Fundamentals of Data Assimilation and Theoretical Advances

Hamid Moradkhani, Grey Nearing, Peyman Abbaszadeh, and Sahani Pathiraja

Contents

H. Moradkhani (✉) · P. Abbaszadeh
Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL, USA
e-mail: hmoradkhani@ua.edu; pabbaszadeh@crimson.ua.edu

G. Nearing
Department of Geological Sciences, The University of Alabama, Tuscaloosa, AL, USA
e-mail: gsnearing@ua.edu

S. Pathiraja
Institute for Mathematics, University of Potsdam, Potsdam, Germany
e-mail: pathiraja@uni-potsdam.de

Q. Duan et al. (eds.), *Handbook of Hydrometeorological Ensemble Forecasting*,
https://doi.org/10.1007/978-3-642-40457-3_30-1

Abstract

Hydrometeorological predictions are not perfect as models often suffer either from inadequate conceptualization of underlying physics or non-uniqueness of model parameters or inaccurate initialization. During the past two decades, Data Assimilation (DA) has received increased prominence among researchers and practitioners as an effective and reliable method to integrate the hydrometeorological observations from in situ measure and remotely-sensed sensors into predictive models for enhancing the forecast skills while taking into account all sources of uncertainties. The successful application of DA in different disciplines has resulted in an ever-increasing publications. This chapter provides a progressive essay covering fundamental and theoretical underpinnings of DA techniques and their applications in a variety of scientific fields. More detailed examples of applications are presented in following chapters in this section.

Keywords

Hydrometeorological predictions · Uncertainty · Data Assimilation (DA)

1 Introduction

1.1 Purpose of Data Assimilation

Forecasting in hydrometeorology is challenging due to the complex, heterogeneous, nonstationary, and nonlinear interactions between water and the environment. Such complexities make precise modeling of hydrometeorological processes infeasible, leading to persistent uncertainty in forecasting systems. Uncertainties are present in all aspects of hydrometeorological modeling and forecasting, due to errors or imprecision in observations of pertinent states and fluxes, gaps in knowledge of the physical science, and spatiotemporal heterogeneities that complicate the highly dynamic nature of water movement through the land and atmosphere. Such prevalence of uncertainty reduces a forecaster's ability to determine the magnitude and timing of catastrophic events (i.e., floods, droughts) and quantify variables of interest (i.e., water supply, soil moisture). Due to these uncertainties, it is advantageous to utilize the full extent of information about the state of the environment in a single unified forecast. This unification of information from models and observations for reduced uncertainty is the premise of data assimilation (DA).

Most generally, DA is defined as the application of Bayes' theorem to probabilistically condition the states of a dynamical model on observations. There are many different computational techniques for implementing DA, and each technique relies on a different set of tractability approximations. Because no tractable DA technique is perfect, it is important to understand the nature of the particular simulation model and observation data set that are being used for a particular forecast problem.

Most commonly, DA results in improvements to model forecasts by improving initial states (DeChant and Moradkhani 2011a); however it is also possible to use DA to improve model parameters (e.g., Moradkhani et al. 2005a, b; Vrugt et al. 2005; Montzka et al. 2011; Pathiraja et al. 2016a, 2017; Abbaszadeh et al. 2018) and model structures (e.g., Bulygina and Gupta 2011; Nearing and Gupta 2015). DA methods that update initial states can typically be implemented sequentially in time; this is called *filtering*, and filtering approximations make DA particularly applicable to real-time forecasting problems. Forecasting agencies often want to make forecasts at regular time intervals, and a method that sequentially improves initial states is especially applicable to this type of situation.

In addition, many popular DA techniques use ensembles, which allow for probabilistic representation of complex systems and quantification of uncertainties in time-evolving simulations. Ensemble-based DA produces an ensemble of initial states, which allows for initializing an ensemble forecast with a range of possible state values. Ensemble-based DA techniques are therefore especially useful when accurately accounting for forecast uncertainty is important.

