

Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework

Yuqiong Liu¹ and Hoshin V. Gupta¹

Received 20 November 2006; revised 20 April 2007; accepted 30 April 2007; published 3 July 2007.

[1] Despite significant recent developments in computational power and distributed hydrologic modeling, the issue of how to adequately address the uncertainty associated with hydrological predictions remains a critical and challenging one. This issue needs to be properly addressed for hydrological modeling to realize its maximum practical potential in environmental decision-making processes. Arguably, the key to properly addressing hydrologic uncertainty is to understand, quantify, and reduce uncertainty involved in hydrologic modeling in a cohesive, systematic manner. Although general principles and techniques on addressing hydrologic uncertainty are emerging in the literature, there exist no well-accepted guidelines about how to actually implement these principles and techniques in various hydrologic settings in an integrated manner. This paper reviews, in relevant detail, the common data assimilation methods that have been used in hydrologic modeling to address problems of state estimation, parameter estimation, and system identification. In particular, the paper discusses concepts, methods, and issues involved in hydrologic data assimilation from a systems perspective. An integrated hierarchical framework is proposed for pursuing hydrologic data assimilation in several progressive steps to maximally reduce uncertainty in hydrologic predictions.

Citation: Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, 43, W07401, doi:10.1029/2006WR005756.

1. Introduction

[2] Hydrologic modeling has benefited from significant developments over the past two decades, including dramatic growths in computational power, ever increasing availability of distributed hydrologic observations, and improved understanding of the physics and dynamics of the hydrologic system. This has led to the building of higher levels of complexity into hydrologic models, and an advance from lumped, conceptual models toward semidistributed and distributed physics-based models. Paradoxically, while these advances reflect our growing understanding, they have also increased the need for concrete methods to deal with the increasing uncertainty associated with the models themselves, and with the observations required for driving and evaluating the models. It is now being broadly recognized that proper consideration of uncertainty in hydrologic predictions is essential for purposes of both research and operational modeling [Wagner and Gupta, 2005]. The value of a hydrologic prediction to water resources and other relevant decision-making processes is limited if reasonable estimates of the corresponding predictive uncertainty are not provided [e.g., Georgakakos et al., 2004].

[3] To adequately address uncertainty in hydrologic modeling, there are three distinct yet related aspects to be considered: understanding, quantification, and reduction of uncertainty. Arguably, understanding uncertainty is an inte-

gral part of any application of uncertainty quantification and/or reduction. Many uncertainty analysis frameworks have been introduced in the hydrologic literature, including the generalized likelihood uncertainty estimation (GLUE) methodology [Beven and Binley, 1992], the Bayesian recursive estimation technique (BaRE) [Thiemann et al., 2001], the Shuffled Complex Evolution Metropolis algorithm (SCEM) [Vrugt et al., 2003a], the multiobjective extension of SCEM [Vrugt et al., 2003b], the dynamic identifiability analysis framework (DYNIA) [Wagner et al., 2003], the maximum likelihood Bayesian averaging method (MLBMA) [Neuman, 2003], the dual state-parameter estimation methods [Moradkhani et al., 2005a, 2005b], and the simultaneous optimization and data assimilation algorithm (SODA) [Vrugt et al., 2005]. However, few of these methods completely address all the above three critical aspects of uncertainty analysis in an explicit and cohesive way.

[4] Methods of probabilistic prediction and data assimilation (DA) for quantification and reduction of state uncertainty have been extensively explored in the atmospheric and oceanic sciences [e.g., Daley, 1991; Courtier et al., 1993; Anderson and Anderson, 1999]. Their application in the hydrological sciences is relatively new, although deterministic hydrological prediction and parameter estimation have become reasonably mature. Nevertheless, the hydrologic literature has seen various applications of data assimilation and/or uncertainty analysis in hydrology ranging from characterization of soil moisture and/or surface energy balance [e.g., Entekhabi et al., 1994; Houser et al., 1998; Entekhabi et al., 1999; Galantowicz et al., 1999; Boni et al., 2001; Walker et al., 2001; Reichle et al., 2001a, 2001b, 2002a, 2002b; Margulis et al., 2002; Dunne and Entekhabi,

¹Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

2005], to rainfall-runoff modeling [e.g., *Restrepo*, 1985; *Moradkhani et al.*, 2005a, 2005b; *Vrugt et al.*, 2005], to flood forecasting [e.g., *Kitanidis and Bras*, 1980; *Young*, 2002], to estimation of hydraulic conductivity [e.g., *Katul et al.*, 1993; *Lee et al.*, 1993], to groundwater flow and transport problems [e.g., *Eigbe et al.*, 1998; *Graham and McLaughlin*, 1991; *McLaughlin et al.*, 1993], to estimation of water table elevations [e.g., *Van Geer et al.*, 1991; *Yangxiao et al.*, 1991], and to water quality modeling [e.g., *Beck*, 1987].

[5] One critical issue for hydrologic modeling is how the DA methods used in atmospheric and related sciences can best be adapted and combined with hydrologic methods to cope with the uncertainties arising from hydrologic modeling in a cohesive, systematic way to maximally reduce and adequately quantify the predictive hydrologic uncertainty [Krzysztofowicz, 1999; Mantovan and Todini, 2006]. Although general principles and techniques on addressing hydrologic uncertainty are emerging in the literature, there exist no well-accepted guidelines about how to actually implement these principles and techniques in various hydrologic settings. In this paper we discuss the sources of uncertainty in hydrological modeling from a systems perspective, illustrate in detail some of the common DA methods that have been used to quantify and reduce hydrological uncertainty, and propose a (preliminary) hierarchical data assimilation framework for systematically addressing the various types of uncertainties as a way to move forward. It is worth noting that this paper does not attempt to provide a comprehensive review of the literature regarding all the methods, applications, and issues related to data assimilation in hydrology; instead, we aim to present to the readers an illustrative and integrated (rather than fragmented) picture of the state of the art of hydrological data assimilation from a systems perspective.

[6] The paper is organized as follows: Section 2 discusses the three important aspects in addressing hydrologic uncertainty, i.e., understanding, quantifying, and reducing uncertainty; in section 3 we present an integrated view of uncertainty in hydrologic modeling from a systems perspective; Bayes' theorem and its application to data assimilation are discussed in section 4; sections 5, 6, 7, and 8 are devoted to reviews of the common methods that have been used to approach problems of system identification, parameter estimation, state estimation, and simultaneous state and parameter estimation, respectively; an integrated Bayesian hierarchical framework for handling all hydrologic uncertainty in a cohesive, systematic manner is proposed in section 9; and the paper closes with some general discussions and recommendations for future research in section 10.

2. Understanding, Quantifying, and Reducing Hydrologic Uncertainty

[7] As mentioned in the introduction, understanding, quantification, and reduction of uncertainty are the three critical aspects to be considered in order to adequately address uncertainty in hydrologic modeling and prediction. For a full uncertainty analyses one may argue that there exists an additional aspect where uncertainty in the predictions are analyzed and interpreted to infer the deficiencies in the model and data, a process that *Wagener and Gupta* [2005] referred to as "uncertainty communication." This,

however, is beyond the scope of the current paper, which focuses on hydrologic data assimilation.

[8] Obviously, without first adequately understanding all the different uncertainty sources and the relationships between them, it is difficult to conduct uncertainty quantification and reduction in a meaningful way. This is because different uncertainty sources may introduce significantly different error characteristics that require different techniques to deal with; and missing important uncertainty sources may lead to misleading uncertainty predictions in the hydrologic outputs. As of today, our understanding of hydrologic uncertainty is still far from complete and there is much room for further efforts in search of cohesive, systematic means to approach this. It is also very important to distinguish modeling uncertainty from predictive uncertainty: While modeling uncertainty comes mainly from the imperfect fit to the truth of the past, predictive uncertainty can also arise from extrapolation errors or temporal prediction errors due to the fact that the future typically does not look exactly like the past [e.g., *Morgan et al.*, 1990; *Krupnick et al.*, 2006]. In other words, predictive uncertainty is related to, but not necessarily equivalent to, modeling uncertainty; and reduction in modeling uncertainty does not necessarily lead to enhanced predictability of the model under changing conditions. In decision-making processes, there may exist other types of uncertainty, such as decision uncertainty, which arises "whenever there is ambiguity and controversy about how to quantify or compare social objective" [Finkel, 1990, p. 16], and scenario uncertainty, which is related to the inability of the scenarios to account for all the factors affecting the key output/decision variables [Cullen and Frey, 1999]. In the context of hydrological data assimilation, addressing modeling uncertainty is of primary interest, which, in turn, will have an impact on predictive uncertainty.

[9] As far as quantifying uncertainty is concerned, a classical and straightforward way presented in the literature is to represent the predictions in terms of a probability distribution, computed by performing probabilistic instead of deterministic prediction/modeling [e.g., *Kuczera and Parent*, 1998; *Krzysztofowicz*, 1999; *Montanari and Brath*, 2004; *Tamea et al.*, 2005]. For example, by producing an ensemble of hydrologic predictions (instead of a single deterministic prediction as does traditional hydrologic modeling), probabilistic prediction seeks to take into account uncertainties in the equations and/or parameters that are used to describe the physical system and in the hydrologic observations that are made on the system and used in the prediction/modeling process. Of course, for effective quantification of the uncertainty, some prior knowledge (estimate) about the error characteristics that describe the probability distribution of the uncertainties is required, indicating that quantification of uncertainty is, indeed, to a large degree dependent on the understanding of uncertainty. In practical applications of probabilistic prediction, the high nonlinearity of the hydrologic system and the complex interactions between different components of the system result in it being highly difficult to estimate and apply probability distributions that accurately represent the true joint distributions of the uncertainties without creating computational and/or mathematical difficulties. Hence, in practice, locally linear assumptions are usually made about

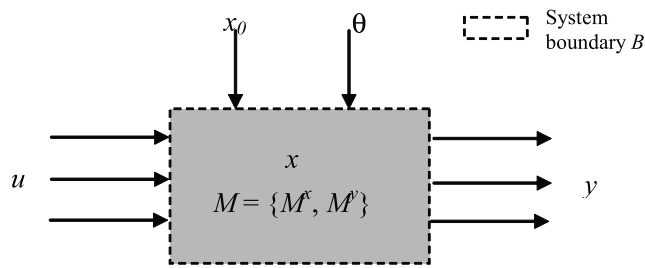


Figure 1. Schematic diagram of model components from a systems perspective.

the system, and uniform or (truncated) Gaussian/normal distributions are typically used to quantitatively represent various sources of uncertainties [e.g., Moradkhani *et al.*, 2005a, 2005b]. To quantify the uncertainty in hydrologic outputs, sampling (sometimes called ensemble) methods are now widely used by taking samples from the assumed error probability density functions (PDFs) and running the model forward for a certain amount of time. With a sufficiently large sample of predictions, statistics describing the uncertainties in model outputs can be easily derived from the sample. In most cases, quantification of uncertainty is embedded in the data assimilation processes aiming to reduce predictive uncertainty as discussed below.

[10] There are three main areas where actions can be taken toward reducing uncertainty in hydrologic predictions: (1) acquisition of more informative and higher quality hydrological data (including data of new types) by developing improved measurement techniques and observation networks; (2) development of improved hydrologic models by incorporating better representations of physical processes and using better mathematical techniques; and (3) development of efficient and effective techniques that can better extract and assimilate information from the available data via the model identification and prediction processes.

