Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation

Jasper A. Vrugt,[1] Cees G. H. Diks,[2] Hoshin V. Gupta,[3] Willem Bouten,[1] and Jacobus M. Verstraten[1]

Received 30 January 2004; revised 12 November 2004; accepted 23 November 2004; published 27 January 2005.

[1] Hydrologic models use relatively simple mathematical equations to conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed. A consequence of process aggregation is that the model parameters often do not represent directly measurable entities and must therefore be estimated using measurements of the system inputs and outputs. During this process, known as model calibration, the parameters are adjusted so that the behavior of the model approximates, as closely and consistently as possible, the observed response of the hydrologic system over some historical period of time. In practice, however, because of errors in the model structure and the input (forcing) and output data, this has proven to be difficult, leading to considerable uncertainty in the model predictions. This paper surveys the limitations of current model calibration methodologies, which treat the uncertainty in the input-output relationship as being primarily attributable to uncertainty in the parameters and presents a simultaneous optimization and data assimilation (SODA) method, which improves the treatment of uncertainty in hydrologic modeling. The usefulness and applicability of SODA is demonstrated by means of a pilot study using data from the Leaf River watershed in Mississippi and a simple hydrologic model with typical conceptual components.

Citation: Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

1. Introduction and Scope

[2] Hydrologic models often contain parameters that cannot be measured directly, but can only be meaningfully inferred by calibration to a historical record of input-output data. During model calibration, the parameters are adjusted in such a way that the behavior of the model approximates, as closely and consistently as possible, the observed response of the hydrologic system over some historical period of time. Because of the time consuming nature of manual trial-and-error model calibration, there has been a great deal of research into the development of automated (computer based) calibration methods [see, e.g., *Duan et al.*, 1992; *Gupta and Sorooshian*, 1994; *Gupta et al.*, 1998; *Yapo et al.*, 1998; *Boyle et al.*, 2000]. Automatic methods for model calibration seek to take advantage of the speed and power of computers, while being relatively objective and easier to implement than manual methods.

[3] While considerable progress has been made in the development and application of automated calibration methods, many of these approaches treat the underlying uncertainty in the input-output representation of the model as being primarily (and explicitly) due to uncertainty in the parameter estimates. A few papers do discuss the treatment of input, state, and model structural uncertainties for environmental models [e.g., *Young and Beck*, 1974; *Beck and Young*, 1975; *Kitanidis and Bras*, 1980a, 1980b; *Beck*, 1985, 1987; *Hebson and Wood*, 1985; *Young*, 1986], but such approaches have not become common practice for nonlinear watershed models. Clearly, uncertainties in the modeling procedure stem not only from uncertainties in the parameter estimates, but also from measurement errors associated with the system input (forcing) and output, and from model structural errors arising from the aggregation of spatially distributed real-world processes into a mathematical model. Not properly accounting for these errors during model calibration, can result in model simulations and associated prediction uncertainty bounds, which do not consistently represent and bracket the measured system behavior. This is evidenced by residuals which exhibit considerable variation in bias (nonstationarity), variance (heteroscedasticity), and correlation structures under different hydrologic conditions. Several contributions to the hydrologic literature have therefore brought into question the continued usefulness of the classical paradigm for estimating model parameters (see, e.g., *Beven and Binley*

[1]Department of Physical Geography, Faculty of Science, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands.
[2]Department of Quantitative Finance, Faculty of Economics and Econometrics, CenDEF, University of Amsterdam, Amsterdam, Netherlands.
[3]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

0043-1397/05/2004WR003059$09.00

[1992], *Gupta et al.* [1998], *Kavetski et al.* [2003], and the papers referenced earlier in this paragraph), especially in the face of the emerging generation of spatially distributed multi-input-output hydrologic models for which multiple (and often conflicting) sources of information are available for model calibration.

[4] Some interesting methods for addressing these problems, particularly in the context of estimating reasonable confidence bands on model simulations for conceptual nonlinear watershed models, have begun to appear in the literature. These methods include the use of classical Bayesian [*Kuczera and Parent*, 1998; *Thiemann et al.*, 2001; *Vrugt et al.*, 2003a; *Kavetski et al.*, 2003], pseudo-Bayesian [*Beven and Binley*, 1992; *Freer et al.*, 1996], set theoretic [*Keesman*, 1990; *Klepper et al.*, 1991; *van Straten and Keesman*, 1991; *Vrugt et al.*, 2001], and multiple criteria [*Gupta et al.*, 1998; *Yapo et al.*, 1998; *Boyle et al.*, 2000; *Vrugt et al.*, 2003b] methods to represent model parameter and prediction uncertainty. Each of these methods, however, uses one or more aggregate statistics of model performance over a large range of hydrologic behaviors, an action that is now understood to result in considerable loss of important information that can be used to distinguish between competing parameter sets. To increase the discriminative power of model calibration strategies, interest has therefore recently begun to switch toward recursive model identification strategies [e.g., *Thiemann et al.*, 2001; *Young*, 2001; *Vrugt et al.*, 2002; *Wagener et al.*, 2003], which sequentially move through a time series of discharge data and can provide estimates of parameter uncertainty. Such approaches can be used to localize the most informative measurements for parameter estimation [see *Vrugt et al.*, 2002], and also, perhaps more importantly, to provide a method for checking for violations of the underlying assumption that parameters are constant [*Wagener et al.*, 2003; *Misirli*, 2003; *Vrugt et al.*, 2003c]. Despite progress made, such methods still assign the uncertainty in the input-output representation primarily to a combination of uncertainty in the parameter estimates plus a residual "model" error term [see *Thiemann et al.*, 2001; *Gupta et al.*, 2003], and therefore lack the conceptual rigor needed to distinguish between all the important sources of uncertainty.

[5] In a separate line of research, considerable progress has been made in the development and application of sequential data assimilation (SDA) techniques. Such methods provide a general framework for explicitly dealing with input, output and model structural uncertainty, and for optimal merging of uncertain model predictions with observations. In contrast to classical model calibration strategies, SDA methods continuously update the states in the model when new measurements become available to improve the model forecast and evaluate the forecast accuracy. The prototype of the SDA methods, the Kalman filter (KF), was developed in the 1960s for optimal control of systems governed by linear equations and (as referenced earlier) was introduced into hydrology during the early 1980s to improve real-time forecasting of river discharges and to recursively estimate model parameters. Note that, for nonlinear dynamics, the extended Kalman filter (EKF) can in principle be used, which linearizes the error covariance equation using a tangent linear operator. However, because third- and higher-order moments in the closure scheme are discarded, this linearization is notoriously unstable if the nonlinearities are strong [*Evensen*, 1992; *Miller et al.*, 1994].

[6] Although the KF and EKF offer a very general framework for accounting for all sources of uncertainty, the filter typically assumes that the optimal values of the model parameters are known prior to the modeling procedure. This is particularly difficult in hydrologic modeling, where parameters often represent conceptual properties whose values can generally not be independently measured or assessed from lookup tables. This uncertainty in the correct choice of the parameter values, results in uncertainty in the estimated state values and therefore the model output prediction. To extend the usefulness and applicability of SDA methods to hydrologic modeling, it is important to explicate a formal methodology, which not only recursively updates model states but simultaneously also estimates model parameters. The major difference between such a framework and the aforementioned classical model calibration strategies is that this approach no longer assigns the primary uncertainty in the input-output representation to uncertainty in the parameters, but explicitly accounts for input, output and model structural errors during model calibration.

[7] This paper presents a simultaneous parameter optimization and data assimilation (SODA) method, which combines the strengths of the parameter search efficiency and explorative capabilities of the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm [*Vrugt et al.*, 2003a], with the power and computational efficiency of the ensemble Kalman filter [*Evensen*, 1994] to provide a better treatment of the input, output, parameter and model structural uncertainties in hydrologic modeling. The implementation of SODA requires a change in philosophy because it means that the model calibration procedure can no longer be considered as finding a single set of parameter values which generate the best possible long-term model forecasts. Instead, the optimal set of parameters identified with SODA only has meaning in combination with a SDA method, and consequently, will (per definition) not generate the best possible forecasts when no state adjustments are allowed, which is typically the case when evaluating the predictive capabilities of the model during an independent evaluation period. We hypothesize that the implementation of SODA will result in (1) meaningful prediction uncertainty bounds on the model simulations, (2) parameter estimates which better represent system properties that are less corrupted by modeling errors, thereby increasing the prospects of finding useful regionalization relationships, and (3) a time series of recursive state and output adjustments, whose interpretation will generate inspiration to improve our model concepts and as such our understanding of the functioning of hydrologic systems.

[8] This paper is organized as follows. Section 2 surveys the limitations of current model calibration strategies, which do not explicitly treat input, output and model structural uncertainty, but assign the uncertainty in the input-output representation primarily to uncertainty in the parameter estimates. In section 3 and 4 we discuss the rationale and architecture of SDA and global optimization methods, and demonstrate how the strengths of these methods can be combined to derive the hybrid SODA framework. In section 5 we illustrate the power and applicability of SODA by

means of a standard mathematical study using the three-parameter highly nonlinear Lorenz model [*Lorenz*, 1963], and a classical rainfall-runoff modeling example. In this section we also introduce a nonparametric variance estimator, which is especially designed to estimate the size of the measurement error of the output data. Finally, in section 6 we summarize the results. Note that although the SODA strategy results in considerable increase in computational burden, this increase is quite easily handled using currently available computing capabilities. The simple watershed model computation performed in this paper required approximately 4 hours on a Pentium 4 desktop personal computer. Application to more complex hydrologic models may, of course, possibly require taking advantage of parallel computing facilities.

2. Basic Inverse Formulation

[9] The fundamental problem with which we are concerned is to simultaneously estimate parameter values and state variables in a hydrologic model using a historical record of input-output data. The formulation of this resulting inverse problem can be expressed in a generic form if we assemble the state variables in the nonlinear hydrologic model at time t into the state vector ψ_t. The evolution of this state vector is described with

$$\psi_{t+1} = \eta(\psi_t, \tilde{X}_t, \theta), \tag{1}$$

where ψ is a vector of m unknown state variables, $\eta(\cdot)$ represents the nonlinear hydrologic model used to simulate the state evolution, $\tilde{X}$ is an observed forcing field, θ is a set of p model parameters, and t denotes time. We further assume that

$$\theta \in \Theta \subseteq \Re^p, \tag{2}$$

where $\Re^p$ denotes the p-dimensional Euclidean space. If Θ is not the entire domain space $\Re^p$, the inverse problem is said to be constrained. In hydrologic modeling, the feasible parameter space Θ can usually be properly constrained using realistic upper and lower bounds on each of the model parameters. Let $\tilde{Y} = \{\tilde{y}_1,\ldots,\tilde{y}_n\}$ denote the vector of streamflow measurement data available at time steps $1,\ldots,n$ and let $Y(\theta) = \{y_1(\theta),\ldots,y_n(\theta)\}$ represent the corresponding vector of model output predictions using the parameter values θ. These model output predictions are related to the model state according to

$$y_t = H(\psi_t), \tag{3}$$

where $H(\cdot)$ is the measurement operator, which maps the state space into the measurement or model output space. The tilde over X and Y emphasizes that these quantities are measured and hence subject to sampling and measurement error. The differences between the model-simulated output and measured data can be represented by the residual vector

$$E(\theta) = G[Y(\theta)] - G[\tilde{Y}] = \{e_1(\theta), \ldots, e_n(\theta)\}, \tag{4}$$

where the function $G(\cdot)$ allows for various user-selected linear or nonlinear transformations.