1.2 State-Space Models

Since the primary purpose of DA is to improve state estimates within the model, it is important to understand models from a state-space perspective. Dynamical models solve systems of partial differential equations (PDEs), and in hydrology, this is typically done in discrete time. We can therefore write a generalized state-transfer function to represent our discrete-time PDE solution:

$$x_t = f\left(x_{t-1}, u_t, \theta^f\right) + \omega_t \tag{1}$$

In Eq. (1), $f(.)$ is a function that governs the evolution of the model state vector x in discrete time. Note that there are generally many model states, so x_t is generally a vector. Since Eq. (1) is an approximate PDE solution, the model requires some boundary conditions, or time-dependent forcing data (u_t), and also a (typically time-independent) parameter vector (θ^f). Since it is unavoidable that the model will have an error, this is accounted for by an additive error term ω_t which is drawn from a distribution. This is referred to as *model error*. The general DA problem is aimed at reducing model error and therefore improving the accuracy of state estimates.

In addition to the forward model operator, $f(.)$, an observation operator, $h(.)$, is necessary to relate observations with system states:

$$y_t = h\left(x_t, \theta^h\right) + \epsilon_t \tag{2}$$

In Eq. (2), $h(.)$ is the *observation operator*, which relies on the current states (x_t) and a parameter vector (θ^h) to translate states into observation space. Similar to the

forward model operator, the observation operator will have some error, which is typically accounted for using an additive error term like ϵ_t. Although ϵ_t could be specified to retrieve the "true" prediction, it is specified as the correction to reach the observation, which will simplify the explanation of the DA process.

1.3 Types of Data Assimilation

As mentioned above, DA is aimed at improving modeled states by conditioning on observation data. The most general expression for the DA problem is:

$$p\left(x_{1:t}|\, y_{1:t}, u_{1:t}, \theta^f, \theta^h\right) \propto p_h\left(y_{1:t}|\, x_{1:t}, \theta^h\right) p_f\left(x_{1:t}|\, u_{1:t}, \theta^f\right). \tag{3}$$

Equation (3) is Bayes' theorem applied to the problem of estimating model states conditional on observations. p_f and p_h are the probability density functions (PDF) implied by Eqs. (1) and (2), respectively.

Analytical solutions to Eq. (3) are infeasible for almost any real-world problem. In addition, $x_{1:t}$ is very high-dimensional (the dimension of the state vector, x_t, multiplied by the number of time steps), which makes it infeasible to estimate the posterior (i.e., the PDF on the right-hand side of Eq. (3)) by sampling. Thus, to implement Eq. (3) for real-world problems, we almost always require some tractability approximations.

1.3.1 Smoothers Versus Filters

The most common tractability approximation is to restrict information from assimilated observations from moving backward in time. That is, an observation from time t will not affect the state values at times $t-s$, where $s > 0$. When this approximation is used in conjunction with a Markovian model, like Eq. (1), the result is DA *filtering*; in contrast, the full DA problem in Eq. (3) is called *smoothing*. The general filtering problem is as follows:

$$p\left(x_t|\, y_{1:t}, u_{1:t}, \theta^f, \theta^h\right) \propto p_h\left(y_t|\, x_t, \theta^h\right) p_f\left(x_t|\, x_{t-1}, u_t, \theta^f, y_{1:t-1}\right). \tag{4}$$

Notice that the dimension of the posterior is greatly reduced – by a multiplicative factor in the number of time steps. This makes it (sometimes) feasible to sample the posterior effectively at each time step. The Markovian property of the dynamical systems model is explicit in the prior (p_f) in Eq. (4), and it is important to recognize that the prior – i.e., $p_f\ (x_t|\ x_{t-1},\ u_t,\ \theta^f,\ y_{1:t-1})$ – is conditional on past observations.