[11] While hydrologic science has witnessed astonishing advances in the availability of hydrologic data (area 1) and the complexity/reliability of hydrological models (area 2), there is an urgent need for techniques that effectively and efficiently assimilate important information from the data into the models to produce improved hydrological predictions (area 3). We will generally refer to such techniques as data assimilation (DA) methods, defined here as

procedures that aim to produce physically consistent representations or estimates of the dynamical behavior of a system by merging the information present in imperfect models and uncertain data in an optimal way to achieve uncertainty quantification and reduction.

[12] It is worth mentioning that this description of the DA problem is broadly encompassing, not being limited only to problems of “state estimation” as the term is often applied to in the literature. Instead, it describes the more comprehensive problem of “merging models with data” and therefore includes the three related problems of system (structure) identification, parameter estimation, and state estimation, which are all critical to the reduction of uncertainty in model predictions. More details on these concepts are provided in section 3.

[13] Arguably, understanding uncertainty should always be an integral part of any application of uncertainty

quantification/reduction; and given the continual arrival of different kinds of observations, one should not stop at the quantification step but continue to reduce the uncertainty by assimilating new observations. In most cases of DA applications, the process of uncertainty reduction inherently involves the quantification of uncertainty in the model inputs, parameters, structure, and observations, and preferably provides quantitative information about uncertainty in model predictions or forecasts. In recognition of this, the focus of this paper is given to understanding and reduction of hydrologic uncertainty from a systems perspective (sections 3–9).

3. Hydrologic Uncertainty From a Systems Perspective

[14] Uncertainty in hydrologic modeling may arise from several sources: model structure, parameters, initial conditions, and observational data used to drive and evaluate the model. In this section, to formally specify the different sources, we will describe a model as being composed of multiple components from the perspective of systems theory. Errors in each of these model components can give rise to uncertainty in hydrologic modeling. In this sense we include within the realm of data assimilation any procedure that assimilates information from observations to reduce the uncertainty associated with one or more of the model components, be it the state, the parameters, or the system structure.

3.1. Model Components in Systems Theory

[15] For the purpose of communication, here we consider a model to be composed of seven different components (Figure 1): system boundary (B), inputs (u), initial states (x_0), parameters (θ), structure (M), states (x), and outputs (y). Note not all the hydrologic applications existing in the literature comply with this definition/terminology of system components (see below).

[16] In this exposition we define the inputs u and outputs y as fluxes of mass and/or energy into and out of the system across the system boundary B ; states x as time-varying quantities of mass and/or energy stored within the system boundary B ; and parameters θ as characteristic properties of the system that are assumed to be “time-invariant” (remain constant over the time duration of interest). Note that in some fields, the system “state” x^* is taken to be some other quantity somehow related to the mass or energy state x ; in such cases the same general equations hold but with some modifications to account for the relationship of x^* to x . Also, we shall return to the issue of time-invariance of model parameters in a moment. For example, in catchment modeling, u may refer to the time-varying two-dimensional spatial distribution of precipitation flux over the catchment; y may refer to the time-varying two-dimensional distribution of streamflow flux at all points along the river network and of evaporation and transpiration from the surface of the catchment; x may refer to the three-dimensional time-varying spatial distribution of surface and subsurface moisture stored within the catchment boundary; and θ may refer to the time-invariant three-dimensional spatial distribution of catchment characteristics such as the soil hydraulic properties.

[17] The model structure M consists of two components: M^x and M^y (i.e., $M = \{M^x, M^y\}$). Here M^x and M^y are (in general) nonlinear vector functional relationships, where M^x represents the input-to-state mapping and M^y represents the state-to-output mapping. For example, M^x may refer to the coupled equations describing the three-dimensional evolution of surface and subsurface moisture in response to catchment inputs and outputs (precipitation, evaporation, transpiration, and outflow), and M^y may refer to the coupled equations describing the dependence of catchment outputs (evaporation, transpiration, and outflow) on the system states. These mappings can be described or constructed in a variety of different ways, including the continuous-time differential equation formulation (using t to represent continuously varying time):

$$dx/dt = M^x(x, u|\theta) \quad (1)$$

$$y = M^y(x|\theta), \quad (2)$$

and the discrete-time difference equation formulation (using k to represent discrete moments in continuously varying time t):

$$x_{k+1} = M_{k+1}^x(x_k, u_{k+1}|\theta) \quad (3)$$

$$y_{k+1} = M_{k+1}^y(x_{k+1}|\theta). \quad (4)$$

[18] Since computer-based implementations are usually constructed to make predictions at discrete moments of time, we shall (without loss of generality) use the discrete time formulation described by equations (3) and (4) in all subsequent discussion. Note that the formulation must implicitly employ the continuity equation $dx/dt = u - y$ to ensure physical consistency in the time-dependent accounting for mass and energy fluxes.

[19] As mentioned above, this formulation assumes that the model parameters θ do not vary with time over the duration of interest. As a conceptual extension, one might wish more generally to permit the system characteristics represented as “parameters” to vary slowly with time, in response to changes in the model state and/or system inputs. In general, we would expect (for reasons of physical consistency) that the rate of this “parameter” variation is slower than that of the variation of the state. To complete the mathematical description, we would then introduce an additional set of mapping relationships that describes, in a manner analogous to the input-state relationship, the time-evolution of the parameters θ (see equations (5)–(7)).

$$\theta_{k+1} = M_{k+1}^\theta(x_k, u_{k+1}, \theta_k|\phi) \quad (5)$$

$$x_{k+1} = M_{k+1}^x(x_k, u_{k+1}|\theta_k) \quad (6)$$

$$y_{k+1} = M_{k+1}^y(x_{k+1}|\theta_k). \quad (7)$$

[20] Note that this revised formulation introduces a new set of (uncertain) time-invariant coefficients ϕ which must be specified a priori or estimated from data; for example, if

θ is believed to take the same Gaussian distribution at all time steps, ϕ might represent the (time-invariant) mean and covariance of that distribution ($\phi \sim N(\mu_\phi, \sigma_\phi^2)$). However, if we define an extended “state” vector $x' = [x, \theta]$ by adjoining the time-varying quantities x and θ into a single variable, and define a new “parameter” vector ϕ , the formulation in (5)–(7) is not fundamentally different from that given in equations (3)–(4). For simplicity of notation, we will therefore proceed by adopting the representation of equations (3)–(4) and let the reader make the appropriate substitutions for the more general case as necessary.

3.2. Errors in Different Model Components

[21] Of the seven model components illustrated in Figure 1, five of them (i.e., B , u , x_0 , θ , and M) must be specified, estimated, or defined before the model can be actually run, while the remaining two (x and y) are computed by running the model. Each of the five predefined components may be uncertain in various characteristic ways, and the consequence of these uncertainties will be mapped into the model states and outputs. Hence input data, parameters, the model structure, initial conditions, and the system boundary represent five major sources of uncertainties in hydrologic modeling. In most cases, model inputs and initial conditions are specified or estimated from in situ observations. Accordingly, errors in these two sources can be collectively considered as data errors or observation errors. Errors in output observations that are used to evaluate the model results should be considered as data errors as well. Note in cases where x_0 are treated as model parameters, errors associated with x_0 can be considered as parameter errors [e.g., Liu et al., 2003]. Definition of the system boundary is part of the model conceptualization process; hence the uncertainty associated with B can be considered as one source of structural uncertainty. In summary, there are three primary types of uncertainties in hydrologic modeling: structural errors, parameter errors, and data errors (see also discussions by Wagener and Gupta [2005]).

[22] 1. Models are assemblies of assumptions and simplifications and thus inevitably imperfect approximations to the complex reality, i.e., the true system that a model seeks to characterize. Conceptualization with inappropriate approximations and omissions can result in large (albeit poorly understood) errors in the conceptual structure of a numerical model. Structure errors can also arise from the mathematical implementation (e.g., spatial and temporal discretizations) that transforms a conceptual model into a numerical model [Neuman, 2003].

[23] 2. Model parameters are conceptual aggregate representations of spatially and temporally heterogeneous properties of the real system. Parameters are an integral part of the equation-based modeling approach, and the use of “effective” parameter values in hydrologic modeling is essential. Errors in the estimates of parameter values can result in huge errors in the model outputs as shown in many modeling studies [e.g., Gupta et al., 1998; Liu et al., 2005]. However, the “conceptual” and spatiotemporal aggregate nature of parameters sometimes makes it difficult to specify them directly and unambiguously from observations made in the field (of course exceptions exist, such as pumping tests to estimate system conductivities). In other words, parameters are not often easily measurable, and must

generally be estimated by indirect means (e.g., prior knowledge or model calibration) with consequent introduction of errors and uncertainties.

[24] 3. Data errors can generate uncertainties in hydrologic predictions through the model inputs and initial conditions, both of which can be estimated from observations [e.g., *Clark and Slater*, 2006]. A data error is also referred to as a measurement error if, as typically is the case, the data of concern is measured. A measurement error usually consists of two components: (1) instrument error due to imperfect measurement devices that do not accurately record the variables they are designed to measure and (2) representativeness error due to scale incompatibility or differences (in time or space) between the variable measured by a device and the corresponding model variable. Representativeness error can be discussed in terms of spacing (distance or interval between samples), extent (overall coverage of measurements in space or time), and support (averaging volume or area of samples) [*Blöschl and Grayson*, 2000]. These two error components tend to have very different characteristics which may vary from variable to variable. To effectively quantify or reduce uncertainty in the predictions, statistics of both errors should be considered and adequately specified.

[25] Structural, parameter, and data errors collectively lead to uncertainties in hydrologic predictions of model outputs and states. Among these three types of errors, structural errors are generally the most poorly understood and the most difficult to cope with; nevertheless, their impacts on hydrologic predictions can be far more detrimental than those of parameter errors and data errors [*Carrera and Neuman*, 1986; *Abramowitz et al.*, 2006].

3.3. Addressing Uncertainty in Different Model Components

[26] Viewing model components and the errors in them within a dynamic systems framework (as described in sections 3.1 and 3.2) helps to better understand and organize the different uncertainty sources in hydrologic modeling. The next critical issue is how to adequately represent (or quantify) the uncertainties in these sources and feed them into a DA framework to effectively and efficiently reduce the predictive uncertainty. A DA application usually requires proper specifications or assumptions of the characteristics of errors associated with the four major error sources (initial conditions, inputs, parameters, and model structure). This, however, is not a trivial issue, because prior knowledge on the error characteristics is usually not available, especially for errors associated with poor specification of the model structure. In the meantime, one should realize that different DA problems may require different techniques/algorithms that best fit into the specific problem setting.

[27] Loosely speaking, there are three types of data assimilation problems based on the model component being considered: state estimation, parameter estimation, and system identification, described as follows.

[28] 1. State estimation seeks to characterize the true “state” of the system by optimally combining state information represented by the model with that inferable from all kinds of available data sources, quantitative or qualitative. In the literature the term data assimilation is commonly used to refer specifically to state estimation only [e.g., *McLaughlin*, 1995, 2002]. Note that the definition (including dimension)

and computation of the model state are conditional on the specification of a model structure and values for the parameters. While current data assimilation methods are praised for their ability to deal with all the three types of errors mentioned above (i.e., structural errors, parameter errors, and measurement errors), most applications of state estimation have been focused on the measurement errors only, without rigorous treatment of structural and parameter errors [e.g., *Reichle et al.*, 2002a, 2002b].

[29] 2. Parameter estimation aims to estimate proper values of the model “parameters” based on available data, so that the model makes sufficiently accurate simulations or predictions of the true input-state-output response. Note that the definition, dimension, and specification of the model parameter set are conditional on the specification of a model structure (the form of the input-state-output relationship). Traditionally, parameter estimation has been conducted by using deterministic (manual or automatic) calibration techniques that tend to ignore model structural errors and measurement errors [e.g., *Duan et al.*, 1992; *Yapo et al.*, 1998]. Recently, stochastic data assimilation methods have been developed and applied to parameter estimation problems [e.g., *Thiemann et al.*, 2001; *Moradkhani et al.*, 2005a, 2005b].