[10] The classical approach to estimating the parameters in equation (1) is to ignore input uncertainty ($X = \tilde{X}$) and to assume that the predictive model η is a correct, or at least accurate, representation of the underlying physical data-generating process. In line with classical statistical estimation theory, the residuals in equation (4) are then assumed to be mutually independent (uncorrelated), Gaussian distributed, with a constant variance. Under these circumstances, the traditional "best" parameter set in equation (1) can be found by minimizing the following lumped simple least square (SLS) objective function with respect to θ:

$$F_{\mathrm{SLS}}(\theta) = \sum_{t=1}^{n} e_t^2. \tag{5}$$

The limitations of neglecting input errors and lumping observed response and model structural error into one single white noise term, becomes immediately apparent when inspecting the time series of residuals after model calibration. To illustrate this, consider Figure 1, which displays a 5 month portion of the measured and simulated hydrographs for the Leaf River basin in Mississippi.

[11] The hydrologic model (HYMOD) conceptual watershed model was calibrated using the Shuffled Complex Evolution (SCE-UA) global optimization algorithm developed by *Duan et al.* [1992] using the SLS criterion and 1 year of calibration data. Notice, that the fit to the observed data can be considered quite good, but that the residuals exhibit considerable variation in bias (nonstationarity), variance (heteroscedasticity) and correlation structure under different hydrologic conditions. Indeed, the common approach of dealing with model structural and data errors, as being "small" or somehow "absorbed" into the output error residual needs serious reconsideration. Moreover, given the presence of these errors, overconditioning of the model to a single parameter set is unreasonable, and cannot be justified.

[12] One set of responses set forth to directly address the problem of overconditioning is to abandon the Frequentists approach of believing that the model parameters in equation (1) are fixed but unknown, and to adopt a Bayesian viewpoint which allows the identification of a plausible set of values for the parameters of the model given the available data. The Bayesian approach treats the model parameters in equation (1) as probabilistic variables having a joint posterior probability density function (pdf), which summarizes our probabilistic belief about the parameters θ in the light of the observed data $\tilde{Y}$. Examples of Bayesian approaches to hydrology include the Generalized Likelihood Uncertainty Estimation (GLUE) framework of *Beven and Binley* [1992] and the Bayesian Recursive Estimation (BaRE) approach of *Thiemann et al.* [2001] for representing model parameter and prediction uncertainty within the context of Monte Carlo analysis, and the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm of *Vrugt et al.* [2003a] for simultaneously estimating the traditional "best" parameter set along with a sample set of parameter values describing the probabilistic representation of the remaining parameter uncertainty (which is then used to generate probabilistic forecasts). We note that the SCEM-UA and BaRE methods operate within the context of the classical Bayesian framework (which makes implicit assumptions

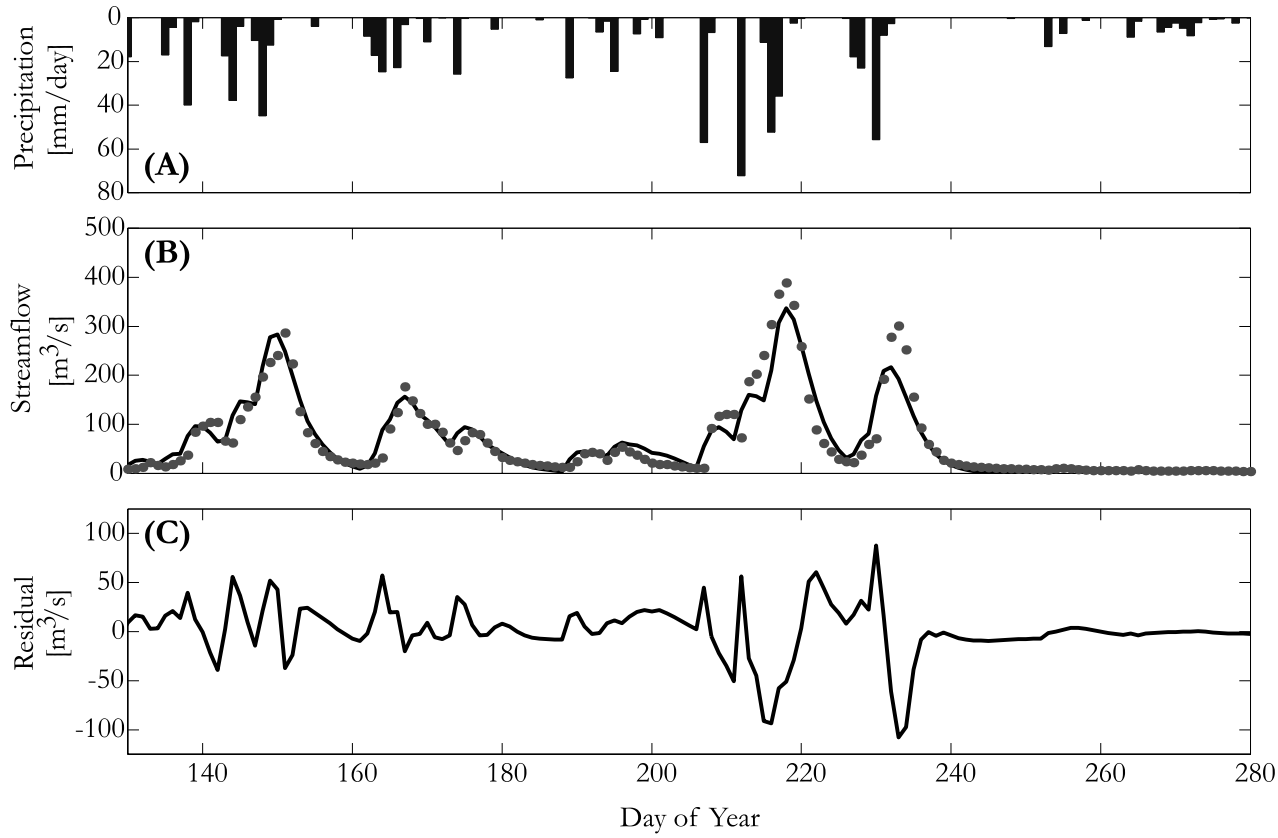

Figure 1. Streamflow predictions of the Shuffled Complex Evolution (SCE-UA)-calibrated hydrologic model (HYMOD) for a representative 5 month period for the Leaf River basin: (a) measured hyetograph; (b) actual model predictions. The solid circles correspond to the observed streamflow data. (c) Model residuals. See color version of this figure in the HTML.

about the correctness or adequacy of the model of the data generating process), while GLUE requires the user to make other kinds of subjective decisions including the selection of a cutoff threshold that separates behavioral from nonbehavioral parameter sets.

[13] An alternative set of responses set forth to identify model structural inadequacies, is to use Recursive Parameter Estimation algorithms, which provide better extraction of the information in the data, because the temporal aggregation associated with traditional batch processing is reduced. Examples of recursive algorithms which can be applied to complex nonlinear conceptual hydrologic models, include the PIMLI and recursive SCEM-UA algorithms [*Vrugt et al.*, 2002, 2003c], the DYNIA approach [*Wagener et al.*, 2003], and the BaRE algorithm [*Thiemann et al.*, 2001]. While these recursive methods have considerable advantages, their usefulness to help diagnose and quantify model structural errors has yet to be satisfactorily demonstrated. The novel instrumental variable techniques developed by *Young* [2001, 2002a, 2002b] does address this problem for the case of transfer function model structures by diagnosing time-varying parameters and showing how this information can be used to update the underlying model structure. However, this approach effectively neglects input and parameter uncertainty.

[14] A third set of responses set forth to more directly confront the problem of model structural errors is to pose the model calibration problem into a multiobjective framework, as was advocated by *Gupta et al.* [1998]. A major consequence of model structural imperfection is that the model is incapable of reproducing all aspects and portions of the output data with a single parameter set. By employing a number of complementary criteria in the optimization procedure, and analyzing the tradeoffs in the fitting of these criteria, the hydrologist is able to better understand the limitations of current hydrologic model structures, and gain insights into possible model improvements. The resulting Pareto solution set of parameters defines a kind of parameter uncertainty attributable to model structural errors. *Yapo et al.* [1998] and later *Vrugt et al.* [2003b] presented the MOCOM and MOSCEM-UA algorithms, respectively, capable of solving, in a single optimization run, the multi-objective problem posed by this approach. Unlike the previous methods, the multiobjective approach explicitly acknowledges the existence of model structural errors during the parameter estimation procedure. However, because of subjectivity in the selection of the measures [*Gupta et al.*, 1998] and data subsets [see *Boyle et al.*, 2000] to be included in the multiobjective vector, considerable fuzziness in the specification of the Pareto optimal parameter space and thus related model structural errors exists. Moreover, multiobjective calibration is typically response focused, and does not explicitly consider the influence of input errors [see *Kavetski et al.*, 2003]. Like the previous methods all sources of uncertainty are treated primarily as parameter uncertainty, giving rise to the concept of Pareto optimality.

[15] Finally, the last set of responses set forth to more directly confront the problem of input errors is to define realistic stochastic input error models. One novel approach is the recently proposed Bayesian Total Error Analysis (BATEA) of *Kavetski et al.* [2003], which explicitly takes into account input errors in the development of the likelihood function. However, it is not yet clear how this likelihood function can be easily extended to include model structural and output errors as well.

[16] In summary, current model calibration strategies are poorly suited to the task of explicitly treating the various important sources of uncertainty associated with the application of hydrologic models. In the next two sections of this paper, we will discuss the challenging question of how to develop a model calibration strategy that explicitly accounts for input, output and model structural errors.

3. Sequential Data Assimilation

[17] Sequential data assimilation methods, at least in principle, specify the uncertainty in the system states that arises from imperfect process representation and from input and output data uncertainty. To facilitate the description of the KF, we start by writing the model dynamics in equation (1) as a stochastic differential equation:

$$d\psi_t = \eta\left(\psi_{t-1}, \tilde{X}_{t-1}, \theta\right)dt + dq_{t-1}, \tag{6}$$

where $dq \in \Re^m$ is a dynamical noise term representing errors in the model formulation. This stochastic forcing term tends to flatten the probability density function of the model states during the integration. In section 3.2 of this paper we further elaborate on the simulation of q_t. The observation equation (3), also has a random additive error ε_t, called the measurement error:

$$y_t = H(\psi_t) + \varepsilon_t \qquad \varepsilon_t \sim N\left(0, \sigma_t^o\right), \tag{7}$$

where σ_t^o signifies the error deviation of the measurements. At each measurement time $t = 1,\ldots,n$ when an observation $\tilde{y}_t$ becomes available, compute the output forecast error, z_t:

$$z_t = \tilde{y}_t - H\left(\psi_t^f\right) \tag{8}$$

and update the forecasted states, ψ_t^f using the standard analysis equation:

$$\psi_t^a = \psi_t^f + K_t\left[\tilde{y}_t - H\left(\psi_t^f\right)\right], \tag{9}$$

where ψ_t^a is the updated or analyzed state, and K_t denotes a matrix of weights (Kalman gain), which is computed as

$$K_t = \sum\nolimits_t^m H^T\left[H\sum\nolimits_t^m H^T + \sum\nolimits_t^o\right]^{-1}, \tag{10}$$

where $\sum_t^m$ and $\sum_t^o$ denote the covariance matrices of the stochastic model error term ($= \overline{qq^T}$) and observations ($= \overline{\varepsilon\varepsilon^T}$), respectively. The difference between the forecasted and updated state is referred to in this paper as the state innovation, ψ_t^i, and is computed as

$$\psi_t^i = \psi_t^a - \psi_t^f, \tag{11}$$

which mapped in the output space will be called the output innovation, I_t:

$$I_t = H\left(\psi_t^a\right) - H\left(\psi_t^f\right). \tag{12}$$

The analyzed state ψ_t^a then recursively feeds the next state propagation step in the model:

$$\psi_{t+1}^f = \eta\left(\psi_t^a, \tilde{X}_t, \theta\right). \tag{13}$$

Since, uncertainty in the model structure and output data can be specified through the stochastic forcing term (q) and output measurement error (ε), respectively, and uncertainty in the input data can be taken into account through stochastic perturbations of the elements of the input ($\tilde{X}$), the KF offers a very general framework for dealing with all sources of uncertainty.

