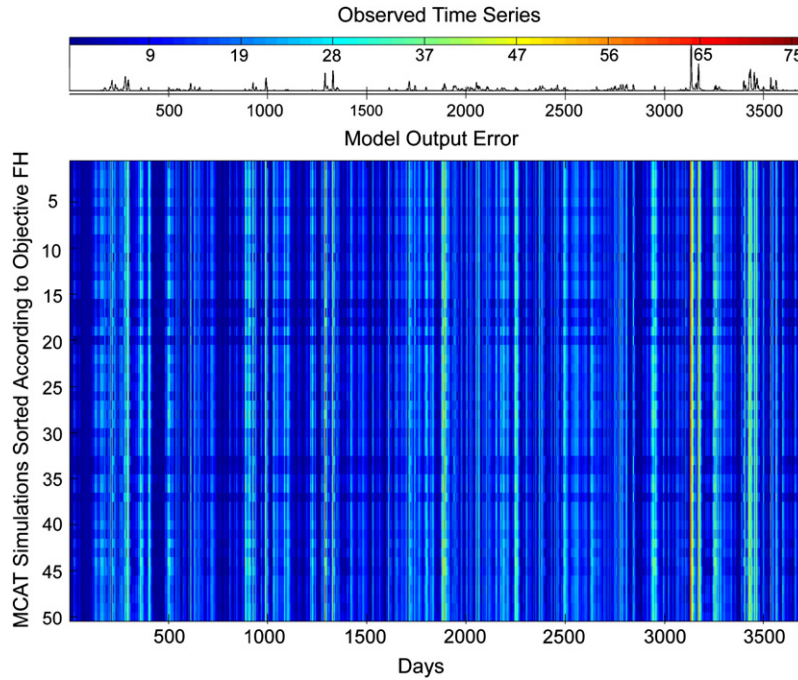
Fig. 9. Pixel plot showing the top 50 simulations ranked in terms of the FH objective values. This type of plot provides an indication of how well the chosen objective function is mapping the performance of the model to a single value of performance. For example, the dark blue bands in this figure indicate simulations which tend to result in an overall better performance than their surrounding simulations (i.e., better than what their corresponding objective function value may be indicating).

analyzed as long as a Monte Carlo type sampling can be performed. A library of tools for the evaluation of performance, identifiability, sensitivity, predictive uncertainty and for the testing of hypotheses is available. The MCAT has thus far

been successfully applied to evaluate models in a research context in hydrological and environmental studies. It has also been used in multiple graduate courses (systems modeling and hydrology) and workshops to teach the fundamentals of
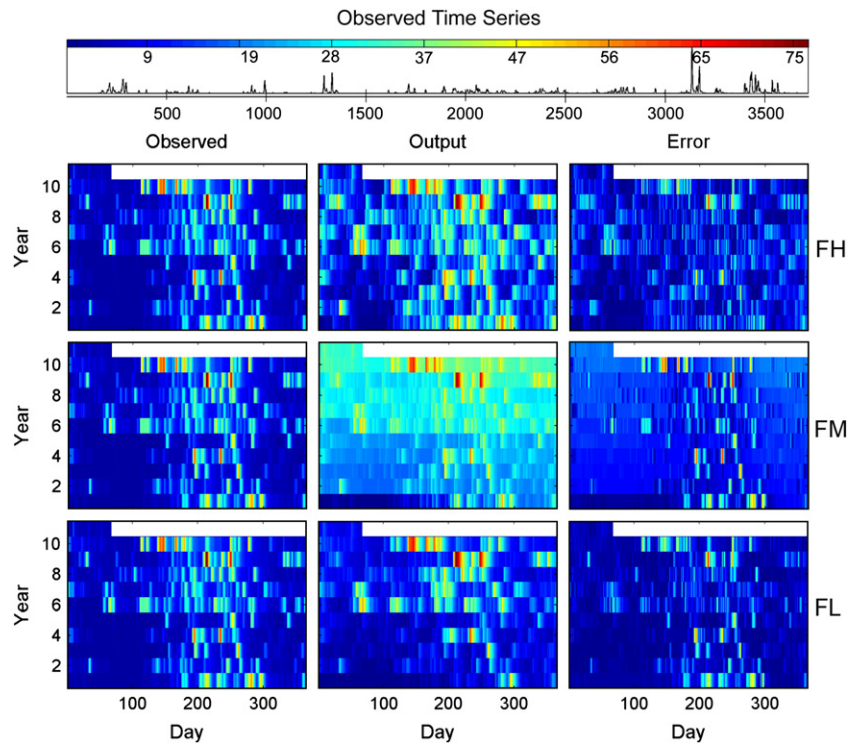