[30] 3. System identification involves the selection of appropriate structures (i.e., conceptual models) for a mathematical or numerical model that aims to represent the real system. More specifically, a system identification process aims to define a set of proper mappings (typically equations, e.g., equations (3) and (4)) that accurately represent the relationships between the model inputs, parameters, states, and outputs [e.g., *Neuman*, 2003].

[31] Among the three types of DA 55.7k0010.0026307(bett522(),-5

consistency) to be related in some manner. We further assume that the probability in the occurrence (or observation) of events A and B can be described by $P(A)$ and $P(B)$. Then, the cooccurrence (or observation) of A and B is represented by the joint probability function $P(A \cap B)$, and this can be further expressed as

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A), \quad (8)$$

where $P(A|B)$ is the conditional probability of occurrence of event A given knowledge that event B has occurred (and similarly for $P(B|A)$). This leads directly to Bayes' theorem:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}. \quad (9)$$

[34] In the Bayesian use of probabilities, the marginal probabilities ($P(A)$ and $P(B)$) and the conditional probabilities ($P(A|B)$ and $P(B|A)$) are referred to as the prior and posterior PDFs, respectively. Bayes' law provides a powerful basis for a full stochastic representation of all the uncertainties in the model and the data in hydrologic modeling. Using Bayes' theory, equation (9) can be reformulated to describe all three aspects of data assimilation, including system identification, parameter estimation, and state estimation.

[35] As discussed in section 2, the five uncertain quantities (B , u , x_0 , θ , and M) must be specified in order to use equations (3) and (4) to compute estimates of the two remaining quantities (x and y). We represent the prior knowledge of the quantities (B , u , x_0 , θ , and M) by the probabilities $p_{\text{prior}}(B)$, $p_{\text{prior}}(u)$, $p_{\text{prior}}(x_0)$, $p_{\text{prior}}(\theta)$, and $p_{\text{prior}}(M)$, respectively. We further assume that there may become available a set of uncertain observations z which may contain information about any of the system aspects of interest. For example, z may consist of direct or indirect measurements on any of the system fluxes (u and y), state variables (x), parameters (θ), or initial conditions (x_0 and B), or more generally can consist of qualitative assessments of any of these quantities, including the model structure (M). Of course, such observations will generally be incomplete, in the sense that they refer to values at a limited and discrete set of points in the four dimensions of space and time. Further, such observations may generally be indirect, in the sense that they actually describe some quantity that is related to the uncertain model quantity of interest. For example, this indirect relationship might arise from inexact correspondence such as scaling differences (e.g., point scale observations are made of spatially distributed soil hydraulic properties, whereas the model representation describes the mean spatial value over some larger scale). Alternatively, it may arise from observing some closely related quantity (e.g., remotely sensed observations are made of radiances which are then related to the system properties of interest via radiative-transfer models). By considering all such factors, we define the following general observation equation:

$$z_k = M^z(y_k, x_k, u_k, x_0, \theta, B, M). \quad (10)$$

[36] To solve the DA problem, we are now interested in the posterior probability distribution (PPD) of the various

quantities of interest: the model structure, parameters, state variables, and outputs. By application of Bayes' theory it can be shown that for given observations z ,

$$p_{\text{posterior}}(M|z) = \frac{p(z|M) \cdot p_{\text{prior}}(M)}{p(z)}. \quad (11)$$

[37] Equation (11) provides a means for identifying appropriate model structures M by describing the posterior probability associated with a selected model structure in terms of the "likelihood" $p(z|M)$ that the observations z might have been generated by the model assuming that the structure M is the correct one, multiplied by the probability that the model structural assumption is correct ($p_{\text{prior}}(M)$). Acknowledging the logical progressive chain of conditional dependence described as $\{M \rightarrow \theta \rightarrow x \rightarrow y\}$, we can further derive

$$p_{\text{posterior}}(\theta|z, M) = \frac{p(z|M, \theta) \cdot p_{\text{prior}}(\theta|M) \cdot p_{\text{prior}}(M)}{p(z, M)} \quad (12)$$

$$\begin{aligned} p_{\text{posterior}}(x|z, M, \theta) \\ = \frac{p(z|x, M, \theta) \cdot p(x|M, \theta) \cdot p_{\text{prior}}(\theta|M) \cdot p_{\text{prior}}(M)}{p(z, M, \theta)} \end{aligned} \quad (13)$$

$$\begin{aligned} p_{\text{posterior}}(y|x, z, M, \theta) \\ = \frac{p(z|y, x, M, \theta) \cdot p(y|x, M, \theta) \cdot p(x|M, \theta) \cdot p_{\text{prior}}(\theta|M) \cdot p_{\text{prior}}(M)}{p(x, z, M, \theta)}. \end{aligned} \quad (14)$$

[38] For simplicity of presentation we have ignored the additional dependence on the system boundary B and initial conditions x_0 . In equation (11), $p(z)$ is a constant that normalizes the posterior probability mass to unity. Equation (12) describes how to compute posterior estimates of the model parameters θ given the model structure M and observations z ; equation (13) describes how to compute posterior estimates of the model states x given the model structure M , parameters θ , and observations z ; and equation (14) describes how to compute posterior estimates of the model outputs y given the model structure M , parameters θ , states x , and observations z . In data assimilation these equations serve as the fundamental basis for system identification (equation (11)), parameter estimation (equation (12)), state estimation (equation (13)), and quantification of uncertainty in hydrologic predictions (equation (14)).

5. Methods for System Identification

[39] In hydrologic modeling or analysis, a system identification problem typically involves selecting or constructing a valid model structure or a set of equally valid model structures (i.e., conceptual models and mathematical implementations) for the hydrologic system of concern. Historically, hydrologic modeling has relied on a single conceptual model of a particular hydrologic environment. *Beven and Freer* [2001, p. 1] point out that for a complex environmental system, there may actually exist "many different

model structures and many different parameter sets within a chosen model structure that may be behavioral or acceptable in reproducing the observed behavior of that system,” a phenomenon that *Beven* [1993] has termed as “equifinality.” This may be partially, if not primarily, due to the limited ability of current conceptual models in representing the complex, heterogeneous hydrologic systems that have unknown, and possibly unique system characteristics [*Beven*, 2000]. In this sense, hydrologic predictions based on a single conceptual model or model structure are invariably subject to statistical bias (if an invalid model is chosen) and underestimation of uncertainty (if equivalent valid models are not included) [*Neuman*, 2003]. Hence the “system identification” problem in hydrologic modeling can be approached through using a suite of “independent” plausible model structures with probability of each structure properly defined so that collectively, these model structures adequately and unambiguously approximate the true underlying system.

[40] In most data assimilation techniques (such as those described later in sections 6 and 7), errors in model structures are usually accounted for by adding an (unbiased) error term to the model transition equation (see section 7.1). However, because of equifinality of models as described above, a full consideration of the model structure error requires involving at least several “independent” alternative model structures that encompass a range of different assumptions [*Beven and Young*, 2003]. In this sense, a multimodel approach based on a suite of conceptual models is better suited for handling uncertainty associated with model structure errors than single-model approaches [*National Research Council*, 2001; *Neuman*, 2003; *Georgakakos et al.*, 2004]. In the hydrologic literature, *Beven and Binley* [1992] introduced the generalized likelihood uncertainty estimation methodology (GLUE) where multiple competing model structures and parameter sets are allowed to account for the possibility of equifinality of models, producing a likelihood-weighted probability distribution of output predictions. GLUE is described in more detail as a model calibration and uncertainty estimation methodology in section 6.2.1.

[41] Along the same line of reasoning, a coherent mechanism for handling structural uncertainty is the concept of Bayesian model averaging (BMA) [*Hoeting et al.*, 1999]. In BMA, the posterior distribution of the prediction on a quantity y given the observation z is approximated by the weighted sum of the posterior distributions of a set of K independent (or mutually exclusive) models $\mathbf{M} = \{M_1, \dots, M_K\}$, i.e.,

$$p(y|z) = \sum_{k=1}^K p(y|M_k, z)p(M_k|z), \quad (15)$$

where the weights are determined by the posterior distributions of the models $p(M_k|z)$ given by Bayes’ theorem as expressed in (11), where the normalization factor $p(z)$ is obtained by

$$p(z) = \sum_{k=1}^K p(z|M_k)p(M_k) \quad (16)$$

and the likelihood of each model M_k (given by $p(z|M_k)$) is calculated as

$$p(z|M_k) = \int p(z|\theta_k, M_k)p(\theta_k|M_k)d\theta_k. \quad (17)$$

[42] In equations (16) and (17), $p(z|M_k)$ is the likelihood of observing the data z given the model M_k ; $p(z|\theta_k, M_k)$ is the joint likelihood of model M_k and its parameter set θ_k ; $p(\theta_k|M_k)$ is the prior density of θ_k given the model structure M_k ; and $p(M_k)$ is the prior probability that the model structure M_k is valid. The BMA framework provides a cohesive way to jointly assess model structure and parameter uncertainties; however, it tends to be computationally demanding/cumbersome and also requires reliable prior information about model parameters. *Neuman* [2002, 2003] proposed a maximum likelihood version of BMA (MLBMA) that proves to be more computationally feasible and capable of dealing with situations where reliable prior information is lacking [*Ye et al.*, 2004, 2005].

[43] In atmospheric science it has recently become very popular to use a multimodel ensemble method (MME) for weather and climate forecasting [e.g., *Doblas-Reyes et al.*, 2000; *Palmer et al.*, 2000; *Ziehmann*, 2000; *Palmer*, 2004; *Hagedorn et al.*, 2005a, 2005b]. In an MME approach a larger ensemble of predictions is composed from a suite of smaller ensembles, each generated based on an independent, plausible model (i.e., several ensembles are generated using each model structure). Instead of computing the probability of each model as in a BMA approach, the goal of MME is to account for uncertainty in the model structure, the assimilated data, and, in particular, the uncertainty associated with knowledge of initial conditions, by means of sampling from the output distributions of several different models. Most MME-based studies have reported that the performance of (properly selected) multimodel ensembles is superior to that of single-model ensembles, due not only to error compensation among different models, but also to the greater consistency and reliability of multimodel ensembles that cover a broad range of possible solutions [e.g., *Georgakakos et al.*, 2004; *Hagedorn et al.*, 2005a, 2005b].

[44] For all the three methodologies mentioned above (i.e., GLUE, BMA, and MME), all probabilities (including the final posterior) are implicitly conditioned on the set of selected models \mathbf{M} . Hence it is critical to select a set of relatively independent, plausible models that are most strongly supported by available data. Otherwise, there is no confidence about whether uncertainty is overestimated or underestimated, and there is no guarantee that the truth will even lie within the range of a model ensemble. This, however, is not straightforward, for there exist no well-accepted guidelines in the literature about how to define “independent” model structures or how many “independent” models are needed to adequately span the model space.

6. Methods for Parameter Estimation

[45] Despite the physical basis of many hydrological models, their parameters are often conceptual, effective quantities that cannot be measured in the field, and must therefore be estimated indirectly. The parameter estimation

problem is referred to by different names in the literature, including model calibration, parameter optimization, data assimilation, inverse problem, parameter tuning, among others. Arguably, an adequate parameter sensitivity analysis should always precede a parameter estimation study to identify sensitive parameters, for including insensitive parameters may render a parameter estimation process ineffective and cumbersome, especially for a complex model that has a large number of parameters [e.g., *Liu et al.*, 2004, 2005]. In this section we review traditional, deterministic model calibration methods as well as the newly emerging, stochastic data assimilation methods for parameter estimation.