[18] Even though the KF or EKF implement a more sensible modeling strategy than traditional deterministic simulation methods by explicitly specifying the uncertainty in the system states that arises from imperfect process representation and from input and output data uncertainty, their widespread application appears to have been limited, both due to the strongly nonlinear effects of threshold type structures commonly found in hydrologic models, and due to the extra programming and heavy computational requirements associated with the storage and forward integration of the error covariance matrix. To resolve these two major problems, *Evensen* [1994] proposed the ensemble Kalman filter (EnKF), which uses a Monte Carlo (MC) method to generate an ensemble of model trajectories from which the time evolution of the probability density of the model states, and related error covariances are estimated. The EnKF avoids many of the problems associated with the traditional EKF method, e.g., there is no closure problem as is introduced in the EKF by neglecting contributions from higher-order statistical moments in the error covariance evolution. Moreover, the conceptual simplicity, relative ease of implementation and computational efficiency of the EnKF make the method an attractive option for data assimilation in the meteorologic, oceanographic and hydrologic sciences [e.g., *Evensen and van Leeuwen*, 1996; *Houtekamer and Mitchell*, 1998; *Lermusiaux*, 1999; *Madsen and Cañizares*, 1999; *Keppenne*, 2000; *Reichle et al.*, 2002].

3.1. Ensemble Kalman Filter

[19] The EnKF uses an ensemble of model trajectories to solve for equations (6)–(13). The description of the basic algorithm of the EnKF below follows *Evensen* [1994] and is further illustrated in Figure 2.

[20] 1. Generate initial ensemble. Sample N combinations of m model states, ψ_j, $j = 1,\ldots,N$ randomly from the prior distribution and store them in a matrix $A[1{:}m,1{:}N]$:

$$A = \left(\psi^1, \ldots, \psi^N\right) \in \Re^{mxN}. \tag{14}$$

[21] 2. Forecast step. Propagate each of the N ensemble members of A forward in time with the nonlinear model $\eta(\cdot)$, using a fixed set of model parameters θ and a corresponding ensemble of N stochastic forcing fields:

$$\psi_t^j = \eta\left(\psi_{t-1}^j, \tilde{X}_{t-1}, \theta\right) + q_{t-1}^j. \tag{15}$$

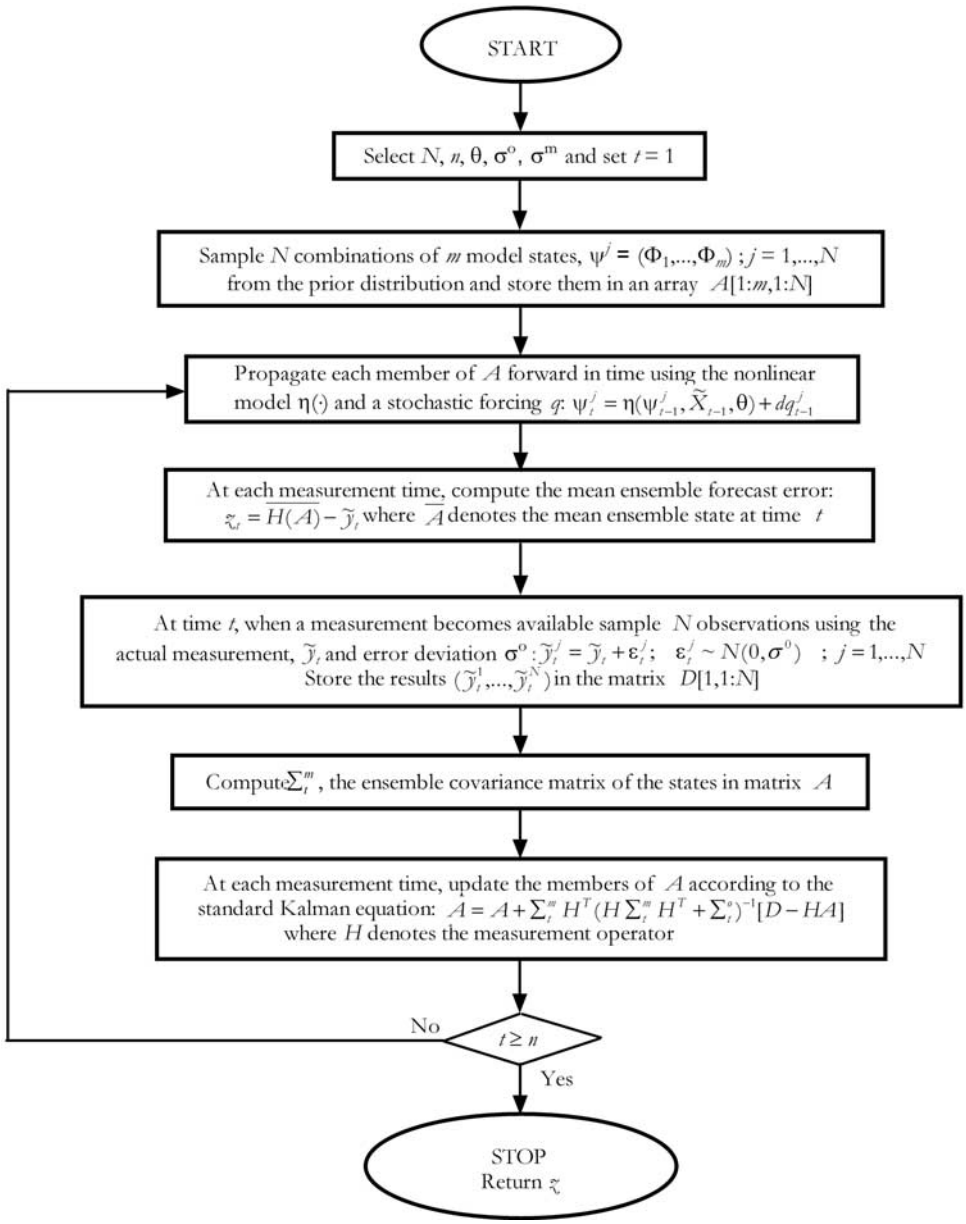

Figure 2. Flowchart of the ensemble Kalman filter (EnKF) used to recursively estimate state variables.

Explicit details on how to simulate q_{t-1} will be given in the next section.

[22] 3. Compute the covariance matrix of the forecasted states. At each time step t compute the covariance matrix, $\sum^m \in \Re^{mxm}$ of the ensemble forecasted states in matrix A:

$$\sum_t^m = \frac{A'(A')^T}{N-1}, \tag{16}$$

where $A' = A - \overline{A}$ and $\overline{A}$ denotes the ensemble mean of the forecasted states.

[23] 4. Compute the mean ensemble forecast error. At time t when a measurement become available, compute the mean forecast error of the ensemble:

$$z_t = \overline{H(A)} - \tilde{y}_t. \tag{17}$$

[24] 5. Generate vector of observations. At each measurement time, generate an ensemble of N observations from a distribution with mean equal to the observation, $\tilde{y}_t$, and covariance equal to $\sum_t^o$,

$$\begin{aligned} \tilde{y}_t^j &= \tilde{y}_t + \varepsilon_t^j, \\ \varepsilon_t^j &\sim N\left(0, \sigma_t^o\right) \qquad j = 1, \ldots, N, \\ \textstyle\sum_t^o &= \overline{\varepsilon_t \varepsilon_t^T}, \end{aligned} \tag{18}$$

and store them in a matrix $D[1,1{:}N]$:

$$D = \left(\tilde{y}^1, \ldots, \tilde{y}^N\right) \in \Re^{1xN}. \tag{19}$$

This particular step ensures that the spread of the updated ensemble is consistent with the true estimation error covariance [*Burgers et al.*, 1998].

[25] 6. State update or analysis step. At each measurement time, update each ensemble member of A using the following analysis equation:

$$A = A + \sum_t^m H^T \left(H \sum_t^m H^T + \sum_t^o \right)^{-1} [D - H(A)], \quad (20)$$

where the product $\sum^m H^T (H \sum^m H^T + \sum^o)^{-1}$ is a numerical approximation of the Kalman gain, K_t in equation (9).

[26] 7. Check stop criterion. If t is less than the total number of time steps, n, return to step 2, otherwise stop.

[27] In summary, the EnKF uses a large ensemble of model trajectories to approximate the probability density of the model states at each time step. The mean of this ensemble represents the "best" state estimate, whereas the variance provides a measure of the spread of the ensemble members (uncertainty). For more information about the derivation of the EnKF and its practical implementation, please refer to *Evensen* [1994].

3.2. Simulation of the Stochastic Model Error Term

[28] To successfully implement the EnKF it is particularly important to specify accurate values of the output measurement error and stochastic forcing term used to simulate the time evolution of dynamical model noise. Hence it is not difficult to see that the size of these terms directly controls the spread among the ensemble members, and as such the mean ensemble model forecasts and associated prediction uncertainty bounds. Fortunately, in most situations a reasonable a priori estimate can be made of the measurement error, σ^o, of the observations. A more delicate issue however, is how to retrieve a reasonable estimate of the stochastic forcing term. This term reflects many sources of uncertainty including model discretization errors and neglected dynamical processes.

[29] On the basis of recommendations in the classic treatise of the EnKF in *Evensen* [1994] we used the following equation to simulate the time evolution of the stochastic forcing:

$$s_t = \phi s_{t-1} + \sqrt{1 - \phi^2} w_{t-1}, \quad (21)$$

where w_{t-1} is an $m \times 1$ vector of draws from a normal distribution, and $\phi \in [0,1)$ represents the first-order decorrelation of the stochastic forcing. The form of equation (21) ensures that, with an increasing number of elements, s will approach a distribution having mean equal to zero and variance equal to one. A value for $\phi = 0$ generates a sequence of white noise terms, while $\phi = 1$ will remove the stochastic forcing. Note that more complex error models could (in principle) be used, and the order of the model determined as part of the identification procedure [e.g., *Young*, 2001], but such procedures will not be pursued in this paper. On the basis of random walk theory, the nonlinear model $\eta(\cdot)$ can then be written as

$$\psi_t = \eta(\psi_{t-1}, \tilde{X}_{t-1}, \theta) + \tau s_{t-1} \sqrt{\Delta t} \sigma, \quad (22)$$

in which Δt denotes the time step, σ is an $m \times 1$ vector of error deviations of individual states and τ is a factor which is computed as [*Evensen*, 1994]

$$\tau = \sqrt{\frac{1}{\Delta t} \frac{(1-\phi)^2}{l - 2\phi - l\phi^2 + 2\phi^{(l+1)}}}, \quad (23)$$

where l denotes the number of steps in each time unit Δt ($l \Delta t = 1$).