1.3.2 Linear Versus Nonlinear

One important category of tractability approximations involves treating all or part of the system as linear. The distinction between linear and nonlinear DA methods is historically important because the original method for DA (i.e., the Kalman filter;

Kalman 1960) was based on strong linearity assumptions. This is limiting, as most models in hydrometeorology are highly nonlinear. This has led to an increasing focus on nonlinear methods in recent decades. If a system is truly linear, then Gaussian uncertainties are preserved through the forward model operator, $f(.)$, and observation operator, $h(.)$. In this case, the DA process is significantly simplified, and there is a known analytical solution to Eq. (4). For dynamic systems that are truly linear, and where uncertainty is truly Gaussian, the Kalman filter is an optimal solution to the filtering problem.

Alternatively, in the presence of a partially nonlinear model/problem, a range of techniques are available. These include Kalman filter-based methods (extended Kalman filter, unscented Kalman filter, ensemble Kalman filter) and more generalized solutions (variational, particle filter). Although a number of techniques are available, none may be considered optimal. Each nonlinear DA technique requires some limiting assumptions, which may complicate the choice of technique when applying DA. The main challenge in nonlinear DA is that there is no perfect inverse model. Therefore, the relationship between the states and the observation must be approximated prior to adjusting the states. Overcoming this challenge has been one of the primary focuses of DA scientists in recent decades.

1.3.3 Deterministic Versus Ensemble

Another distinction is between deterministic and ensemble DA methods. Deterministic methods perform updates on a single model realization, and the result is some metric (typically a mean field or maximum likelihood estimate) from the full posterior of either Eq. (3) or (4). Examples of deterministic DA methods are the Kalman filter, extended Kalman filter, and variational filters and smoothers. Deterministic methods are often more computationally efficient than ensemble methods, but have limitations when applied to complex models. Deterministic techniques require stricter assumptions about the forms of model error distributions and sometimes require model derivatives (e.g., for maximum likelihood estimation over nonlinear models). As a consequence, deterministic methods typically assume the state space at any particular time step is represented by a multivariate Gaussian distribution, which may be questionable in highly nonlinear models.

Alternatively, ensemble DA methods utilize multiple stochastic realizations of the model to represent uncertainties. Examples of ensemble DA methods include the ensemble Kalman filter (EnKF), the particle filter (PF), and the maximum likelihood ensemble filter (MLEF). Ensemble methods have the benefit of estimating the full PDF over model error, as it manifests in the state and/or prediction variables. In many cases, it is easier to apply ensemble methods, as opposed to deterministic methods, to nonlinear models. The primary drawback of ensemble methods is the increased computational demand of simulating the forward model and observational operators multiple times. Although these simulations are easily parallelizable, which reduces the computational demand, ensemble-based techniques still generally require increased simulation run time.

2 Error Characterization

2.1 Uncertainty Quantification

Uncertainty quantification is a key component of many DA systems. Since the intent of any DA system is to reduce uncertainty with respect to some pertinent model value, it is essential to understand the prior probabilistic characteristics of that uncertainty. This generally requires understanding the different sources of uncertainty in the full modeling and DA system. Generally, uncertainties will break down into three categories: boundary conditions, parameters, and model structure (i.e., process uncertainty). Boundary conditions include the initial state (x_0) and the model forcing data (u_t). Boundary conditions are required for solving any PDE system, and as mentioned previously, estimating improved initial conditions is one of the primary motivations for state estimation with DA. Uncertainty in model forcing data (e.g., precipitation, temperature, radiation) is typically estimated a priori. It is theoretically possible to use DA, or something like Eqs. (3) and/or (4), to condition PDFs over model forcing data, but this is not done regularly in hydrology.

Model parameters (e.g., hydraulic conductivity, streambed roughness) also inevitably contribute some uncertainty to the forecast system. DA can also be used to help reduce parameter uncertainty (see references above), but these methods are not yet common in operational hydrology forecasting.