6.1. Model Calibration Methods

[46] As an illustration to the general concept of model calibration, we consider a physically based model with p parameters ($\theta = \{\theta_1, \dots, \theta_p\}$), which is to be calibrated by assimilating the information from N different time series of observations $\{Z_n, n = 1, \dots, N\}$ corresponding to N model outputs $\{Y_n, n = 1, \dots, N\}$. The parameter estimation problem can be most generally stated as a vector optimization problem as follows [*Gupta et al.*, 1998]:

$$\text{Minimize } F(\theta) = \{f_n(\theta), n = 1, \dots, N\} \text{ subject to } \theta \in \Theta, \quad (18)$$

where $f_n(\theta)$ is an objective function (also called a criterion) for measuring the distance between the n th model output and the n th observation; Θ is the physically feasible p -dimensional parameter space; and $F(\theta)$ is a vector in the case of a multiobjective parameter estimation problem ($N \geq 2$) and a scalar in single-objective cases ($N = 1$).

[47] The two major strategies used for parameter estimation have been the “manual-expert” approach and the “automatic” approach. While manual-expert strategies rely on the informed but subjective judgment and skill of an experienced hydrologist, automatic strategies utilize the power of computer-based optimization techniques based in nonlinear regression theory. With the emergence of increasingly complex hydrological models with larger numbers of model parameters, effective and efficient automatic approaches have become more popular than the time-consuming, expertise-demanding manual approaches. *Duan et al.* [1992] introduced the Shuffled Complex Evolution algorithm (SCE), a global optimization strategy applicable to a broad class of single-criterion calibration problems. This algorithm was extended to the multiobjective complex optimization method (MOCOM) by *Yapo et al.* [1998], thereby enabling the use of multiple complementary measures for better extraction of information from the data, resulting in improved parameter estimates.

[48] The single- and multiple-criteria methods for parameter estimation mentioned above rely on deterministic nonlinear optimization techniques that seek to identify a single (few) “best” parameter set (sets), thus implicitly ignoring the uncertainties associated with observed data, model structure and parameters. In the case of significant system and data noise or bias, such methods can lead to parameter estimates that provide biased model predictions. Recently, *Vrugt et al.* [2003a] presented an efficient Markov Chain Monte Carlo (MCMC) sampler called the Shuffled Complex Evolution Metropolis algorithm (SCEM, with a multiobjective extension MOSCEM presented by *Vrugt et*

al. [2003b]), which converges to an ensemble of parameter sets that approximates the posterior distribution of model parameters. This posterior description of parameter uncertainty obtained through SCEM or MOSCEM can be used to assess the uncertainty in hydrological outputs arising from parameter uncertainty, representing an improvement over traditional deterministic optimization methods (e.g., SCE and MOCOM) in accounting for uncertainties associated with model parameters.

[49] Nevertheless, application of any of the methods mentioned above is implicitly based on an assumption that there exists a feasible parameter set for which the specific model structure under consideration is able to provide unbiased estimates of the model states and outputs at each time step. When this is not true (as is generally the case), we must acknowledge the existence of model structural and data errors and combine the (stochastic) parameter estimation methods with methods for system identification as described below in section 6.2. In addition, it should be mentioned that most (traditional) parameter estimation methods do not exploit the full power of the Bayesian framework, because they rely on “batch” processing of long-term historical data, and therefore lack the ability to recursively reduce parameter (and hence prediction) uncertainty as new data become available. An exception is the data-based mechanistic (DBM) approach to stochastic modeling, which is based on advanced recursive methods of time series analysis and has been successfully applied to hydrological systems modeling and data assimilation [e.g., *Young*, 2003, and references therein]. When considered in Bayesian terms, the DBM approach has the advantage of quantifying the uncertainty in the model and the data without resort to Monte Carlo methods, resulting in comparatively simple online implementation for flood forecasting and warning [e.g., *Young*, 2002; *Romanowicz et al.*, 2006].

6.2. Parameter Estimation Based on Stochastic Methods

[50] In recognition of the two major limitations of the model calibration methods mentioned above, there has been recent growing interest in the use of stochastic, sequential data assimilation techniques for parameter estimation. Such techniques operate within the Bayesian updating framework for estimation of predictive uncertainty. Examples include the generalized likelihood uncertainty estimation method (GLUE, [*Beven and Binley*, 1992]), the Bayesian recursive estimation method (BaRE [*Thiemann et al.* [2001]]), and other more recent techniques for simultaneous state and parameter estimation (see relevant details in section 8).

6.2.1. Generalized Likelihood Uncertainty Estimation (GLUE)

[51] *Beven and Binley* [1992] introduced the generalized likelihood uncertainty estimation (GLUE) methodology for model calibration that takes into account the effects of uncertainty associated the model structure and parameters. A fundamental assumption underlying GLUE is the “equifinality” or “nonuniqueness” concept [*Beven*, 1993], where multiple model structures and many parameter sets within a chosen structure are considered equally likely as simulators of the system. In other words, it is assumed that there exists no optimal model or parameter set due to structural and parameter uncertainties. This has introduced a different

philosophy to the venue of model calibration where the primary goal had historically been identifying an optimal parameter set based on a single model.

[52] To implement the GLUE methodology, several alternative model structures are selected and appropriate prior parameter uncertainty distributions are assumed for each model. Samples are then taken from these parameter distributions (coupled with their corresponding model structures) to generate Monte Carlo simulations. To evaluate the degree of correspondence between each simulation and the observed system behavior, a likelihood value is calculated based on a predefined likelihood measure (i.e., a measure of goodness of fit). The likelihood values are then used to determine whether a model structure-parameter set is “behavioral” or “nonbehavioral” according to a subjectively defined threshold of likelihood values; and only behavioral model structure-parameter sets are retained to provide predictions of the system behavior. To assess the uncertainty associated with the predictions, weights of the behavioral sets of model structure and parameters are calculated by normalizing the corresponding likelihood values so that all the weights sum up to one; the distribution of these weights is then taken as the probabilistic distribution of the predicted variables to reflect the uncertainty impacts of structural and parameter errors on model predictions.

[53] The primary improvement of the GLUE methodology over the deterministic calibration methods lies in its ability to explicitly account for the combined effects of model structure and parameter uncertainty, by using multiple models and assuming proper prior distributions for each parameter. Moreover, when a new observation period arrives or there exist different observation types (quantitative or qualitative), the likelihood values can be updated to estimate the posterior distribution of parameter sets (and thus that of model predictions), based on Bayes’ theorem. One concern that has been raised is that the Bayesian equation may not properly apply in GLUE in certain cases, because in GLUE, a certain likelihood measure, or essentially an objective function, is used in place of a formal likelihood function that is consistent within the framework of Bayes’ theorem [e.g., *Thiemann et al.*, 2001; *Mantovan and Todini*, 2006]. In addition, in the GLUE procedure, uncertainties associated with input data and output data (i.e., data errors) are not explicitly and/or formally considered.

6.2.2. Bayesian Recursive Estimation (BaRE)

[54] *Thiemann et al.* [2001] introduced the Bayesian recursive parameter estimation (BaRE) methodology that poses the parameter estimation problem within the context of a formal Bayesian framework. Unlike in GLUE where error sources are only implicitly considered with a likelihood measure, BaRE makes strong, explicit assumptions about the characteristics of errors in the observations by using an exponential power density error model. Like in GLUE, proper parameter ranges and prior probability distributions are specified; and the Monte Carlo approach is used to sample from the predefined distributions to represent parameter uncertainty.

[55] Once the error model is defined and model structure-parameter selections are initialized from their prior distributions, the BaRE methodology consists of two recursive steps that are common to the other data assimilation methods for state estimation (see section 7): prediction

and update. At time t_k , BaRE predicts the outputs and the uncertainty in the outputs by running the model forward to the next observation time t_{k+1} (i.e., when the observation z_{t+1} is available) for each set in the model structure-parameter ensemble. To update the probability of the model structure-parameter sets, a recursive version of the Bayesian equation for parameter estimation (equation (12)) is used to obtain the posterior probability of each model structure-parameter set i as follows:

$$p_{\text{posterior}}(M^i, \theta_{k+1}^i | z_{k+1}) \propto p(z_{k+1} | M^i, \theta_k^i) p_{\text{prior}}(\theta_k^i | M^i) p_{\text{prior}}(M^i). \quad (19)$$

[56] After updating, the model system continues to run forward to the next observation time, using the posterior model structure-parameter distribution at time t_{k+1} as the prior distribution. With a well-posed modeling system, this recursive process of conditioning parameters on available observations would gradually reduce uncertainty associated with the model structure-parameter set and lead to a progressively smaller region of high probability density (HPD) in the model-parameter space. In some cases, the sampling limitation of the Monte Carlo approach may lead to the HPD parameter region converging to one single point [*Beven and Young*, 2003; *Gupta et al.*, 2003]. *Misirli* [2003] proposed an improvement on the BaRE methodology by including a resampling technique to reduce the effect of the sampling limitation.

[57] Like GLUE, the BaRE methodology introduced a broader paradigm for parameter estimation without resorting to traditional optimization techniques. By adopting a recursive rather than “batch” approach, BaRE allows model parameters to behave as though time-variant and also reduces the dependence on availability of substantial input and output data before estimation can begin. More important, BaRE explicitly considers the uncertainties associated with model-parameter selection and output measurements, which has not been possible for most previous model calibration studies through parameter optimization, and explicitly represents these in the state and output predictions.

[58] Nevertheless, the BaRE methodology is not the final word on what can be achieved for model-parameter estimation. First of all, input data uncertainty and model structural uncertainty are not specifically separated out and are only implicitly considered, by expanding the predictive uncertainty bounds in a somewhat subjective manner. In addition, in the current BaRE methodology for which parameter estimation is the primary focus, the outputs and associated uncertainty remain un-updated after the posterior parameter distributions are obtained; in other words, the effects of reduction on parameter uncertainty (through incorporating new knowledge from available observation) do not properly propagate to the estimation of outputs and associated uncertainty in a timely manner. Accordingly, it would be beneficial to conduct simultaneous state and parameter estimation to generate unbiased parameter estimates, as well as more accurate state estimates. Several such approaches are reviewed in section 8. Finally, given that system structures and model parameters naturally vary slowly in time, it would be more appropriate to employ a time interval

sufficiently larger than the typical observation time step when performing model structure-parameter estimation. In other words, better results may be achieved by adopting an estimation algorithm that combines the advantages of batch and recursive methods through using an assimilation time interval of proper length.

7. Methods for State Estimation

[59] State estimation for dynamic systems is a process where information is extracted from observations and accumulated in time into the model, propagating to all state variables. For a well-behaved model with consistent constraints of physical properties of the system, improved state estimates can be obtained through data assimilation. This section focuses on state estimation methods assimilating observations that are distributed in time. Given observations available up to the current time, there are three types of state estimation problems: (1) smoothing problems that seek to characterize system states at a past time; (2) filtering problems that seek to characterize system states at the current time; and (3) forecasting problems that seek to characterize system states at a future time point [Gelb, 1974; McLaughlin, 2002]. Smoothing problems are usually found in reanalysis or retrospective studies, while filtering and forecasting problems are most commonly seen in real-time or operational forecasting applications. In dealing with these problems, batch-processing methods (or smoothers) are employed to estimate model states in a batch mode through least squares approximations, while sequential methods (or filters) are typically used for recursive estimation/correction of the states of a system each time an observation becomes available. In hydrologic data assimilation the most commonly used methods are Kalman filtering, particle filtering, and variational data assimilation. These methods are explained in detail below, with an introduction to the state-space formulation commonly used for state estimation applications.