[30] While this equation is frequently used in meteorology and oceanography to simulate the time evolution of the stochastic forcing, our attempt to apply this equation to conceptual hydrologic modeling has revealed one major drawback. When the states in the model do not have a clear physical meaning, as is typically the case with conceptual hydrologic models, it is difficult to specify accurate values of the error deviations of individual states. To circumvent this problem, we specify the error deviation in the output space of the model. It can be shown that the error deviation in the state space is related to the output error deviation σ^m (also referred to in the remainder of this paper as the error deviation of stochastic forcing) by the relationship $\sigma = \sum_t^m H [H \sum_t^m H^T]^{-1} \sqrt{\Delta t}\, \sigma_t^m$. Therefore we adjust the states in the model as follows:

$$\psi_t = \eta(\psi_{t-1}, \tilde{X}_{t-1}, \theta) + \tau s_{t,1} \sum_t^m H \left[H \sum_t^m H^T \right]^{-1} \sqrt{\Delta t} \sigma_t^m, \quad (24)$$

where $\sum_t^m$ represents the covariance matrix of the forecasted states. Among various alternatives, this implementation was shown to work well for simulating the time evolution of the stochastic forcing for a variety of test problems.

4. Combining Global Optimization and Sequential Data Assimilation

[31] Even though the EnKF provides a general framework to account for input, output, and model structural uncertainty, a defining characteristic of the filter is its focus on time recursive assimilation of the output measurements to update estimates of the state variables, while the parameters are fixed at predefined values, thereby explicitly ignoring the effects of parameter uncertainty and interaction. In contrast, parameter estimation using a global optimization method does not use the output data to recursively update the state variables, but uses batch processing of the data to search for the parameter estimates that minimize the overall (statistical) variance of the model residuals, thereby improving the prediction accuracy in an overall statistical sense. This section discusses how the strengths of a global optimization method can be merged with EnKF state estimation to derive a hybrid framework, which recursively tracks the model states, while simultaneously estimating the values of the model parameters (and their associated uncertainty). This hybrid framework, simultaneous optimization and data assimilation (SODA), consists of an inner EnKF loop for recursive state estimation conditioned on an assumed parameter set, and an outer global optimization loop for batch estimation of the high probability region of the posterior density of the parameters.

[32] To implement SODA, we must first specify a mathematical criterion that measures the "closeness" between the EnKF-derived mean ensemble model forecast and the corresponding measurement, which can be used to extract the information content in the data and transform it into estimates for the model parameters. This issue will be addressed in this paragraph. Second, we must implement a robust global optimization method to effectively and efficiently estimate the posterior parameter density region,

thereby generating the desired estimate of parameter uncertainty. We will tackle this problem in the next paragraph. To estimate the model parameters, we adopt a classical Bayesian approach using the following posterior density criterion, $p(\theta|\tilde{Y})$ [*Box and Tiao*, 1973]:

$$p(\theta|\tilde{Y}) \propto \left(\sum_{t=1}^{n} [z_t(\theta)]^2 \right)^{-\frac{1}{2}n}, \quad (25)$$

in which $z(\theta)$ denotes the time series of mean ensemble forecast errors corresponding to the parameter set θ. Using this criterion, the best set of model parameters is defined as that set which is on average associated with the best one-observation-ahead model forecasts.

[33] To generate samples from equation (25), and to summarize the posterior parameter pdf using statistical moments and histograms, we use an implementation of the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm. The SCEM-UA algorithm is a general purpose global optimization algorithm that provides an efficient estimate of the most likely parameter set and its underlying posterior probability distribution within a single optimization run [see *Vrugt et al.*, 2003a]. The algorithm is a Markov Chain Monte Carlo (MCMC) sampler, which generates a sequence of parameter sets $\{\theta^{(1)},\theta^{(2)},\ldots,\theta^{(k+1)}\}$ that convergences to the stationary posterior distribution for a large enough number of simulations k. The SCEM-UA algorithm is related to the successful SCE-UA global optimization method, but uses the Metropolis-Hastings (MH) search strategy [*Metropolis et al.*, 1953; *Hastings*, 1970] instead of the downhill simplex method for population evolution, and is therefore able to simultaneously infer both the most likely parameter set and its underlying posterior probability distribution within a single optimization run. A detailed description and explanation of the method appears in *Vrugt et al.* [2003a] and so will not be repeated here.

[34] In brief, the SCEM-UA method involves the initial selection of a population of points distributed randomly throughout the p-dimensional feasible parameter space. In the absence of prior information about the location of the maximum likelihood value a uniform sampling distribution is used. For each point the complete system of equations of the EnKF, as outlined in steps 1–7 in section 3.1, is executed and the time series of mean ensemble forecast errors returned to the SCEM-UA algorithm. This time series is subsequently inserted in equation (25) to calculate the posterior density for each point. The population of parameter sets is subsequently partitioned into a number of complexes, and in each complex a parallel sequence is launched from the point that exhibits the highest posterior density. A new candidate point in each sequence is generated using a multivariate normal distribution either centered on the current draw of the sequence or the mean of the points in the complex augmented with the covariance structure induced between the points in the complex. The Metropolis annealing [*Metropolis et al.*, 1953] criterion is used to test whether the candidate point should be added to the current sequence. Finally, the new candidate point is shuffled into the original population of complexes. The evolution and shuffling procedures are repeated until the Gelman-Rubin convergence diagnostic for each of the parameters demonstrates convergence to a stationary posterior target distribution [*Gelman and Rubin*, 1992]. Experiments conducted using standard mathematical test problems have shown that the SCEM-UA-derived posterior distribution closely approximates the target distribution [*Vrugt et al.*, 2003a].

[35] The overriding characteristic of the SODA methodology is the improved treatment of input, output, parameter and model structural uncertainties, during model calibration. Theoretically, if the model structure is correct, and the input and output data are observed without error, no state adjustments are needed during model calibration, and the set of parameters that maximizes the posterior density in equation (25), will also be associated with, on average, the smallest output residuals, when performing traditional batch calibration without state adjustments (as done in section 2). However, because of errors in the model structure and input (forcing) and output data this will generally not be the case. The major objective of this paper is to demonstrate how the hybrid SODA strategy can be used to simultaneously estimate model parameters and state variables, resulting in reliable model prediction uncertainty bounds and a time series of state and output innovations whose interpretation will generate inspiration to improve our model concepts and as such our understanding of the functioning of hydrologic systems.

5. Case Studies

[36] We illustrate the power and applicability of SODA by means of two different case studies. The first is a classical mathematical study using the three-parameter highly nonlinear Lorenz model [*Lorenz*, 1963], to demonstrate the ability of SODA to simultaneously estimate state variables and parameter values when confronted with highly nonlinear model dynamics. The second case study explores the usefulness of SODA by application to hydrologic modeling using the simple HYMOD conceptual watershed model and historical streamflow data from the Leaf River watershed in Mississippi. In this study we are especially concerned with the estimation of model structural errors and present a novel nonparametric method, which is especially designed to estimate the measurement error of streamflow data.

5.1. Case Study 1: A Simple Demonstration Using the Lorenz Equations

[37] To demonstrate that SODA is indeed able to simultaneously estimate state variables and model parameters when confronted with nonlinear model dynamics, the first case study considers the three-parameter Lorenz model [*Lorenz*, 1963]. This model consists of a system of three nonlinear and coupled ordinary differential equations:

$$\begin{aligned} \frac{dx}{dt} &= \gamma(y - x), \\ \frac{dy}{dt} &= \rho x - y - xz, \\ \frac{dz}{dt} &= xy - \beta z, \end{aligned} \quad (26)$$

where $x(t)$, $y(t)$, and $z(t)$ denote the dependent variables and γ, ρ, and β represent model parameters. This model is a relatively simple example, being nonlinear in the state dynamics but linear in the parameters, and is interesting to

our study because of the effects of the interactions between the parameters and states. Because of its nonlinear nature, this model has served as a test bed in the field of data assimilation for examining the properties for various data assimilation methods. A reference solution of $(x(t), y(t), z(t))$ with output print step $\delta t = 0.25$ was computed for $t \in [0,40]$ by solving equation (26) starting from the initial condition $(x_0, y_0, z_0) = (1.508870, -1.531271, 25.46091)$ and fixed values for the parameters of $\gamma = 10$, $\rho = 28$, and $\beta = 8/3$. Synthetic observations were subsequently computed by imposing normally distributed noise, with mean zero and variance equal to 2.0, to the x, y, and z values in this reference solution.

[38] The error covariance of the stochastic forcing is defined to be diagonal with variances equal to 2.00, 12.13, and 12.31 for the three equations in (26) respectively, and is included in the model through the stochastic forcing term in equation (22). The error variances of the initial conditions and observations were taken to be the same values as those used for generating the synthetic observations. The values for the parameters, initial conditions, and error variances used in the setup of this case study are identical to those used in similar studies reported in the literature [see, e.g., *Evensen*, 1994; *Miller et al.*, 1994; *Evensen*, 1997; *Evensen and van Leeuwen*, 2000].

[39] The SODA algorithm was used to simultaneously estimate the state variables and model parameters using the synthetically generated observations and the density criterion in equation (25). A search population of 50 points in combination with five parallel sequences was selected based on recommendations in our previous work. The feasible parameter space was taken to be a uniform hypercube between [0–30], [0–50], and [0–10] for the parameters γ, ρ, and β, respectively. Furthermore, an ensemble of 100 members was used to compute the error statistics with the EnKF. Experiments with larger ensemble sizes (not shown here) gave only marginal improvements in the approximation of the probability density of the model states. The estimated ensemble prediction uncertainty ranges associated with the computed x values in equation (26) and corresponding to the most likely SODA identified parameter set is shown in Figure 3a. The solid circles correspond to the observed x values, whereas the dashed black line represents the evolution of the mean ensemble prediction. Note that the prediction uncertainty ranges generally bracket the observations very well, indicating that the EnKF does a good job in tracking the state transitions. Further, the width of the ensemble prediction uncertainty bounds is consistent with the specified error variance of the model and observations.

[40] The transitions of the parameters γ, ρ, and β in each of the five sequences (Markov chains) during the evolution of SODA to the stationary posterior distribution is illustrated in the Figures 3b–3d respectively.

[41] For clarity, the three different parallel sequences are coded with different symbols. The asterisks at the right-hand side indicate the "true" values of the parameters. The 1-D scatterplots of the sampled parameter space demonstrate that at early stages during the evolution, the individual sequences tend to occupy different regions of the parameter space. At a later stage during the evolution, however, all of the individual sequences have been able to fully explore the feasible parameter space, and indeed they assign the highest probability to the "true" parameter values used for generating the synthetic observations. All in all, we can conclude that SODA is able to simultaneously estimate state variables and model parameters when confronted with nonlinear model dynamics, so that we can proceed with the next case study.

5.2. Case Study 2: The HYMOD Conceptual Watershed Model

[42] We illustrate the power and applicability of the SODA algorithm to hydrologic modeling by application to the HYMOD conceptual watershed model using historical data from the Leaf River watershed (1950 km^2) located north of Collins, Mississippi. The data, obtained from the Hydrologic Research Laboratory (HL), consists of mean areal precipitation (mm/day), potential evapotranspiration (mm/day), and streamflow (m^3/s). Forty consecutive years of data (1948–1988) are available for this watershed, representing a wide variety of hydrologic conditions. The HYMOD model, consisting of a relatively simple rainfall excess model, described in detail by *Moore* [1985], connected with two series of linear reservoirs (three identical quick and a single reservoir for the slow response). A schematic overview of HYMOD appears in Figure 4.

[43] The model has five parameters: the maximum storage capacity in the catchment, C_{max} (L), the degree of spatial variability of soil moisture capacity within the catchment, b_{exp}, the factor distributing the flow between the two series of reservoirs, Alpha, and the residence time of the linear quick and slow flow reservoirs, R_q (days) and R_s (days), respectively. The upper and lower bounds that define the prior uncertainty ranges of these parameters appear in Table 1.