Fig. 10. Temporal pixel plots which portray the observed flow, the modelled flow, and the model error using a color scheme in terms of the three objective functions. These types of plots can reveal temporal trends in data. For this case study, stream flow generally seems to be increasing over the 10 year period analyzed.

model identification and evaluation. Its simplicity of use and the ease with which model results can be visualized make it a valuable tool in this context. Further evaluation capabilities are continuously added to the MCAT to increase its versatility in analyzing dynamic mathematical models.

The MCAT toolbox and other computational tools for hydrological and environmental modeling are available from the Hydroarchive website (Wagener et al., 2003; http://www.sahra.arizona.edu/software/). Tools for download include optimization algorithms, rainfall-runoff models, artificial neural networks, etc. GLUE is available as freeware from http://www.es.lancs.ac.uk/hfdg/hfdg_freeware.htm. SIMLAB can be downloaded from http://webfarm.jrc.cec.eu.int/uasa/primer/index.asp.

Acknowledgements

References

Anctil, F., Perrin, C., Andréassian, V., 2004. Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. Environmental Modelling & Software 19 (4), 357–368.

Andres, T.H., 1997. Sampling methods and sensitivity analysis for large parameter sets. Journal of Stochastic Computation and Simulation 57, 75–110.

ASCE, 1993. Criteria for evaluation of watershed models. Journal of Irrigation and Drainage Engineering 119 (3), 429–442.

Beck, M.B., 1985. Structures, failures, inference and prediction. In: Barker, M.A., Young, P.C. (Eds.), Identification and System Parameter Estimation. Proceedings 7th Symposium vol. 2, July 1985. IFAC/IFORS, York, UK, pp. 1443–1448.

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

Beck, M.B., 2005. Environmental foresight and structural change. Environmental Modelling & Software 20 (6), 651–670.

Beven, K.J., 2000. Rainfall-Runoff Modeling: The Primer. John Wiley & Sons, Chichester, UK.

Beven, K.J., 2002. Towards a coherent philosophy for environmental modeling. In: Proceedings of the Royal Society of London A 458, 2465–2484.

Beven, K.J., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320 (1–2), 18–36.

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

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

Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research 36 (12), 3663–3674, doi:10.1029/2000WR900207.

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

Croke, B.F.W., Jakeman, A.J., 2004. A catchment moisture deficit module for the IHACRES rainfall-runoff model. Environmental Modelling & Software 19 (1), 1–5.

Dessai, S., Lu, X., Hulme, M., 2005. Limited sensitivity analysis of regional climate change probabilities for the 21st century. Journal of Geophysical Research 110, D19108, doi:10.1029/2005JD005919.

Dingman, S.L., 2002. Physical Hydrology. Prentice Hall, Upper Saddle River, NJ.

Diskin, M.H., Simon, E., 1977. A procedure for the selection of objective functions for hydrologic simulation models. Journal of Hydrology 34, 129–149.

Dooge, J.C., 1972. Mathematical models of hydrologic systems. In: Proceedings of the International Symposium on Modeling Techniques in Water Resources Systems, Ottawa, Canada, vol. 1, 171–189.

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

Freer, J., Beven, K.J., 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.

Frey, H.C., Patil, S.R., 2002. Identification and review of sensitivity analysis methods. Risk Analysis 22 (3), 553–578.

Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Kluwer, Dordrecht, NL, 229 pp.

Gooseff, M.N., Bencala, K.E., Scott, D.T., Runkel, R.L., McKnight, D.M., 2005. Sensitivity analysis of conservative and reactive stream transient storage models applied to field data from multiple-reach experiments. Advances in Water Resources 28 (5), 479–492, doi:10.1016/j.advwatres.2004.11.012.

Grayson, R.B., Moore, I.D., McMahon, T.A., 1992. Physically based hydrologic modelling: II. Is the concept realistic? Water Resources Research 26 (10), 2659–2666.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information. Water Resources Research 39 (5), 751–763.

Gupta, H.V., Beven, K.J., Wagener, T., 2006. Model calibration and uncertainty estimation. In: Anderson, M.G. (Ed.), Encyclopedia of Hydrological Sciences. John Wiley & Sons Ltd., Chichester, UK, pp. 1–17.

Hamby, D.M., 1994. A review of techniques for sensitivity analysis of environmental models. Environmental Modelling and Assessment 32 (2), 135–154.

Helton, J.C., 1997. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. Journal of Stochastic Computation and Simulation 57 (1–4), 3–76.

Helton, J.C., Davis, F.J., 2002. Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Analysis 22 (3), 622–691.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering and System Safety 81 (1), 23–69.

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

Hornberger, G.M., Beven, K.J., Cosby, B.J., Sappington, D.E., 1985. Shenandoah watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. Water Resources Research 21, 1841–1859.

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

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21 (5), 602–614.

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

Jolma, A., Norton, J., 2005. Methods of uncertainty treatment in environmental models. Environmental Modelling & Software 20 (8), 979–980.

Kavetski, D., Franks, S., Kuczera, G., 2002. Confronting input uncertainty in environmental modelling. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibration of Watershed Models. Water Science and Application 6. American Geophysical Union, pp. 49–68. 10/1029/006WS04.

Keim, D.A., 1996. Pixel-Oriented Database Visualizations. Sigmond Record, Special Issue on Information Visualization.

Kottegoda, N.T., Rosso, R., 1997. Statistics, Probability, and Reliability for Civil and Environmental Engineers. McGraw-Hill, New York, USA.

Lee, H., McIntyre, N., Wheater, H.S., Young, A., 2005. Selection of conceptual models for regionalisation of the rainfall-runoff relationship. Journal of Hydrology 312, 125–147.

Lence, B.J., Takyi, A.K., 1992. Data requirements for seasonal discharge programs: an application of a regionalized sensitivity analysis. Water Resources Research 28 (7), 1781–1789.

Letcher, R.A., Croke, B.F.W., Jakeman, A.J., 2007. Integrated assessment modeling for water resource allocation and management: A generalized conceptual framework. Environmental Modelling & Software 22 (5), 733–742.

Martinec, J., Rango, A., 1989. Merits of statistical criteria for the performance of hydrological models. Water Resources Bulletin 25 (2), 421–432.

McIntyre, N.R., Wheater, H.S., 2004. A tool for risk-based management of surface water quality. Environmental Modelling & Software 19 (12), 1131–1140.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Merritt, W.S., Croke, B.F.W., Jakeman, A.J., 2005. Sensitivity testing of a model for exploring water resources utilization and management options. Environmental Modelling & Software 20, 1013–1030.

Moore, R.J., Clarke, R.T., 1981. A distribution function approach to rainfall runoff modeling. Water Resources Research 17 (5), 1367–1382.

Osidele, O.O., Beck, M.B., 2001. Identification of model structure for aquatic ecosystems using regionalized sensitivity analysis. Water Science and Technology 43 (7), 271–278.

Pappenberger, F., Iorgulescu, I., Beven, K.J., 2006. Sensitivity analysis based on regional splits and regression trees (SARS-RT). Environmental Modelling & Software 21, 976–990.

Pareto, V., 1906. Manual of Political Economy. 1971 translation. Augustus M. Kelley, New York.