7.1. State-Space Formulation

[60] For the convenience of illustrating the different state estimation methods, let us consider the following generic dynamic state-space formulation of a stochastic model:

$$x_{k+1} = M_{k+1}(x_k, \theta, u_{k+1}) + \eta_{k+1} \quad (20)$$

$$z_{k+1} = H_{k+1}(x_{k+1}, \theta) + \varepsilon_{k+1}, \quad (21)$$

where x_k and x_{k+1} represent the true system state vectors at time t_k and t_{k+1} , respectively; the nonlinear operator M_{k+1} (equivalent to the model structure mentioned earlier in section 3) expresses the system propagation from time t_k to t_{k+1} in response to the model input vector u_{k+1} ; θ is a vector of time-invariant model parameters; the observation vector z_{k+1} is related to the model parameters and states through an observation operator H_{k+1} (equivalent to \mathcal{M} mentioned in equation (10)); η_{k+1} denotes the model error with mean $\bar{\eta}_{k+1}$ and covariance Q_{k+1} ; and ε_{k+1} denotes the observation error with mean $\bar{\varepsilon}_{k+1}$ and covariance R_{k+1} . In the context of Bayesian updating (equation (13)), the state equation (20) represents the model prior at time t_{k+1} , while the observation equation (21) can be used to calculate the likelihood of the

observation z_{k+1} . Note in the literature, the state equation is also referred to as “transition equation,” “forward model,” “forecast model,” or “dynamic system”; and the observation equation is often referred to as “measurement equation/model/system.”

[61] To set up the assimilation system using the above state-space formulation, some assumptions have to be made on the statistics of the two error terms η and ε , based on the prior knowledge of the deficiencies in the assimilating system. For example, the mean values of η and ε (i.e., biases) reflect the systematic errors in the modeling and observation systems, while the error covariances Q_k and R_k in particular reflect the uncertainty in the model predictions and observations. In practice, since these error characteristics cannot be observed directly and are difficult to estimate via indirect methods such as calibration, approximations to the error PDFs are typically unavoidable [e.g., Reichle *et al.*, 2001a, 2001b]. One popular approach is to assume that the errors are zero-mean white noise sequences with a normal (i.e., Gaussian) probability distribution. In addition, it is typically assumed that the model error and observation error are uncorrelated in order to obtain optimal estimates.

7.2. Kalman Filtering

[62] In the case of Gaussian model and measurement errors and linear model and observation operators, the data assimilation problem presented in (20) and (21) can be easily solved by an optimal recursive data processing algorithm known as the Kalman filter (or KF [Kalman, 1960]). The KF algorithm originates from the optimal least squares analysis and consists of recursive implementation of a prediction step (equations (22) and (23)) and an update step (equations (24) and (25)) as follows:

$$x_{k+1}^- = M_{k+1}(x_k^+, \theta, u_{k+1}) \quad (22)$$

$$P_{k+1}^- = \mathbf{M}_{k+1} P_k^+ \mathbf{M}_{k+1}^T + Q_{k+1} \quad (23)$$

$$x_{k+1}^+ = x_{k+1}^- + K_{k+1} d_{k+1} \quad (24)$$

$$P_{k+1}^+ = P_{k+1}^- - K_{k+1} \mathbf{H}_{k+1} P_{k+1}^-, \quad (25)$$

where P is the error covariance matrix of the state variables; \mathbf{M} and \mathbf{H} stand for the linear (or “linearized” in nonlinear cases) model operator and observation operator presented in matrix forms, respectively; the minus and plus superscripts are used to discriminate the states and the error covariance matrix before and after updating, respectively; T stands for transpose; and d is the innovation vector and is defined as the difference between the actual observation z and the model forecast of z (denoted as z^-), i.e.,

$$d_{k+1} = z_{k+1} - z_{k+1}^- \quad (26)$$

$$z_{k+1}^- = H_{k+1}(x_{k+1}^-, \theta). \quad (27)$$

K is called the Kalman gain and can be calculated as follows:

$$K_{k+1} = \frac{P_{k+1}^- \mathbf{H}_{k+1}^T}{\mathbf{H}_{k+1} P_{k+1}^- \mathbf{H}_{k+1}^T + R_{k+1}}. \quad (28)$$

[63] The calculations of (22)–(28) can be repeated at the next time step $k + 2$ to assimilate a new observation available at that time; and this process can progress sequentially into the future to assimilate all available observations if desired. Note by updating the states with equation (24), the assimilation algorithm does not explicitly comply with fundamental physical principles such as conservation of mass, momentum, and energy within the model system.

[64] Equation (28) shows that the Kalman gain K is determined by the relative magnitudes of the state error covariance P and the observation error covariance R and acts as a weighting factor on the innovation term. In other words, the larger the observation error covariance, the smaller the Kalman gain, and the smaller the update correction applied to the forecast state vector. This indicates that the assimilation results can be highly sensitive to the choice of the priors, i.e., the statistics of model structural, parameter, and measurement errors. It is worth noting that low correlation between model states and observations will also result in a small Kalman gain, suggesting the importance of using appropriate observations in an assimilation study.

[65] The KF algorithm described above is easy to implement and has proved effective and efficient in the case of linear system dynamics [e.g., *Eigbe et al.*, 1998; *Galantowicz et al.*, 1999]. However, in practice, hydrologic systems are often inevitably highly nonlinear, limiting the use of Kalman filtering. Hence variations of the KF algorithm have been developed to make it applicable to nonlinear problems, including the commonly used extended Kalman filter (EKF [*Jazwinski*, 1970]) and ensemble Kalman filter (EnKF [*Evensen*, 1994]).

[66] In the EKF algorithm, local (tangent linear) approximation of the nonlinear state and measurement equations (i.e., the model operator M and the observation operator H) is performed each time data assimilation is conducted. When implementing the EKF, the same equations (22)–(27) for the KF algorithm will be used; however, the linearized forms of the model and observation operators (\mathbf{M} and \mathbf{H}) will be used in those equations. Some successful applications of the EKF have been seen in the hydrological literature [*Katul et al.*, 1993; *Entekhabi et al.*, 1994; *Walker and Houser*, 2001]; the EKF, however, may produce instabilities or even divergence due to closure approximation by neglecting the second- and higher-order derivatives of the model [*Evensen*, 1994].

[67] *Evensen* [1994] introduced the ensemble Kalman filtering (or EnKF) algorithm as an alternative to the EKF to address difficulties arising from high-dimensional nonlinear filtering problems. By making a Monte Carlo generation from random input perturbations, EnKF nonlinearly propagates an ensemble of model states using (20), maps them to an ensemble of prior estimates of the observations using (21), and then updates the prior ensemble based on the Kalman gain. The EnKF still consists of a prediction step

(equation (29)) and an update step (equation (30)) as follows:

$$x_{k+1}^{-i} = M_{k+1}(x_k^{+i}, \theta, u_{k+1}^i) \quad i = 1, \dots, n \quad (29)$$

$$x_{k+1}^{+i} = x_{k+1}^{-i} + K_{k+1} d_{k+1}^i \quad i = 1, \dots, n, \quad (30)$$

where n is the size of the ensemble; the input ensemble u_{k+1}^i is obtained by adding a noise term ζ_{k+1}^i to the nominal input u_{k+1} , i.e., $u_{k+1}^i = u_{k+1} + \zeta_{k+1}^i$ ($\zeta_{k+1}^i \sim N(0, U_{k+1})$); and U_{k+1} is the error covariance of u_{k+1} . A noise term ε_{k+1}^i can also be added to the nominal observation z_{k+1} to calculate the innovation ensemble using the following two equations (as compared with (26) and (27)):

$$d_{k+1}^i = z_{k+1} + \varepsilon_{k+1}^i - z_{k+1}^{-i} \quad \varepsilon_{k+1}^i \sim N(0, R_{k+1}), i = 1, \dots, n \quad (31)$$

$$z_{k+1}^{-i} = H_{k+1}(x_{k+1}^{-i}, \theta) \quad i = 1, \dots, n. \quad (32)$$

[68] Unlike in the EKF, no linearization of M or H is needed. More important, the prior (or prediction) error covariance P_{k+1}^- of the state variables can be directly calculated from the ensemble $\{x_{k+1}^{-i}\}$ as expressed in (33), saving substantial computation resources in propagating and updating P using (23) and (25),

$$P_{k+1}^- = \Sigma_{k+1}^{xx} = E[X_{k+1}^- (X_{k+1}^-)^T], \quad (33)$$

where Σ denotes covariance and $X_{k+1}^- = \{x_{k+1}^{-i}\}_{i=1}^n$. In fact, the state error covariance P is never explicitly needed in EnKF, for the $P_{k+1}^- H_{k+1}^T$ term in (28) is essentially the cross error covariance of the state prediction $\{x_{k+1}^{-i}\}$ and the observation prediction $\{z_{k+1}^{-i}\}$, i.e.,

$$P_{k+1}^- H_{k+1}^T = \Sigma_{k+1}^{xz} = E[X_{k+1}^- Z_{k+1}^-], \quad (34)$$

where $Z_{k+1}^- = \{z_{k+1}^{-i}\}_{i=1}^n$. Similarly, the prediction error covariance in the observation space (i.e., the $H_{k+1} P_{k+1}^- H_{k+1}^T$ term in (28)) can be calculated from $\{z_{k+1}^{-i}\}$ as follows:

$$H_{k+1} P_{k+1}^- H_{k+1}^T = \Sigma_{k+1}^{zz} = E[Z_{k+1}^- (Z_{k+1}^-)^T]. \quad (35)$$

Consequently, the Kalman gain in the EnKF algorithm can be easily derived by substituting (34) and (35) into the following equation:

$$K_{k+1} = \Sigma_{k+1}^{xz} (\Sigma_{k+1}^{zz} + R_{k+1})^{-1}. \quad (36)$$

Similar to the standard KF and EKF algorithms, EnKF can also be implemented recursively in time to sequentially assimilate observations as they become available.

[69] The applicability to nonlinear problems and easy implementation of the EnKF method has led to extensive applications of this DA technique in hydrology, meteorology, and other fields [e.g., *Burgers et al.*, 1998; *Margulis et al.*, 2002; *Reichle et al.*, 2002a, 2002b; *Moradkhani et al.*, 2005a; *Vrugt et al.*, 2005].

7.3. Particle Filtering

[70] Particle filtering (PF) is another commonly used data assimilation algorithm for recursive estimation of model states. In the literature the algorithm is also known as bootstrap filtering, the condensation algorithm, sequential Monte Carlo (SMC) sampling, interacting particle approximations, and survival of the fittest [Arulampalam *et al.*, 2002]. In particle filtering, the posterior probability distribution (PPD) of model states at time t_{k+1} is characterized by a set of discrete random particles ($\{x_{k+1}^i\}_{i=1}^n$) with associated importance weights ($\{w_{k+1}^i\}_{i=1}^n$) as follows:

$$p(x_{k+1}|z_{1:k+1}) \approx \sum_{i=1}^n w_{k+1}^i \delta(x_{k+1} - x_{k+1}^i), \quad (37)$$

where n is the number of particles and δ denotes the Dirac delta function. If n is sufficiently large, the discrete expression on the left-hand side of (37) becomes an effective approximation to the PPD of the true state space at time t_{k+1} .