[44] Because the HYMOD model and the Leaf River data have been discussed extensively in previous work [e.g., *Yapo et al.*, 1998; *Boyle et al.*, 2000; *Vrugt et al.*, 2002; *Misirli*, 2003], we will not describe the details of these here. To implement the SODA framework outlined above, it is further necessary to specify an appropriate value for the parameter ϕ in equation (21). As our knowledge about a correct value for this first-order decorrelation coefficient is quite poor, we added this parameter to the optimization problem, using a feasible range between [0] and [1]. On the basis of our experience with the model we used 3 years of streamflow data (28 July 1952–1955) for calibration purposes, to illustrate that the SODA method is relatively simple to implement and helps to distinguish between the various sources of uncertainty in hydrologic modeling. An additional data period of approximately 5 years (29 July 1955–30 September 1960) was subsequently used for evaluation purposes. To generate the one-step-ahead streamflow forecasts with the HYMOD model during the calibration and evaluation period we assumed full knowledge of the rainfall time series by using measured rainfall amounts of the next day.

5.2.1. Derivation of the Measurement Error of Streamflow Data

[45] To solve the model calibration problem using SODA, we must do two things. First, a realistic estimate of the error deviation of the output (streamflow) data must be retrieved. This issue will be addressed in this section. Second, we must find a way to obtain a reasonable estimate of input and model structural uncertainty. We will tackle this problem in the next section.

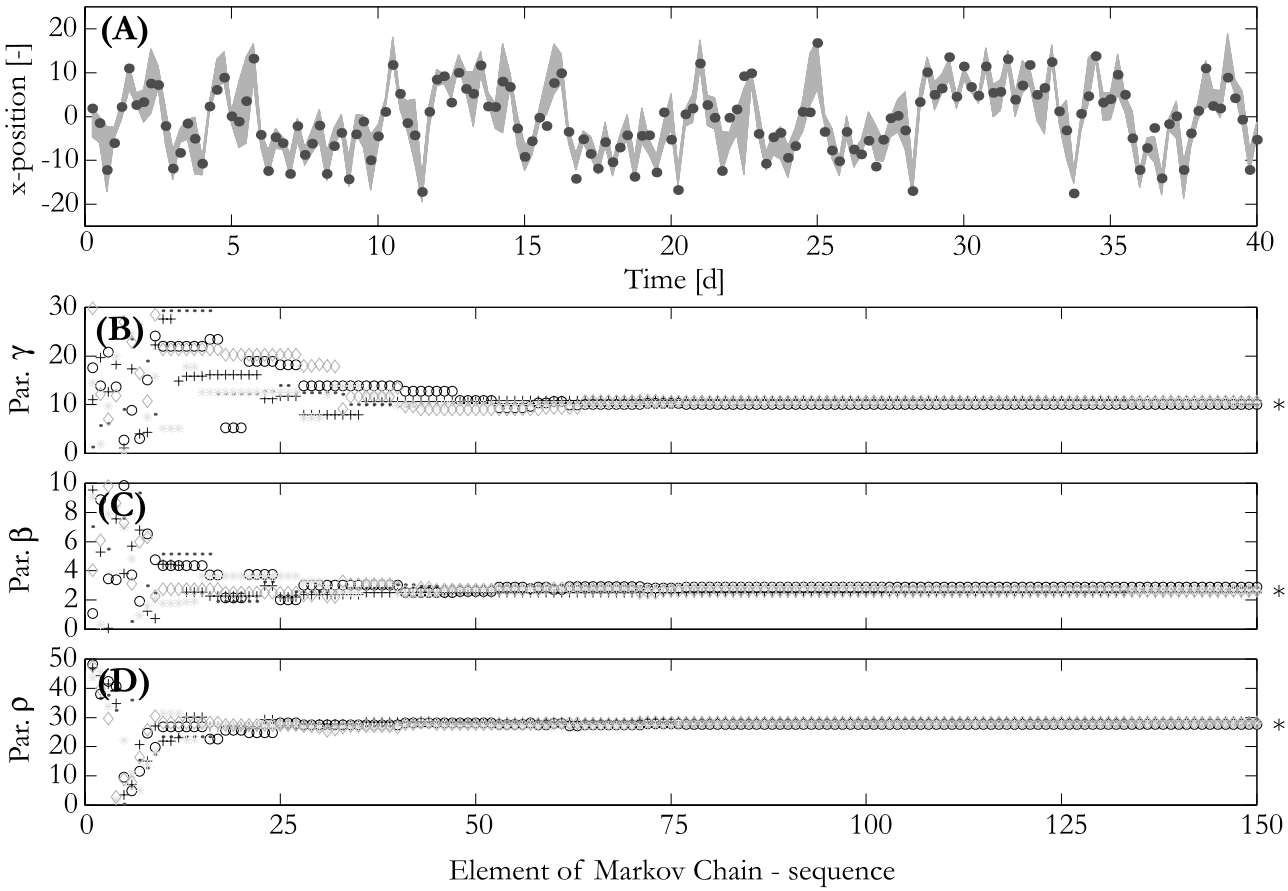

Figure 3. Results obtained with the simultaneous optimization and data assimilation (SODA) method for the three-parameter highly nonlinear Lorenz model. (a) Prediction uncertainty ranges corresponding to the best parameter set identified with SODA and 100 ensemble members. The solid dots denote the observations. (b–d) Sampled γ, β, and ρ parameters, respectively, in five different sequences (color coded with different symbols) during the evolution of SODA to the posterior target distribution. The asterisks at the right-hand side of these figures denote the "true" parameter values used to generate the synthetic observations. See color version of this figure in the HTML.

[46] Unfortunately, the problem of estimating the measurement error of streamflow data has not proved to be simple. To further elaborate on this problem, consider observations obtained through

$$\tilde{y}_t = \hbar(t) + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_t^o\right), \qquad (t = 1, \ldots, n), \qquad (27)$$

where $\hbar(t)$ is the actual streamflow at time t and the errors ε_t are independent random variables with zero mean and unknown variance, σ_t^o. Without any doubt, most attention in hydrologic modeling has been given to a correct estimation of the function $\hbar$. However, estimation of the error deviation is nearly as important as the estimation of $\hbar$ itself, to obtain reasonable confidence intervals on the model predictions.

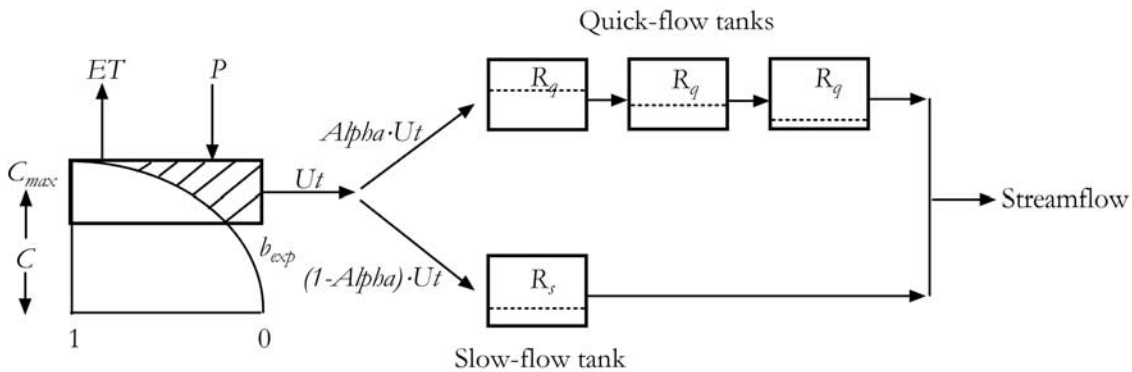

Figure 4. Schematic of the HYMOD conceptual watershed model.

Table 1. Prior Uncertainty Ranges and Description of the Hydrologic Model (HYMOD) Parameters

Parameter	Description	Minimum	Maximum	Unit
C_{max}	maximum storage watershed	200.00	500.00	mm
b_{exp}	spatial variability soil moisture capacity	0.10	2.00	
Alpha	distribution factor between two reservoirs	0.50	0.99	
R_s	residence time slow flow reservoir	0.00	0.10	days
R_q	residence time quick flow reservoir	0.30	0.70	days

[47] During the last two decades several nonparametric methods have been proposed which estimate the error deviation in equation (27) [see, e.g., *Rice*, 1984; *Hall et al.*, 1990; *Seifert et al.*, 1993; *Dette et al.*, 1998]. All these methods involve differencing the original time series, $\tilde{y}_t$, and estimating the error deviation as

$$\hat{\sigma}_t^o = \sqrt{\frac{1}{2(n-1)} \sum_{t=2}^{n} (\tilde{y}_t - \tilde{y}_{t-1})^2}. \quad (28)$$

The underlying assumptions of this approach include that (1) $\hbar(t)$ as a function of t is sufficiently smooth, (2) the sampling interval is high compared to the typical timescale of $\hbar(t)$, and (3) the error terms, ε_t, exhibit a constant variance (homoscedastic). While the first two assumptions can be considered quite reasonable for a time series of daily streamflow data, the assumption of homoscedasticity of the streamflow error terms is not very realistic. For instance, *Sorooshian and Dracup* [1980] commented that large flows tend to have larger error variances compared to smaller flows, partially because of the nonlinear nature of the rating curve used to transform stage measurements to flow volume estimates. To enable the identification of these nonhomogeneous (heteroscedastic) streamflow errors, we decided not to compute an average standard error estimate of the total time series, as suggested in equation (28), but instead, to apply the following nonparametric error deviation estimator locally in the time series

$$\hat{\sigma}_t^o = \sqrt{\binom{2u}{u}^{-1} (\Delta^u \tilde{y}_t)^2}, \quad (29)$$

where Δ^u denotes the difference operator applied u times. It can be readily verified that the estimator in equation (29) is insensitive to polynomial trends in $\hbar(t)$ up to order u. In the literature more sophisticated higher-order differencing procedures have been proposed [*Hall et al.*, 1990]. However, investigations with numerically generated streamflow data showed that equation (29) with the choice $u = 3$ works well in practice.

[48] To demonstrate the validity of the proposed approach for estimating the error deviation of streamflow data, we synthetically generated two different test problems. Synthetic daily streamflow data, $\{y_1,\ldots,y_n\}$ were first generated for the period 28 June 1952–30 September 1960, by driving the HYMOD model with observed mean areal rainfall of the Leaf River watershed and values for the parameters identical to those used in Figure 1. This output time series was subsequently corrupted by adding one of the two following error models to the data to yield two different time series of "observed" streamflow data; (1) a homoscedastic error model: $\varepsilon_t \sim N(0,10)$; and (2) a heteroscedastic error model: $\varepsilon_t \sim N(y_t,0.1y_t)$. Both these time series were subsequently used in conjunction with equation (29) to estimate the error deviation of the streamflow measurements. The results of this analysis are summarized in Figure 5, which presents scatterplots of "observed" streamflow data versus estimated error deviations for the (Figure 5a) homoscedastic and (Figure 5b) heteroscedastic error case.

[49] To further facilitate graphic interpretation of the results, the dark black line in each of the figures represents the optimal fit of a spline function through the data, whereas the dashed line corresponds to the original error model used to corrupt the streamflow data. Notice, that the agreement between the predicted and actual error model can be considered quite excellent. This is especially true for the homoscedastic error case. Although not further reported here, we performed a variety of synthetic experiments with different error distributions, and found that the nonparametric estimator in equation (29) provides a close approximation to the actual error model, when the time series of streamflow data is sufficiently long (>2 years of daily data).