Patil, S.R., Frey, H.C., 2004. Comparison of sensitivity analysis methods based upon applications to a food safety risk model. Risk Analysis 23 (3), 573–585.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C, second ed. Cambridge University Press, Cambridge, UK.

Refsgaard, J.C., Henriksen, H.J., 2004. Modelling guidelines – terminology and guiding principles. Advances in Water Resources 27 (1), 71–82.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Wiley, ISBN 0-470-87093-1.

Sieber, A., Uhlenbrook, S., 2005. Sensitivity analyses of a distributed catchment model to verify the model structure. Journal of Hydrology 310 (1–4), 216–235.

Sincock, A.M., Wheater, H.S., Whitehead, P.G., 2003. Calibration and sensitivity analysis of a river water quality model under unsteady flow conditions. Journal of Hydrology 277 (3–4), 214–229.

Singh, V.P., 1988. Hydrologic Systems – Rainfall-Runoff Modeling, vol. 1. Prentice Hall, Englewood Cliffs, New Jersey.

Singh, V.P., Woolhiser, D.A., 2002. Mathematical modeling of watershed hydrology. Journal of Hydrologic Engineering 7 (4), 270–292.

Smith, R.M.S., Wheater, H.S., 2004. Multiple objective evaluation of a simple phosphorus transfer model. Hydrological Processes 18 (9), 1703–1720.

Smith, R.M.S., Evans, D.J., Wheater, H.S., 2005. Evaluation of two hybrid metric-conceptual models for simulating phosphorus transfer from agricultural land in the river Enborne, a lowland UK catchment. Journal of Hydrology 304 (1–4), 366–380.

Spear, R.C., 1993. Regional sensitivity analysis in environmental systems. In: Young, P.C. (Ed.), Concise Encyclopedia of Environmental Systems. Pergamon Press, Oxford, UK, pp. 476–479.

Spear, R.C., Hornberger, G.M., 1980. Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research 14, 43–49.

Uhlenbrook, S., Seibert, J., Leibundgut, C., Rohde, A., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structures. Hydrological Sciences Bulletin 44 (5), 779–797.

Van Asselt, M.B.A., Rotmans, J., 1996. Uncertainty in perspective. Global Environmental Change 6 (2), 121–157.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39 (8), 1201, doi:10.1029/2002WR001642.

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

Wagener, T., Gupta, H.V. Model identification for hydrological forecasting under uncertainty. Stochastic Environmental Research and Risk Assessment 19, 378–387, doi:10.1007/s00477-005-0006-5.

Wagener, T., McIntyre, N., 2005. Identification of hydrologic models for operational purposes. Hydrological Sciences Journal 50 (5), 735–751.

Wagener, T., Wheater, H.S., 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology 320 (1–2), 132–154.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrology and Earth System Sciences 5 (1), 13–26.

Wagener, T., Lees, M.J., Wheater, H.S., 2002. A toolkit for the development and application of parsimonious hydrologic models. In: Singh, V.P., Frevert, D.K. (Eds.), Mathematical Models of Large Watersheds Hydrology. Water Resources Publications LLC, USA, pp. 91–140.

Wagener, T., Gupta, H.V., Carpenter, K., James, B., Vázquez, R., Sorooshian, S., Shuttleworth, J., 2003. A hydroarchive for the free exchange of hydrological software Website. Hydrological Processes 18 (2), 389–391.

Wagener, T., Wheater, H.S., Gupta, H.V., 2004. Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London, UK.

Ye, W., Jakeman, A.J., Young, P.C., 1998. Identification of improved rainfall-runoff models for an ephemeral low-yielding Australian catchment. Environmental Modelling & Software 13 (1), 59–74.

Young, P.C., 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. Environmental Modelling & Software 13 (2), 105–122.

Young, P.C., 2001. Data-based mechanistic modelling and validation of rainfall-flow processes. In: Anderson, M.G., Bates, P.D. (Eds.), Model Validation: Perspectives in Hydrological Sciences. John Wiley & Sons, Ltd.