[71] The PPD is best represented if the particles are directly sampled from the posterior distribution of the states, which, however, is generally not possible. To circumvent this obstacle, a sequential importance sampling (SIS) strategy has typically been adopted, where a proposal distribution $q()$ (referred to as importance density in the literature) is used and the importance weights are calculated as follows:

$$w_{k+1}^{i(*)} = p(x_{k+1}^i|z_{1:k+1})/q(x_{k+1}^i|z_{1:k+1}) \quad i = 1, \dots, n, \quad (38)$$

where $\{w_{k+1}^{i(*)}\}_{i=1}^n$ are the weights before normalization (i.e., $w_{k+1}^i = w_{k+1}^{i(*)}/\sum_{i=1}^n w_{k+1}^{i(*)}$). In practice, equation (38) can be rearranged as below to allow recursive evaluation of the importance weights as successive observations become available (see Arulampalam *et al.* [2002] for detailed derivation):

$$w_{k+1}^{i(*)} = w_k^{i(*)} \frac{p(z_{k+1}|x_{k+1}^i)p(x_{k+1}^i|x_k^i)}{q(x_{k+1}^i|x_k^i, z_{k+1})} \quad i = 1, \dots, n. \quad (39)$$

[72] Choice of an appropriate proposal importance density is crucial in the SIS algorithm as reported by several studies [e.g., Doucet *et al.*, 2000; Arulampalam *et al.*, 2002]. In a generic approach the importance density is often conveniently chosen to be the prior; and the weight calculation in (39) simplifies to

$$w_{k+1}^{i(*)} = w_k^{i(*)} p(z_{k+1}|x_{k+1}^i) \quad i = 1, \dots, n. \quad (40)$$

This renders the importance weights proportional to the likelihood $p(z_{k+1}|x_{k+1}^i)$ calculated using the observation equation (21).

[73] Particle filtering based on the above SIS algorithm consist of recursively propagating the particles using (20) and updating the importance weights associated with each particle using (21) and (40) as successive observations become available in time. Compared with the Kalman filtering algorithms discussed earlier (i.e., the standard KF,

EKF, and EnKF), PF performs updating on the particle weights instead of the state variables. In addition, PF has the desirable characteristics of being applicable to any state-space model of any form, linear or nonlinear, Gaussian or non-Gaussian.

[74] Implementation of the SIS particle filter in practice, however, may often be complicated by the well-known degeneracy problem where many particles are found to have negligible weights after a few iterations, thus making little or no contribution to the final representation of the posterior distribution [Doucet *et al.*, 2000] (Note that this same problem arose in the implementation of the BaRE algorithm, which has conceptual and implementational similarities.) As a result, only a small number of particles effectively participate in the filtering process according to the following measure [Doucet *et al.*, 2000; Arulampalam *et al.*, 2002]:

$$N_{eff} \approx 1 / \sum_{i=1}^n (w_{k+1}^i)^2, \quad (41)$$

where N_{eff} is the effective sample size that can be used to measure the degree of degeneracy in the filter. In general, the required number of particles n is likely to increase with the dimension of the state vector, the overlap between the prior and the likelihood, and the required number of time steps for filter operation; there exists, however, no universal provable criterion for defining the minimum effective sample size required to achieve a satisfactory approximation to the true PPD of the state vectors [Gordon *et al.*, 1993].

[75] In practice, to reduce the effect of the degeneracy problem, a resampling procedure is usually added to the SIS algorithm when there exists significant degeneracy (i.e., when N_{eff} is below a certain predefined threshold). The resampling step involves eliminating particles with small weights by replacing them with high-weight particles and then applying uniform weights to all the particles [e.g., Arulampalam *et al.*, 2002; Moradkhani *et al.*, 2005b]. When resampling is applied at each step (without evaluating N_{eff}), the standard SIS algorithm becomes the sampling importance resampling (SIR) filter, a special case of the SIS filter.

[76] Although resampling can reduce the effect of degeneracy, it also introduces another practical problem known as sample impoverishment due to loss of diversity among particles, especially for systems with small noises. In the case of severe sample impoverishment, all particles may converge to one single point in the state space, rendering poor final representation of the posterior distribution. Musso *et al.* [2001] introduced a modified PF known as the regularized particle filter (RPF) to solve the above problem by resampling from a continuous approximation to the importance density, instead of a discrete approximation as the SIR does.

[77] For more details on the implementation and applications of particle filtering and various SMC methods, the readers are referred to Gordon *et al.* [1993], Carpenter *et al.* [1999], Crisan *et al.* [1999], Doucet *et al.* [2001], Arulampalam *et al.* [2002], and Djurić *et al.* [2003].

7.4. Variational Data Assimilation (VDA)

[78] Unlike Kalman filtering and particle filtering, which approach the assimilation in a sequential manner, variational

methods operate in a batch-processing manner over a given time window which contains a sequence of observation time points. Hence variational methods are smoothers and mostly suitable for solving smoothing problems. Theoretically, VDA methods can also be used for filtering problems if a new smoothing problem is defined sequentially at each observation time point; this, however, can be computationally inefficient for real-time applications where the measurement vector z needs to be expanded indefinitely as new observations arrive continually. Depending on the spatial and temporal dimensions of the state variable, VDA methods can be one-dimensional (1D-Var), three-dimensional (3D-Var), or four-dimensional (4D-Var) (see unpublished lecture available at www.ecmwf.int/newsevents/training/rcode_notes/pdf_files/Assim_concepts.pdf).

[79] For illustration purposes we assume that the prior estimate of state variables at time t_0 is x_0^- (with error covariance Q_0); and the assimilation is to operate over the time interval $[t_1, t_n]$, with observations $[z_1, z_2, \dots, z_n]$ available at the n discrete time points $[t_1, t_2, \dots, t_n]$. A general variational data assimilation problem can then be defined as the minimization of the following cost function J , which represents the aggregated error over the entire assimilation window (assuming that errors at different times are independent and additive):

$$\begin{aligned} J(x, u, \theta) = & \sum_{i=1}^n \eta_i^T Q_i^{-1} \eta_i + \sum_{i=1}^n (z_i - H_i[x_i])^T R_i^{-1} (z_i - H_i[x_i]) \\ & + (x_0 - x_0^-)^T Q_0^{-1} (x_0 - x_0^-) \\ & + \sum_{i=1}^n (u_i - u_i^-)^T C_{uu}^{-1} (u_i - u_i^-) \\ & + (\theta - \theta^-)^T C_\theta^{-1} (\theta - \theta^-) \\ = & J_M + J_O + J_0 + J_u + J_\theta, \end{aligned} \quad (42)$$

where η_i represents the model error at t_i ; u_i^- and θ^- denote the prior model inputs at t_i and the prior time-invariant parameters, respectively; and C_{uu} and C_θ are the time-invariant error covariances of inputs and parameters, respectively. The purpose of variational data assimilation is, by means of minimizing J , to obtain the least squares estimates of state variables x_i and input variables u_i for each time point within the assimilation window and the time-invariant parameters θ . The minimization problem is subject to the strong constraint that the state, input, and parameter estimates obtained by VDA must be consistent with the state equation (20). Alternatively, one can turn the constrained minimization problem into an unconstrained one by adjoining the state equation to the cost function (42) with a Lagrange multiplier λ as follows:

$$J(x, u, \theta) = J_M + J_O + J_0 + J_u + J_\theta + \sum_{i=1}^n \lambda_i^T [x_i - M_i(x_{i-1}, u_i, \theta)]. \quad (43)$$

[80] In the above general formulations of the cost function, the first term J_M penalizes the difference between the estimated model error vector η_i and its prior mean (assumed to be zero in this case); the second term J_O is used to penalize the differences between model predictions and

observations at all time points within the assimilation window; and J_0 , J_u , and J_θ are included to measure the errors associated with the initial conditions, model inputs and parameters, respectively. When summed together to form the aggregated cost function, each of the errors is weighted by the corresponding error covariance (i.e., Q , R , Q_0 , C_{uu} , or C_θ). In this general VDA framework, errors from various sources (e.g., the model, observations, initial conditions, inputs, and parameters) can be collectively taken into account.

[81] In practice, however, nonlinear, high-dimensional hydrologic applications render the comprehensive optimization problem as represented by (42) very difficult, and often impossible, to solve. Consequently, simplifications and approximations are often introduced by, for example, neglecting model/parameter errors and/or linearizing the state and observation equations. Even with simplifications, solving a VDA problem analytically is not easy, and often a numerical algorithm such as the adjoint model technique is used to obtain solutions in an iterative manner.

[82] To illustrate the implementation process of variational data assimilation, we consider a simple VDA system where the objective is to minimize the following cost function with only the measurement term J_O considered:

$$J(x) = J_O = \sum_{i=1}^n (z_i - H_i[x_i])^T R_i^{-1} (z_i - H_i[x_i]). \quad (44)$$

[83] According to the state equation (20), given θ , u_i , and η_i (assumed to be zero in this case), the state prediction at t_i (i.e., x_i) is solely dependent on prediction at the previous time step t_{i-1} (i.e., x_{i-1}), which is in turn solely dependent on x_{i-2} . This indicates that x_i is ultimately determined by the initial condition x_0 , the only fundamental unknown in this VDA problem. The objective of VDA is then to find the best estimate of x_0 that minimizes $J_O(x_0)$. The optimization process requires the evaluation of both the cost function and its gradient $\nabla J_O(x_0)$, which can be computed as follows using the adjoint technique (see detail derivations by *Huang and Yang* [1996]; see also unpublished note available at <http://citeseer.ist.psu.edu/huang96variational.html>):

$$\nabla J_O(x_0) = \sum_{i=1}^n \left[\prod_{k=0}^{i-1} \mathbf{M}_k^T \right] \mathbf{H}_i^T d_i \quad (45)$$

$$d_i = R_i^{-1} (z_i - H_i(x_i, \theta))$$

where \mathbf{M}_k^T is the transpose of the tangent linear model of M_k at point x_k ; similarly, \mathbf{H}_i^T denotes the transpose of the tangent linear model of H_i at point x_i and d_i is the normalized difference between model prediction and the observation at time t_i . For computational efficiency, we define the adjoint model at time t_i as

$$\tilde{x}_{i-1} = \mathbf{M}_i^T (\tilde{x}_i + \mathbf{H}_i^T d_i), \quad (46)$$

where \tilde{x}_i is called the adjoint variable. It can be proved that if we start from the end point of the assimilation interval t_n with \tilde{x}_n initialized to zero and then integrate the adjoint model

(46) backward in time to the initial time t_0 , we obtain \tilde{x}_0 , which is exactly equal to $\nabla J_O(x_0)$ defined by (45). With this method of computing the cost function gradient, a VDA problem can be solved through an iterative minimization process to identify a best estimate of x_0 , which can then be used to compute the value of x at any time point within the assimilation window by integrating the state equation (20) forward in time.

[84] A number of examples of designing and solving a VDA problem can be found in the literature. *Huang and Yang* [1996] discussed in detail the general procedure to construct a VDA system using the adjoint technique based on a nonlinear mathematical model, with only observation errors considered in the cost function. Applications of similar VDA techniques are given by *McLaughlin* [1995, 2002], *Bouttier and Courtier* [1999], *Reichle et al.* [2001a, 2001b], and *Seo et al.* [2003]. The readers are referred to these references for more details of using the VDA algorithms.