[50] To unravel the relationship between flow level and measurement error for the Leaf River watershed, Figure 6, presents a similar scatterplot as Figure 5 for this real world data set. Again, the black line denotes the optimal fit of a spline function through the scattered data using nonlinear regression.

[51] The pattern of the scatterplot and slope of the regression function clearly indicate the presence of heteroscedastic errors in the streamflow data. In general, the size of the streamflow error almost log-linearly increases from 0.01 to about 80 m^3/s in the flow range of 2 to 1000 m^3/s respectively. These results provide strong support for the claim by *Sorooshian and Dracup* [1980], that streamflow data exhibit nonhomogeneous (heteroscedastic) errors. To optimally exploit the information contained in the scatterplot of Figure 6, it would seem most productive to first cluster the data into various flow levels, and then to approximate the corresponding error probabilities with Gaussian mixture models. However, because the purpose of this case study is illustrative, we decided not to follow such an approach, but rather to implement the optimal fit nonlinear regression function (see Figure 6) into SODA, thereby relating the error deviation of the streamflow data to the flow level.

5.2.2. Derivation of Input Error Model and Error Deviation of the Stochastic Forcing

[52] In the absence of a compelling basis for the assignment of a reasonable input error model to stochastically perturb the elements of $\tilde{X}$, we decided to merge input and model structural errors into a single forcing term. The approach that we propose here to estimating the total error deviation of this term, σ^m in equation (24), is to conduct a classical deterministic calibration of the model, and to use the resulting model residuals (total error) for the subsequent

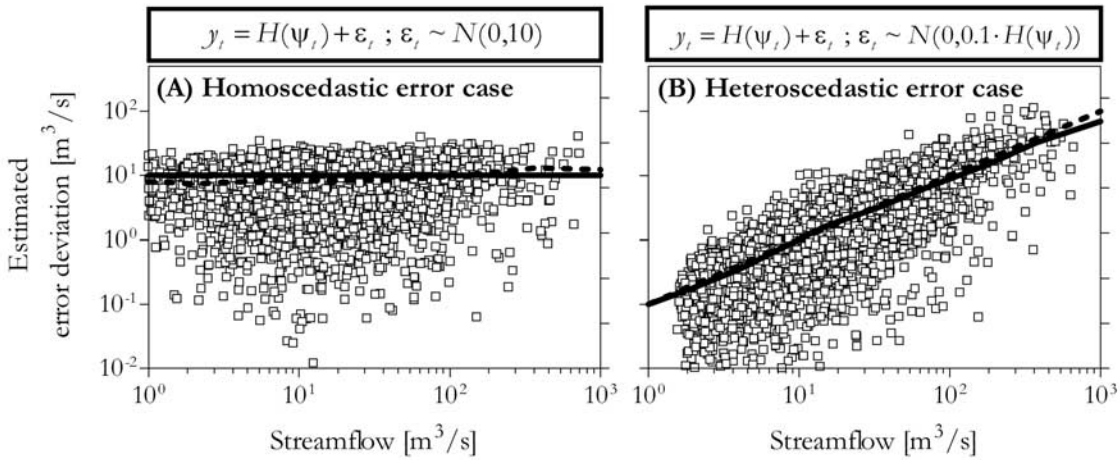

Figure 5. Synthetic case studies to demonstrate the validity of the nonparametric estimator to derive the error deviation of streamflow measurements. (a) Homoscedastic error case. (b) Heteroscedastic error case. The solid and dashed black lines in each of the figures denote the estimated and actual error models, respectively.

error analysis and decomposition into component terms as described below.

[53] If we proceed with classical model calibration the output is a set of residuals which constitute a combination of model structural, input and output errors (see Figure 1). The difference between this residual and the associated measurement error of the observation represents an estimate of the combined effect of model structural and input error. To illustrate this, consider Figure 7, which presents a scatterplot of the error deviation of the residuals of the previously calibrated HYMOD model (see section 2) against flow level, created using 10 years of streamflow data for the Leaf River watershed.

[54] The dashed black line denotes the estimated measurement error model (see Figure 6), whereas the solid line represents the optimal fit of a spline function through the scattered data. Accordingly, the size of the model structural and input error can be estimated from the difference between the solid and dashed black lines. A spline function was fitted through this difference function and subsequently implemented in SODA to relate the error standard deviation of the stochastic forcing to flow level.

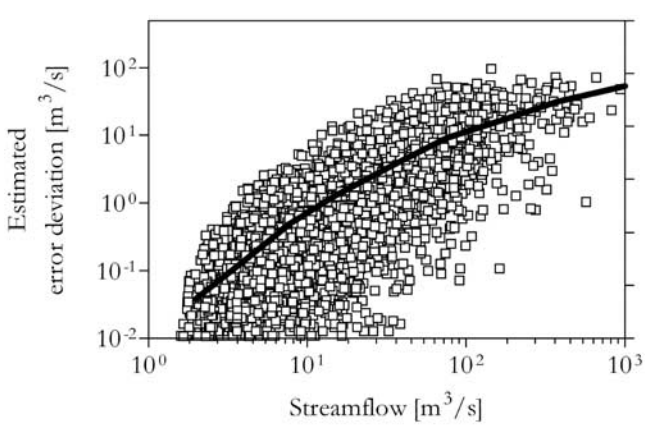

Figure 6. Scatterplot of observed streamflow data versus estimated error deviation of the measurements using 10 years of data from the Leaf River watershed. The solid line represents the fit of a spline function through the data.

5.3. Use of SODA for Hydrologic Model Calibration: Model Precision and Accuracy

[55] The SODA algorithm was used with the particular settings discussed in the previous two sections to simultaneously estimate the HYMOD state variables and model parameters using the daily streamflow data of the Leaf River basin and the density criterion specified in equation (25). A search population of 250 points in combination with five parallel sequences and 100 ensemble members was selected. The procedure used about 1500 parameter evaluations to converge to a reasonable estimate of the optimal model parameters, including their underlying posterior distribution. The results of this analysis are summarized in Table 2 and Figures 8–10 and are discussed below.

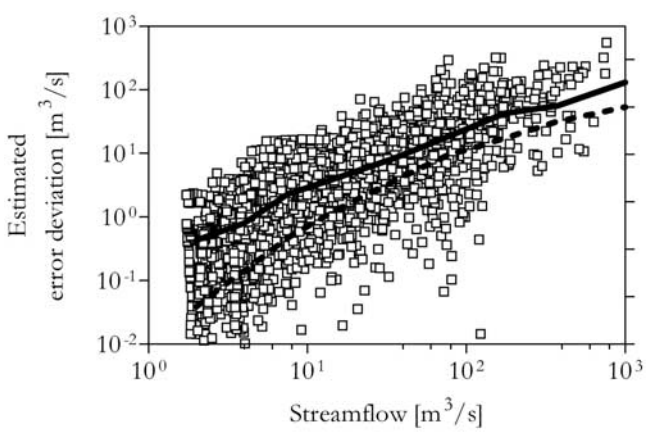

Figure 7. Scatterplot of the computed error deviation of the residuals of the calibrated HYMOD model against flow level for a 10 year period of daily streamflow data for the Leaf River watershed. The solid and dashed black lines represent the best possible fit of a spline function through the scattered data and the previously estimated error model of the measurements, respectively.

Table 2. Summary Statistics (RMSE, CORR, and BIAS) of the 1-Day-Ahead Streamflow Forecasts Using the Shuffled Complex Evolution (SCEM-UA) and Simultaneous Optimization and Data Assimilation (SODA) Methods

	Calibration (WY 1952–1955)			Validation (WY 1955–1960)		
Statistics	RMSE	CORR	BIAS, %	RMSE	CORR	BIAS, %
SCEM-UA	17.25	0.903	−5.04	20.48	0.88	−4.65
SODA	13.14	0.96	0.65	14.32	0.95	0.82

[56] The usefulness of the implementation of SODA can be demonstrated in a number of different ways. One of the most straightforward is to compare the hydrograph prediction uncertainty ranges of the HYMOD model using SODA, with those obtained using a classical Bayesian SCEM-UA calibration, which ignores input uncertainty and lumps output and model error into a single white noise term. The size of the prediction uncertainty ranges is a measure of the precision of the model. Figure 8 presents the results of this analysis.

[57] The hydrograph prediction uncertainty ranges (lightly shaded region) of the HYMOD model corresponding to the probability distribution of SODA identified parameter sets, each using an ensemble size of 100 members for state estimation, is presented in Figure 8a. The observed streamflows are indicated with solid circles, the parameter uncertainty is indicated by the darkly shaded region, the total uncertainty (parameter, structure plus input, and output) is indicated by the lightly shaded region, and the prediction corresponding to SODA identified "best" parameter set is indicated with the dashed black line. Figure 8b presents the prediction uncertainty ranges for the HYMOD simulated streamflows corresponding to only the SCEM-UA-derived posterior parameter estimates. From visual inspection of the two plots we see (as expected) that the input-output behavior of the HYMOD model when properly accounting for input, output and model structural errors, is more consistent with the observations. Hence in the case of SODA, the prediction uncertainty ranges are reasonably small and bracket the observations. On the contrary, when applying a Bayesian approach (Figure 8b), without state adjustments, the model seems to be unable to match large portions of the hydrograph; this is indicated by sections where the darkly shaded region does not bracket the observed streamflow data. Although one might argue that this is a problem of overconditioning, and that better results can be obtained by assigning more relaxed cutoff thresholds to determine what can be considered an acceptable parameter set or not (as is

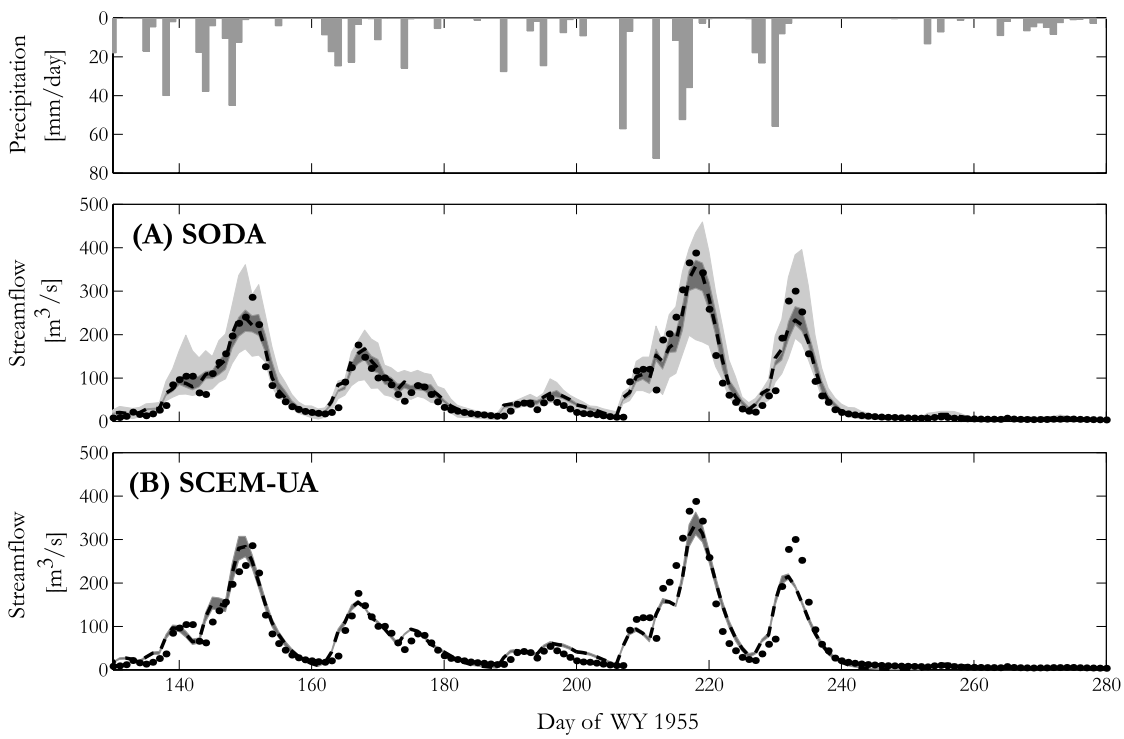

Figure 8. Comparison of the results of SODA and SCEM-UA by application to hydrologic modeling using the HYMOD conceptual watershed model and historical streamflow data from the Leaf River watershed in Mississippi. (a) One-step-ahead hydrograph prediction uncertainty ranges corresponding to the posterior parameter probability density function (pdf) identified with SODA and an ensemble size of 100 members for state estimation. The darkly shaded region denotes the parameter uncertainty, whereas the total uncertainty (input, output, model structural and parameter) is indicated with the lightly shaded region. The dashed line denotes the prediction of the SODA identified "best" parameter set. (b) Prediction uncertainty ranges of 95% of the HYMOD model forecasts associated with the SCEM-UA-derived posterior parameter distribution. Observed streamflows are indicated with solid circles. One-step-ahead streamflow forecasts were generated assuming complete knowledge of the rainfall time series.

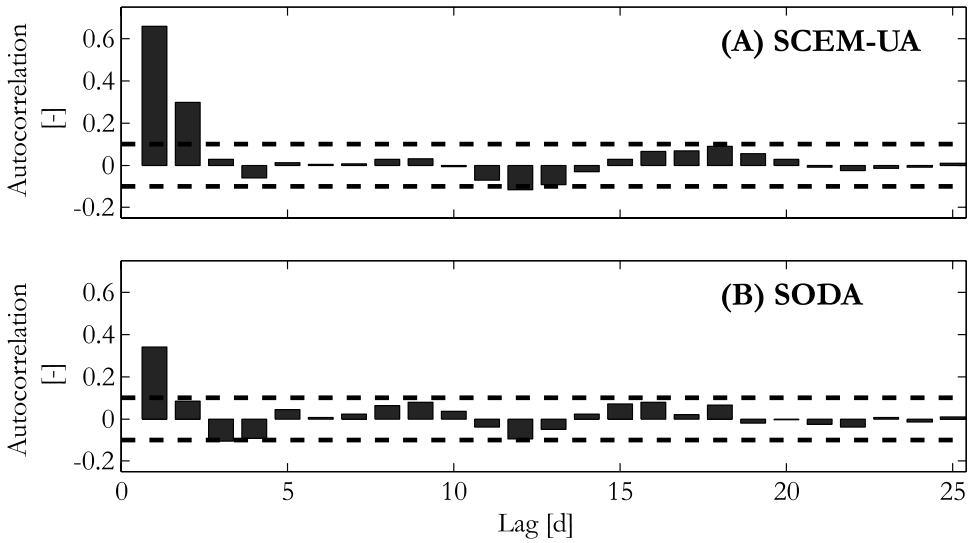

Figure 9. Autocorrelation functions of the time series of forecast errors using the (a) SCEM-UA algorithm and the (b) SODA framework. The dashed lines denote the theoretical upper and lower 99% significance intervals of a time series of white residuals. See color version of this figure in the HTML.

done in GLUE), that in fact leads to considerably larger hydrograph uncertainty bounds than those presented for SODA, particularly during the storm events. Moreover, as argued in the first two sections of this paper, it is not realistic to attribute all uncertainty in the modeling exercise to uncertainty in the parameter estimates.

[58] Another diagnostic measure, which can be used to contrast the results of SODA with those obtained using classical Bayesian model calibration, is to compare the autocorrelation functions of the SCEM-UA (see Figure 1c) and SODA-derived time series of forecast errors. Autocorrelation is a measure of the accuracy of the model predictions. Figure 9 presents the results of this analysis.

[59] To benchmark against a time series of white forecast errors, the dashed lines in these plots represent the theoretical 99% confidence intervals of a white error time series [*Box and Jenkins*, 1976]. When performing a traditional SCEM-UA calibration, there is significant autocorrelation between the residuals at the first few lags, confirming our earlier findings reported in Figure 1. Notice, however, that there is considerably less autocorrelation between the forecast residuals when using the SODA framework, suggesting that recursive state adjustments remove a large part of the bias in the model predictions.

[60] This is further demonstrated in Table 2, which presents summary statistics of the 1-day-ahead streamflow

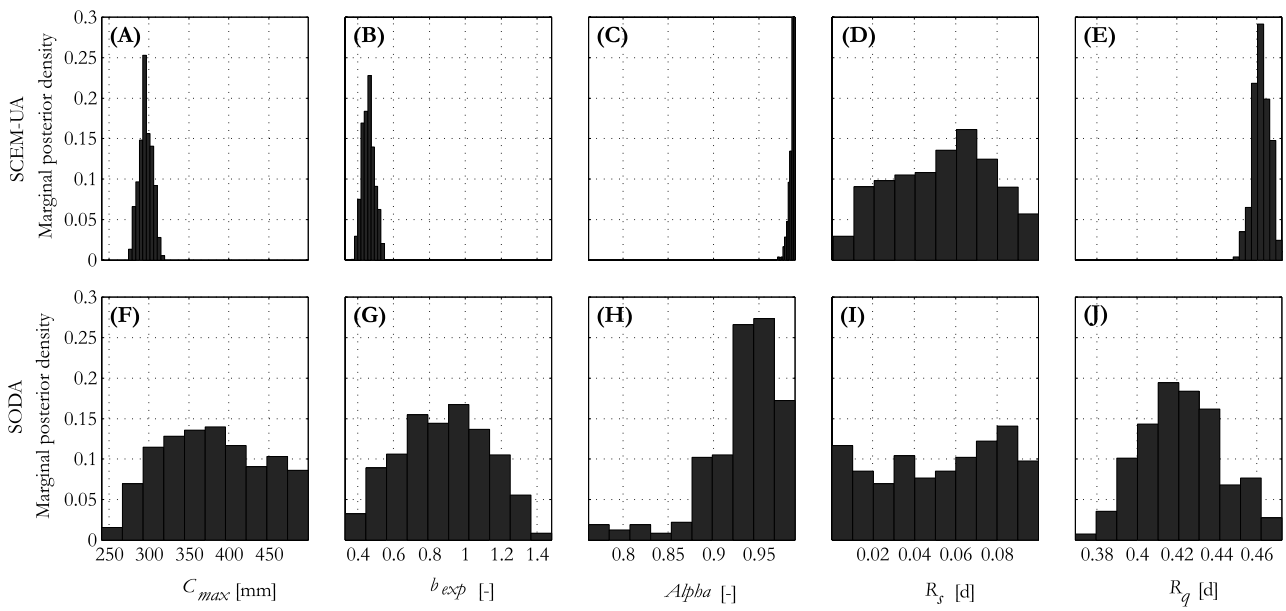

Figure 10. Marginal posterior probability distributions of the HYMOD model parameters derived for the Leaf River watershed using the (a–e) SCEM-UA algorithm and the (f–j) SODA framework. See color version of this figure in the HTML.

prediction performance of the SODA and SCEM-UA methods for the 3 year calibration (28 July 1952–1955) and 5 year evaluation (29 July 1955–30 September 1960) period.

[61] The results in this table indicate that SODA consistently provides better values of the Root Mean Square Error (RMSE), correlation (CORR), and bias (BIAS) statistics than conventional batch calibration for both time periods. Although not further demonstrated here, the performance of SODA with a simple five-parameter conceptual model compares favorably well with other modeling approaches, including the 13-parameter Sacramento Soil Moisture Accounting model of *Burnash et al.* [1973], and the multiparameter Self-Organizing Linear Output (SOLO) map, recently developed by *Hsu et al.* [2002]. Thus explicitly accounting for uncertainty in the input, output and model structure, during model calibration will lead to both an improved assessment of predictive uncertainty, and a simultaneous improvement in the model forecasts.

5.4. Use of SODA for Hydrologic Model Calibration: Parameter Uncertainty

[62] Figure 10 presents posterior marginal probability density distributions for each of the HYMOD model parameters estimated for the Leaf River watershed using the SCEM-UA (Figures 10a–10e) and SODA (Figures 10f–10j) algorithms.

[63] The results presented in this figure highlight two interesting observations. First, notice that for some of the model parameters the SCEM-UA and SODA algorithms have assigned the highest probability to different locations in the parameter space. If the model structure would be correct and the input and output data were observed without error, no state adjustments would be needed during model calibration and the SCEM-UA and SODA method would identify the same location for the mode of the posterior probability distribution in the full parameter space. Not properly accounting for model structural and input errors during model calibration leads to corrupted parameter estimates, which are compensating for these errors.

[64] On the basis of these propositions, we therefore suggest that the model parameters identified with SODA are less corrupted by model structural and input errors and, therefore, better represent the underlying properties of the physical system. Our view has also been supported by additional numerical experiments with synthetically generated streamflow data, and a corrupted input (rainfall) data during the inverse identification (see, e.g., *Kavetski et al.* [2003] for a similar approach), which demonstrated that model parameters identified with SODA seem to be less sensitive to errors in the input and converge in close proximity of the true parameter values used to generate the synthetic streamflow data. To increase the prospects of finding useful regionalization relationships, it would therefore seem most productive to use the SODA identified mode of the parameter space.

[65] A second significant and interesting observation is that the HYMOD model parameters become less well defined when allowing for recursive state adjustments during model calibration (compare histograms of identical counterparts in Figure 10). For most of the parameters, the dispersion around the mode of the distribution has significantly increased. In the systems theoretic sense, this can be explained as part of the ''information'' from the streamflow measurements being used in SODA to recursively estimate the state variables.

5.5. Use of SODA for Hydrologic Model Calibration: Toward Model Structural Improvements

[66] An interesting byproduct of SODA, which deserves further investigation, is the computed time series of state and output innovations. Any inability of the model to represent the input-output behavior of the underlying hydrologic system, will result in state updates in the model, when new measurements of the system are assimilated and processed. We are therefore left with the intuitively reasonable hypothesis that the computed time series of state updates should contain valuable information about model structural errors. In the last section of this paper, we verify the validity of this hypothesis by closer examination of the time series of output innovations.

[67] To facilitate this process, we begin by partitioning the hydrograph into a driven, nondriven quick, and a nondriven slow component, in a method similar to *Boyle et al.* [2000]. Another strategy to exploit the information contained in the innovations for the case of transfer function models, is presented in *Young* [2001, 2002a, 2002b]. For each of these portions of the hydrograph, we plot the SODA-computed mean ensemble output innovation against the mean ensemble streamflow prediction (prior to the update) for the calibration period. The results of this analysis are presented in Figure 11.

[68] For the driven and nondriven quick part of the hydrograph, these plots do not reveal any simple relationship between the ensemble mean predicted flow level (x axis) and associated output innovations (y axis). Nonlinear time series analysis (not presented here) using artificial neural networks (ANN) indicates that some structure is, indeed, present in the time series of output innovations, but the process of how to translate these findings into simple mathematical equations which can be used to improve the conceptual model and as such our understanding of hydrologic processes, remains the topic of future work. A more easily interpretable result is obtained for the nondriven slow part of the hydrograph (Figure 11c). Here, a clear systematic linear relationship between flow level and associated output innovation is apparent. This relationship can easily be exploited to improve the performance of the HYMOD model during low flows. Indeed, when incorporating a linear relationship between flow level and output innovation into HYMOD, the predictive capabilities of the model during nondriven slow flow increases from an average error of 0.20 to 0.09 m^3/s. In contrast, the classical model calibration methodologies, which assign all the uncertainty in the input-output representation to uncertainty in the model parameters, provide little or no such guidance.