[85] Compared with the sequential equivalents KF and EKF, the VAR methods are preferable for data assimilation in a realistic, complex system (e.g., a numerical weather prediction framework) because they are much less expensive computationally than KF and EKF methods. In addition, by using observations inside the assimilation interval all at once, VDA methods are also more optimal than KF and EKF methods inside (within) the interval (at the end of the interval, VDA and KF methods are expected to give the same results for linear systems; in the presence of high nonlinearity, the results from the two methods may diverge because VDA gives the mode of an uncertain variable while KF estimates the expected value). However, the sequential KF methods are more suitable for real-time data assimilation to process observations that arrive continuously in time, while VDA methods can only be run for a finite time interval; also, KF methods provide error covariance estimates for the prediction, while a VDA method itself does not provide any estimate of the predictive uncertainty. When the assimilation system is nonlinear, both EKF and VDA methods rely on using the tangent-linear models M and H to approximate the state and observation equations; if the nonlinearity is important, it makes more sense to use ensemble (or Monte Carlo) approaches such as EnKF and PF for data assimilation.

8. Simultaneous State and Parameter Estimation

[86] In general, parameter estimation tends to focus on uncertainty in the parameter estimates only, while neglecting partial or all of the other uncertainty sources. On the other hand, state estimation via data assimilation methods, although having the potential for explicitly handling various uncertainties arising from model inputs and observations, typically does not take into account the uncertainties associated with model parameters. In either case, there is a tendency to generate biased model predictions due to biased parameter and/or state estimates. Hence it would be desirable to combine parameter estimation with state estimation to account for all kinds of uncertainties.

[87] *Vrugt et al.* [2005] applied a simultaneous optimization and data assimilation (SODA) approach to estimate both states and parameters of two hydrological models. The SODA approach estimates model parameters using the

batch calibration strategy SCEM, with EnKF updating of state estimates performed at each time step in each model run during the calibration process. This way of combining optimization with data assimilation is conceptually simple and easy to implement. Preliminary results show that the SODA approach is able to produce both less biased parameters and less biased model states, compared with the results from using only SCEM or EnKF. The SODA approach is certainly a step forward from traditional parameter estimation and data assimilation methodologies, in that it reasonably uses calibration to correct long-term systematic biases due to parameter uncertainties and uses ensemble data assimilation to correct short-term or instantaneous system biases associated with model states, data, and other sources of errors. The SODA approach, however, is still not wholly satisfactory in that it optimizes model parameters in one single batch, without allowing parameters to vary over time while also requiring considerable computational time.

[88] *Moradkhani et al.* [2005a, 2005b] presented two dual state-parameter estimation methods based on EnKF and PF, respectively. These two methods were designed to recursively estimate both states and parameters using two parallel filters. In these methods, Monte Carlo sampling and sequential updating (via EnKF or PF techniques) are applied to not only a vector of state variables, but also to a different vector of model parameters at each assimilation time step. Accordingly, the probability distributions of both model states and parameters are (independently) recursively updated each time a new observation is available. In these approaches, better state and parameter estimates enable the modeling system to evolve consistently over time and make improved predictions with proper uncertainty bounds. Along the same lines, *Labarre et al.* [2006] presented an approach to jointly estimate the model states and the hyperparameters of the data assimilation algorithm using the mutually interactive state and parameter estimation (MISP) technique [*Todini*, 1978a, 1978b] with two conditionally linked Kalman filters running in parallel.

[89] Another way of conducting joint state-parameter estimation is to extend the current state vector with the model parameters, a technique known as “state augmentation” [e.g., *Gelb*, 1974; *Drécourt et al.*, 2005]. Here the model parameters are recast as state variables to form an extended state vector; and the simultaneous state and parameter estimation problem is reduced to a state estimation problem. If we assume the parameters are time-variant with normally distributed errors ξ_k at time t_k ($\xi_k \sim N(0, V_k)$), then with state augmentation, the new model equation and observation equation can be expressed as follows (as opposed to the original equations (20) and (21)):

$$\begin{bmatrix} x_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_k & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} \eta_k \\ \xi_k \end{bmatrix} \text{ or } x'_{k+1} = \mathbf{M}'_k x'_k + \eta'_k \quad (47)$$

$$\begin{bmatrix} y_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_k & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_k \\ \xi_k \end{bmatrix} \text{ or } y'_{k+1} = \mathbf{H}'_k x'_k + \varepsilon'_k, \quad (48)$$

where x' , y' , \mathbf{M}' , \mathbf{H}' , η' , and ε' are the new state vector, observation vector, model operator, observation operator, model error, and observation error, respectively. These new

quantities should be used in place of the old ones when a data assimilation algorithm (e.g., EKF or EnKF) is applied to recursively update the states and parameters simultaneously. Being conceptually simple, this method, however, may render the estimation process unstable and intractable because of complex interactions between states and parameters in nonlinear dynamic systems [Todini, 1978a, 1978b]. In addition, since parameters generally vary much more slowly than the system states, unstable problems may also result from the fact that both model states and parameters are updated at each observation time step in this method. This same argument may apply to the dual state-parameter estimation methods presented by Moradkhani *et al.* [2005a, 2005b].

9. An Integrated Uncertainty Framework for Hydrological Modeling

[90] The data assimilation methods introduced in sections 5–8 are designed for system identification, parameter estimation, state estimation, and combined state and parameter estimation, respectively. As described in section 4, Bayes' theorem is the fundamental basis of these DA methods. However, one criticism of Bayesian methods is that the computation of posteriors depends on prescribed priors that could be wrong, rendering the possibility of unrealistic uncertainty estimation. For example, state estimation is often conducted under the assumption that the model structure and parameters are correct (i.e., assuming the model/parameter priors to be unity), which is hardly the case in hydrologic modeling. To adequately quantify the total uncertainty in hydrologic predictions and to maximally reduce it, we shall consider an integrated uncertainty framework that can facilitate the implementation of all the three types of DA applications in a cohesive, systematic manner.

[91] Berliner [1996] introduced a Bayesian hierarchical modeling (BHM) approach to complex environmental DA problems, where one can obtain the joint distribution of the model process x and parameters θ , given observational data z , by computing a hierarchy of conditional models based on Bayesian rules as follows (see also Winkle [2003]):

$$p(x, \theta | z) \propto p(z | x, \theta) p(x | \theta) p(\theta). \quad (49)$$

With the logical progressive chain of conditional dependence $\{M \rightarrow \theta \rightarrow x \rightarrow y\}$ as described in section 4, we can derive the following integrated hierarchical framework in a manner analogous to the BHM approach described above:

$$\begin{aligned} p(y, x, \theta, M | z^y, z^x, z^\theta, z^M) &\propto p(M | z^M) \quad (1) \text{system identification} \\ &\cdot p(\theta | z^\theta, M) \quad (2) \text{parameter estimation} \\ &\cdot p(x | z^x, \theta, M) \quad (3) \text{state estimation} \\ &\cdot p(y | z^y, x, \theta, M) \quad (4) \text{output prediction} \end{aligned} \quad (50)$$

where z^M , z^θ , z^x , z^y denote independent observations on which the model structure, parameters, states, and outputs are progressively conditioned, respectively. To assess the total uncertainty of a hydrological model, it is really critical

that the joint output-state-parameter-structure distribution $p(y, x, \theta, M)$ (instead of the individual distributions) be examined, considering the complex interactions between the model outputs, states, parameters, and structure. The purpose of the hierarchical approach is to obtain this complicated joint distribution, which is difficult to compute directly, by factoring it into a sequence of conditional probabilities that are easier to characterize utilizing available knowledge and data.

[92] In the framework defined in (50), one starts with (1) system identification by computing $p(M | z^M)$ given the observation z^M and some prior knowledge of the model parameters; (2) the model parameters are then estimated through a parameter estimation technique given a different set of observations z^θ and the models M (and their probability distribution obtained in the previous step; (3) with the observation z^x , and M and θ obtained in (2), the distribution of the model states x can be updated; and (4) finally, with M , θ , and x defined in the last three steps and the observation z^y , one can predict the final uncertainty reflected in the model outputs y . The joint distribution of the model outputs, states, parameters, and structures, which captures the complex interactions among these components, can be obtained by multiplying the four conditional probability terms together as shown in (50). When a period of new observations arrives, the above progressive steps can be repeated to update the conditional probability distributions and the overall joint distribution.

[93] In implementing this uncertainty framework, one should choose a most suitable data assimilation method for each of the four steps. In particular, special attention should be paid to the appropriate timescales when deciding which DA method to use. For example, as discussed in section 3, we expect model parameters to vary much more slowly in time than the states and outputs. In this sense, it would be appropriate to adopt a smoothing or variational approach for parameter estimation in step 2 and a sequential or filtering approach for state estimation in step 3. In principle, a method used for state estimation is expected to be applicable to output prediction as well, for model states and outputs generally share the same dynamic characteristics/frequencies. We can also reasonably assume that the model structure does not vary with time or varies at a very low frequency (even lower than that of the parameters). Precisely how to adjust the existing DA methods so that they can be cohesively nested within the above integrated hierarchical framework remains an important topic for further research.

10. Summary and Discussions

[94] Application of data assimilation techniques to hydrologic modeling is relatively new, and there is a lack of general guidance in the hydrologic literature on how to choose and implement a suitable DA technique so that the hydrologic uncertainty is properly considered. This may come to limit the extensive application of hydrologic data assimilation. On the other hand, it has not been realized by the general hydrologic community that the traditional data assimilation focus on state estimation alone is not sufficient for adequate consideration of the uncertainties associate with all sources in hydrologic modeling. In most cases,

uncertainty in model structures and parameters is ignored in data assimilation applications.

[95] In this analysis we have discussed the three critical aspects of addressing hydrologic uncertainty, namely, understanding, quantifying, and reducing uncertainty, to arrive at a general context for hydrologic data assimilation. The intention of this paper is to provide not an extensive review of all the data assimilation techniques and applications existing in the hydrologic literature, but a discussion of the main recent developments, potential future directions, and some open issues in hydrologic data assimilation. We explore uncertainties associated with different sources from a systems perspective, leading to the definition of the three major types of DA problems: system identification, parameter estimation, and state estimation. Bayesian techniques for addressing these uncertainty problems and typical methods used in the hydrologic literature were then described in relevant detail to provide sufficient guidance on how to properly implement these methods. To adequately quantify the hydrologic predictive uncertainty and reduce it to a maximum degree, we call for the adoption of an integrated framework such as the one proposed in section 9, where system identification, parameter estimation, state estimation, and ultimately output prediction are progressively conducted in a cohesive, systematic manner. Proper implementation of all these DA problems within such a single, integrated framework would greatly improve the effectiveness and efficiency in extracting information from available data and assimilating it into hydrologic predictions.

[96] Nevertheless, there remain critical issues that need to be properly addressed before the proposed integrated framework can be implemented to realize its maximum potential. For example, the exposition assumes that we have successfully described the information extraction process via the definition of suitable likelihood functions [e.g., *Beven and Young*, 2003; *Gupta et al.*, 2003], a topic that merits rethinking and further research. Also, the fundamental Bayesian rule requires that the error models used as inputs to data assimilation applications be properly prescribed from prior information, which is often difficult to satisfy for real-world hydrologic applications. Although we all recognize that real hydrologic systems are seldom linear or close to linear, there often has been no other choice but to prescribe the error distributions as Gaussians; better strategies (e.g., via using mixtures of Gaussians [*Wójcik et al.*, 2006]) are desired to avoid this subjectivity in prescribing various types of errors, and new mathematical developments might be necessary to circumvent this difficulty. Also, we must recognize that a hydrological system (i.e., a model structure) and its physical properties (i.e., model parameters) naturally tend to vary much more slowly in time than the states and fluxes of the system. In other words, proper timescales (or time intervals) should be identified and utilized in the different steps within the framework. In addition, implementing such an integrated framework would require a substantial amount of data, encompassing various observation types that are suitable for different kinds of DA problems including system identification, parameter estimation, and state estimation. Other issues to be addressed may include, for example, handling temporally and spatially correlated errors in hydrological variables [e.g., *Drécourt et al.*, 2005] and resolving the scale differences between model states and

observations. To facilitate addressing these issues, the next generation of hydrologic models should be developed in coordination with developments in DA techniques to facilitate the implementation of the proposed integrated hierarchical framework. Finally, it is useful to be aware that complete accounting for model structure errors might never be achieved, as there exists no means to define a model space (with a set of truly independent model structures) that can perfectly represent the reality.

[97] **Acknowledgments.** This study was supported by SAHRA (an NSF center for Sustainability of Semi-Arid Hydrology and Riparian Areas) under grant EAR-9876800. We thank Gab Abramowitz, Andras Bardossy, Martyn Clark, Shlomo Neuman, Amilcare Porporato, Peter Troch, Ezio Todini, Thorsten Wagener, and the anonymous reviewer for their constructive comments that greatly improved this paper.

References

- Abramowitz, G., H. Gupta, A. Pitman, Y. Wang, R. Leuning, H. Cleugh, and K.-L. Hsu (2006), Neural error regression diagnosis (NERD): A tool for model bias identification and prognostic data assimilation, *J. Hydrometeorol.*, 7, 160–177.
- Anderson, J. L., and S. L. Anderson (1999), A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Mon. Weather Rev.*, 127, 2741–2758.
- Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Processes*, 50(2), 174–188.
- Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, 23(8), 1393–1442.
- Berliner, L. M. (1996), Hierarchical Bayesian time series models, in *Maximum Entropy and Bayesian Methods*, edited by K. Hanson and R. Silver, pp. 15–22, Springer, New York.
- Beven, K. J. (1993), Prophecy, reality, and uncertainty in distributed hydrological modeling, *Adv. Water Resour.*, 16, 41–51.
- Beven, K. J. (2000), Uniqueness of place and non-uniqueness of models in assessing predictive uncertainty, in *Computational Methods in Water Resources XIII*, edited by L. R. Bentley et al., pp. 1085–1091, A. A. Balkema, Rotterdam, Netherlands.
- Beven, K. J., and A. M. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298.
- Beven, K. J., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249, 11–29.
- Beven, K. J., and P. Young (2003), Comment on “Bayesian recursive parameter estimation for hydrologic models” by M. Thieman, M. Trosset, H. Gupta, and S. Sorooshian, *Water Resour. Res.*, 39(5), 1116, doi:10.1029/2001WR001183.
- Blöschl, G., and R. Grayson (2000), Spatial observations and interpolation, in *Spatial Patterns in Catchment Hydrology: Observations and Modeling*, edited by A. Grayson and G. Blöschl, pp. 17–50, Cambridge Univ. Press, New York.
- Boni, G., F. Castelli, and D. Entekhabi (2001), Sampling strategies and assimilation of ground temperature for the estimation of surface energy balance components, *IEEE Trans. Geosci. Remote Sens.*, 39(1), 165–172.
- Burgers, G. P., J. Van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126, 1719–1724.
- Carpenter, J., P. Clifford, and P. Fearnhead (1999), Improved particle filter for nonlinear problems, *IEE Proc. Radar Sonar Navig.*, 146(1), 2–7.
- Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data, *Water Resour. Res.*, 22(2), 228–242.
- Clark, M. P., and A. G. Slater (2006), Probabilistic quantitative precipitation estimation in complex terrain, *J. Hydrometeorol.*, 7(1), 3–22.
- Courtier, P., J. Derber, R. Errico, J.-F. Louis, and T. Vukicevic (1993), Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology, *Tellus, Ser. A*, 45, 342–357.
- Crisan, D., P. Del Moral, and T. J. Lyons (1999), Non-linear filtering using branching and interacting particle systems, *Markov Processes Related Fields*, 5(3), 293–319.
- Cullen, A. C., and H. C. Frey (1999), *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing With Variability and Uncertainty in Models and Inputs*, Springer, New York.

- Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge Univ. Press, New York.
- Djurić, P. M., J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez (2003), Particle filtering, *IEEE Signal Processing Mag.*, 9, 19–38.
- Doblas-Reyes, F. J., M. Deque, and J.-P. Pielou (2000), Multi-model spread and probabilistic seasonal forecasts in PROVOST, *Q. J. R. Meteorol. Soc.*, 126(567), 2069–2088.
- Doucet, A., S. Godsill, and C. Andrieu (2000), On sequential Monte Carlo sampling methods for Bayesian filtering, *Stat. Comput.*, 10, 197–208.
- Doucet, A., N. D. Freitas, and N. Gordon (Eds.) (2001), *Sequential Monte Carlo Methods in Practice*, 581 pp., Springer, New York.
- Drécourt, J.-P., H. Madsen, and D. Rosbjerg (2005), Bias aware Kalman filters: Comparison and improvements, *Adv. Water Resour.*, 29(5), 707–718.
- 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.
- Dunne, S., and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation, *Water Resour. Res.*, 41, W02013, doi:10.1029/2004WR003449.
- Eigbe, U., M. B. Beck, H. S. Wheater, and F. Hirano (1998), Kalman filtering in groundwater flow modeling: Problems and prospects, *Stochastic Hydrol. Hydraul.*, 12, 15–32.
- Entekhabi, D., H. Nakamura, and E. Njoku (1994), Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations, *IEEE Trans. Geosci. Remote Sens.*, 32, 438–448.
- Entekhabi, D., G. R. Asrar, A. K. Betts, K. J. Beven, R. L. Bras, C. J. Duffy, T. Dunne, R. D. Koster, D. P. Lettenmaier, and D. B. McLaughlin (1999), An agenda for land surface hydrology research and a call for the second international hydrological decade, *Bull. Am. Meteorol. Soc.*, 80(10), 2043–2058.
- 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.
- Finkel, A. M. (1990), Confronting uncertainty in risk management: A guide for decision-makers, report, Cent. for Risk Manage., Resour. for the Future, Washington, D. C.
- Galandowicz, J., D. Entekhabi, and E. Njoku (1999), Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness, *IEEE Trans. Geosci. Remote Sens.*, 37, 1860–1870.
- Gelb, A. (Ed.) (1974), *Applied Optimal Estimation*, 374 pp., MIT Press, Cambridge, Mass.
- Georgakakos, K. P., D.-J. Seo, H. V. Gupta, J. Schaake, and M. B. Butts (2004), Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, 298, 222–241.
- Gordon, N., D. Salmond, and A. F. M. Smith (1993), Novel approach to nonlinear and non-Gaussian Bayesian state estimation, *Proc. IEE, Part F*, 140, 107–113.
- Graham, W., and D. McLaughlin (1991), A stochastic model of solute transport in groundwater: Applications to the Borden Ontario tracer test, *Water Resour. Res.*, 27(6), 995–1004.
- 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. Thieman, 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.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005a), The rationale behind the success of multi-model ensemble in seasonal forecasting: I. Basic concept, *Tellus, Ser. A*, 57, 219–233.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005b), The rationale behind the success of multi-model ensemble in seasonal forecasting: II. Calibration and combination, *Tellus, Ser. A*, 57, 234–252.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999),

- elling (ModelCARE 2002), Charles Univ., Prague, Czech Republic, 16 – 20 June.
- Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, *17*, 291–305.
- Palmer, T. N. (2004), Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*(6), 853–872.
- Palmer, T. N., C. Brankovic, and D. S. Richardson (2000), A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations, *Q. J. R. Meteorol. Soc.*, *126*(567), 2013–2033.
- Reichle, R., D. Entekhabi, and D. McLaughlin (2001a), Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach, *Water Resour. Res.*, *37*, 2353–2364.
- Reichle, R., D. B. McLaughlin, and D. Entekhabi (2001b), Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications, *IEEE Trans. Geosci. Remote Sens.*, *39*, 1708–1718.
- Reichle, R., D. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114.
- Reichle, R., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*, 728–740.
- Restrepo, P. J. (1985), A view of maximum-likelihood estimation with large conceptual hydrologic models, *Appl. Math. Comput.*, *17*(4), 375–403.
- Romanowicz, R. J., P. C. Young, and K. J. Beven (2006), Data assimilation and adaptive forecasting of water levels in the River Severn catchment, *Water Resour. Res.*, *42*, W06407, doi:10.1029/2005WR004373.
- Seo, D. J., V. Koren, and N. Cajina (2003), Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting, *J. Hydrometeorol.*, *4*(3), 627–641.
- Tamea, S., F. Laio, and L. Ridolfi (2005), Probabilistic nonlinear prediction of river flows, *Water Resour. Res.*, *41*, W09421, doi:10.1029/2005WR004136.
- Thiemann, T., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535.
- Todini, E. (1978a), Mutually interactive state/parameter estimation (MISP) in hydrological applications, in *Identification and Control in Environmental Systems*, edited by G. C. Vansteenkiste, Elsevier, New York.
- Todini, E. (1978b), Mutually interactive state/parameter estimation (MISP), in *Application of Kalman Filter to Hydrology, Hydraulics and Water Resources*, edited by C.-L. Chiu, Univ. of Pittsburgh, Pittsburgh, Pa.
- Van Geer, F. C., C. B. M. Te Stroet, and Z. Yangxiao (1991), Using Kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations: 1. The role of system noise and its calibration, *Water Resour. Res.*, *27*(8), 1987–1994.
- 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 multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.
- 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.
- Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochastic Environ. Res. Risk Assess.*, *19*, doi:10.1007/s00477-005-0006-5.
- 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.
- Walker, J. P., and P. R. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *J. Geophys. Res.*, *106*, 11,761–774.
- Walker, J. P., G. R. Willgoose, and J. D. Kalma (2001), One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms, *Adv. Water Resour.*, *24*(6), 631–650.
- Wikle, C. K. (2003), Hierarchical Bayesian models for predicting the spread of ecological processes, *Ecology*, *84*(6), 1382–1394.
- Wójcik, R., P. A. Troch, H. Stricker, P. Torfs, E. Wood, H. Su, and Z. Su (2006), Mixtures of Gaussians for uncertainty description in bivariate latent heat flux proxies, *J. Hydrometeorol.*, *7*, 330–345.
- Yangxiao, Z., C. B. M. Te Stroet, and F. C. Van Geer (1991), Using Kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations: 2. Application to monitoring network design, *Water Resour. Res.*, *27*(8), 1995–2006.
- Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrological models, *J. Hydrol.*, *204*, 83–97.
- Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, *40*, W05113, doi:10.1029/2003WR002557.
- Ye, M., S. P. Neuman, P. D. Meyer, and K. Pohlmann (2005), Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff, *Water Resour. Res.*, *41*, W12429, doi:10.1029/2005WR004260.
- Young, P. C. (2002), Advances in real-time flood forecasting, *Philos. Trans. R. Soc. London, Ser. A*, *360*, 1433–1450.
- Young, P. C. (2003), Top-down and data-based mechanistic modeling of rainfall-flow dynamics at the catchment scale, *Hydrol. Processes*, *17*, 2195–2217.
- Ziehmann, C. (2000), Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models, *Tellus, Ser. A*, *52*(3), 280–299.

H. Gupta and Y. Liu, Department of Hydrology and Water Resources, University of Arizona, 845 North Park Avenue, Marshall Building, Room 535, Tucson, AZ 85721-0158, USA. (yqliu@hwr.arizona.edu)