6. Summary

[69] The objective of model identification is to obtain a model where the input-state-output behavior is consistent with the measurements, and where the model predictions are accurate and precise. In practice, however, because of errors in the model structure, input (forcing), and output data, this has proven to be quite difficult, leading to considerable uncertainty in the model predictions. Classical model calibration strategies typically ignore input uncertainty, lump

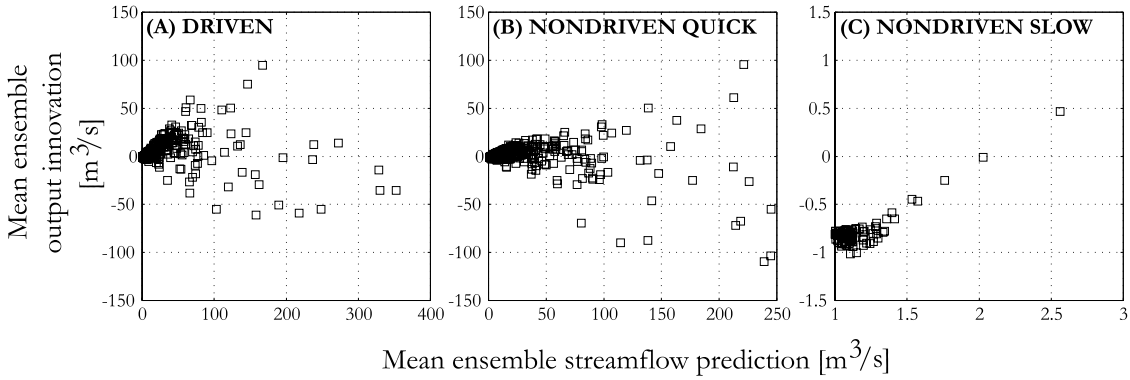

Figure 11. SODA computed mean ensemble output innovations (y axis) versus mean ensemble streamflow prediction for the (a) driven, (b) nondriven quick, and (c) nondriven slow part of the hydrograph.

model structural and output errors into a single white noise term, and assign the uncertainty in the input-output representation of the model primarily to the parameters, therefore lacking the conceptual rigor required to properly account for all the important sources of uncertainty.

[70] In this paper we present a combined global optimization and data assimilation method, which improves the treatment of input, output, parameter, and model structural uncertainty in hydrologic model calibration. The method, entitled simultaneous optimization and data assimilation (SODA), merges the strengths of the parameter search efficiency and explorative capabilities of the Shuffled Complex Evolution Metropolis algorithm and the power and computational efficiency of the ensemble Kalman filter to simultaneously estimate parameters and state variables in hydrologic models.

[71] The usefulness and applicability of SODA has been demonstrated for two preliminary case studies. The first case study considered the highly nonlinear three-parameter Lorenz model, and demonstrated that SODA is indeed successfully able to simultaneously estimate state variables and model parameters when confronted with highly nonlinear model dynamics. The second case study explored the usefulness of SODA by application to hydrologic modeling using the HYMOD conceptual watershed model and historical streamflow data from the Leaf River watershed in Mississippi. The ability of SODA to deal with input, output, parameter and model structural errors, results in improved estimates of the parameter and model prediction uncertainty ranges. With this basis, a detailed investigation of the state and output innovation time series, can be used to investigate improvements to our model concepts and as such our understanding of the functioning of hydrologic systems.

[72] Research aimed at further improvements of SODA is ongoing. This includes (1) the derivation and implementation of realistic input error models to further distinguish between input and model structural errors, (2) the extension to multicriteria problems; SODA does not exclude the use of different parameter sets to match different portions of the hydrograph, (3) the use of recursive parameter estimation methods to further minimize the short-term bias using the state augmentation technique, and (4) the development of improved filtering methods which further relax the Gaussian assumptions of the error distributions in classical KF implementations. Potential improvements related to these issues will be reported in future papers. Also, one of the reviewers of this paper suggested that the use of probabilistic approaches that define uncertainty boundaries, such as SODA, might motivate and support the use of structurally simpler hydrologic models. This is an interesting conjecture that is beyond the scope of the current work and is definitely worthy of further investigation. Software used in this and related work can be found at http://www.science.uva.nl/ibed/cbpg/products.

[73] **Acknowledgments.** The Earth Life Sciences and Research Council (ALW) partly supported the investigations of the authors with financial aid from the Netherlands Organization for Scientific Research (NWO). The third author received support from the Center for Sustainability of Semi-arid Hydrology and Riparian Areas (SAHRA) under the STC Program of the National Science Foundation, agreement EAR-9876800. The constructive review comments of the Associate Editor, Peter Young, and Jean-Philippe Drecourt significantly improved the current version of this paper.

References

Beck, M. B. (1985), Late eutrophication: Identification of tributary nutrient loading and sediment resuspension dynamics, *Appl. Math. Comput.*, *17*, 433–458.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*, 1393–1442.

Beck, M. B., and P. C. Young (1975), A dynamic model for DO-BOD relationships in a non-tidal stream, *Water Res.*, *9*, 769–776.

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

Box, G. E. P., and G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, Boca Raton, Fla.

Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analyses*, Addison-Wesley, Boston, Mass.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*, 3663–3674.

Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724.

Burnash, R. J. E., R. L. Ferral, and R. A. McQuire (1973), A generalized streamflow simulation system, report, Joint Fed. State River Forecasting Cent., Sacramento, Calif.

Dette, H., A. Munk, and T. Wagner (1998), Estimating the variance in nonparametric regression—What is a reasonable choice?, *J. R. Stat. Soc. B., Part 4*, *60*, 751–764.

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

Evensen, G. (1992), Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model, *J. Geophys. Res.*, *97*, 17,905–17,924.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162.

Evensen, G. (1997), Advanced data assimilation for strongly nonlinear dynamics, *Mon. Weather Rev.*, *125*, 1342–1354.

Evensen, G., and P. J. van Leeuwen (1996), Assimilation of Geostat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model, *Mon. Weather Rev.*, *124*, 85–96.

Evensen, G., and P. J. van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, *128*, 1852–1867.

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

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, *7*, 457–472.

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

Gupta, H., M. Thiemann, M. Trosset, and S. Sorooshian (2003), Reply to comment by K. Beven and P. Young on "Bayesian recursive parameter estimation for hydrologic models," *Water Resour. Res.*, *39*(5), 1117, doi:10.1029/2002WR001405.

Gupta, V. K., and S. Sorooshian (1994), Calibration of conceptual hydrologic models: Past, present and future, in *Trends in Hydrology*: *Research Trends*, pp. 329–346, Counc. of Sci. Res. Integration, Trivandrum, India.

Hall, P., J. W. Kay, and D. M. Titterington (1990), Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika*, *77*, 521–528.

Hastings, W. K. (1970), Monte-Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109.

Hebson, C., and E. F. Wood (1985), Partitioned state and parameter estimation for real-time flood forecasting, *Appl. Math. Comput.*, *17*, 357–374.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*, 796–811.

Hsu, K.-L., H. V. Gupta, X. Gao, S. Sorooshian, and B. Imam (2002), Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, *Water Resour. Res.*, *38*(12), 1302, doi:10.1029/2001WR000795.

Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting input uncertainty in environmental modeling, in *Calibration of Watershed Models*, *Water. Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 49–68, AGU, Washington, D. C.

Keesman, K. J. (1990), Set theoretic parameter estimation using random scanning and principal component analysis, *Math. Comput. Simul.*, *32*, 535–543.

Keppenne, C. L. (2000), Data assimilation into a primitive equation model with a parallel ensemble Kalman filter, *Mon. Weather Rev.*, *128*, 1971–1981.

Kitanidis, P. K., and R. L. Bras (1980a), Adaptive filtering through detection of isolated transient errors in rainfall-runoff models, *Water Resour. Res.*, *16*, 740–748.

Kitanidis, P. K., and R. L. Bras (1980b), Real-time forecasting with a conceptual hydrological model: 1. Analysis of uncertainty, *Water Resour. Res.*, *16*, 1025–1033.

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

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*, 69–85.

Lermusiaux, P. F. J. (1999), Data assimilation via error subspace statistical estimation, part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation, *Mon. Weather Rev.*, *127*, 1408–1432.

Lorenz, E. N. (1963), Deterministic non-periodic flow, *J. Atmos. Sci.*, *20*, 130–141.

Madsen, H., and R. Cañizares (1999), Comparison of extended and ensemble Kalman filter for data assimilation in coastal area modeling, *Int. J. Numer. Methods Fluids*, *31*, 961–981.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1091.

Miller, R. N., M. Ghil, and F. Ghautiez (1994), Advanced data assimilation in strongly nonlinear dynamical system, *J. Atmos. Sci.*, *51*, 1037–1055.

Misirli, F. (2003), Improving efficiency and effectiveness of Bayesian recursive parameter estimation for hydrologic models, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson.

Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, *30*, 273–297.

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

Rice, J. (1984), Bandwidth choice for nonparametric kernel regression, *Ann. Stat.*, *12*, 1215–1230.

Seifert, B., T. Gasser, and A. Wolf (1993), Nonparametric estimation of residual variance revisited, *Biometrika*, *80*, 373–383.

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

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrological models, *Water Resour. Res.*, *37*, 2521–2535.

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

Vrugt, J. A., A. H. Weerts, and W. Bouten (2001), Information content of data for identifying soil hydraulic parameters from outflow experiments, *Soil Sci. Soc. Am. J.*, *65*, 19–27.

Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian (2002), Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.*, *38*(12), 1312, doi:10.1029/2001WR001118.

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

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

Vrugt, J. A., S. C. Dekker, and W. Bouten (2003c), Identification of rainfall interception model parameters from measurements of throughfall and forest canopy storage, *Water Resour. Res.*, *39*(9), 1251, doi:10.1029/2003WR002013.

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

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

Young, P. C. (1986), Time-series methods and recursive estimation in hydrological systems analysis, in *River Flow Modeling and Forecasting*, edited by D. A. Kraijenhoff and J. R. Moll, pp. 129–180, Springer, New York.

Young, P. C. (2001), Data-based mechanistic modeling and validation of rainfall-flow processes, in *Model Validation*, edited by M. G. Anderson and P. G. Bates, pp. 117–161, John Wiley, Hoboken, N. J.

Young, P. C. (2002a), Advances in real-time flood forecasting, *Philos. Trans. R. Soc. Phys. Eng. Sci.*, *360*, 1433–1450.

Young, P. C. (2002b), Advances in real-time flow forecasting, *CRES Rep. TR/176*, Cent. for Res. on Environ. Syst. and Stat. Univ. of Lancaster, U. K.

Young, P. C., and M. B. Beck (1974), The modeling and control of water quality in a river system, *Automatica*, *10*, 455–468.

W. Bouten, J. M. Verstraten, and J. A. Vrugt, Department of Physical Geography, Faculty of Science, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, Netherlands. (wbouten@science.uva.nl; jmverstraten@science.uva.nl; j.vrugt@science.uva.nl)

C. G. H. Diks, Department of Quantitative Finance, Faculty of Economics and Econometrics, CenDEF, University of Amsterdam, Roetersstraat 11, 1018 WB, Amsterdam, Netherlands. (diks@fee.uva.nl)

H. V. Gupta, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA. (hoshin@hwr.arizona.edu)