Finally, process uncertainty is the uncertainty due to incomplete knowledge of the underlying processes within the model. This manifests as errors in the forward model operator $f(.)$ and observation operator $h(.)$. Due to the requirement to discretize processes both spatially and temporally, the model cannot perfectly simulate reality, and therefore the model itself will have uncertainty. There are methods for using DA to infer or condition model structural uncertainty distributions (e.g., Ghahramani and Roweis 1999) – some of which have been applied to river forecasting models (references above); however this is a relatively immature area of DA research and will not be discussed further in this essay.

2.1.1 Probabilistic Simulations

Applying any approximation of either Eq. (3) or (4) requires estimating all relevant uncertainties in the simulation system. Thus, DA inherently requires some type of probabilistic simulation to quantify that uncertainty. Depending on the complexity of this problem, that may be difficult or computationally expensive. Early DA methods targeted linear systems, with the assumption that errors were Gaussian. Probabilistic simulations for linear-Gaussian systems may be performed by propagating the expected value and covariance structures of the modeled state estimates forward in time. More generally, deterministic methods require partial derivatives of $f(.)$ for locating extremum of the posterior state PDF in nonlinear systems.

2.1.2 Ensemble Simulations

Ensemble simulations allow sampling of complex uncertainty distributions. This is beneficial when working with strongly nonlinear models and/or non-Gaussian

uncertainties. In an ensemble DA framework, the forecast PDF is represented by multiple stochastic realizations of a model, where each ensemble member is a sample from the forecast density. All of the input uncertainties (parameters, structure, boundary conditions) to the model are sampled, and each sample is propagated through the model, generating an ensemble forecast at each time step. This ensemble forecast is made up of N ensemble members, each with a weight that may be nonuniform, which is dependent on the DA technique applied. We can notate this situation as follows:

$$p(x_t) = \sum_{i=1}^{N} w_{t,i}\delta\left(x_t, x_{t,i}\right). \tag{5}$$

where $x_{t,\,i}$ is the state from ensemble member i of N, $\delta(.)$ is the Dirac delta function, and $w_{t,\,i}$ is the weight of ensemble member i. To simulate the ensemble of states, a model is run for N ensemble members, according to Eq. (6):

$$x_{t,i} = f\left(x_{t-1,i}, u_{t,i}, \theta_i^f\right) + \omega_{t,i} \tag{6}$$

In Eq. (6), $x_{t\,-\,1,\,i}$ is the state vector from the previous time step, $u_{t,\,i}$, is the current forcing sample, and $\omega_{t,\,i}$ is the current model error sample, each for ensemble member i. Each ensemble member represents a specific point within the state probability distribution. From this ensemble of model states, an ensemble of model-predicted observations may be generated:

$$y_{t,i} = h\left(x_{t,i}, \theta_i^h\right) + \epsilon_{t,i}. \tag{7}$$

In Eq. (7), $y_{t,\,i}$ and $\epsilon_{t,\,i}$ are the model prediction in observation space and error sample, respectively, for ensemble member i. The quantity $\epsilon_{t,\,i}$ captures deficiencies in the observation operator h and possibly also parameters θ_i^h (in some cases, these are considered separately with the parameters treated as random variables). Evaluation of Eqs. (6) and (7) for a large ensemble size allows for propagation from uncertainty distributions over parameters, forcing data, and model structure to uncertainty in model states and model-simulated observations. The uncertainty from the various aforementioned sources can be treated individually or lumped together as a "total uncertainty" term quantified by the additive errors ω_t and ϵ_t. The total uncertainty approach can be useful whenever quantifying the uncertainty in the individual sources is challenging. Often the additive errors are assumed to be zero mean Gaussian, although this assumption is seldom appropriate for hydrologic applications. Pathiraja et al. (2018a) presented a data-driven approach to estimate ω_t and ϵ_t from a total uncertainty perspective using only partial observations of the system and without relying on distributional assumptions on the errors. The approach is particularly suited to cases where model error characteristics are dependent on the system states and when the model-observed variables are of principal interest. It works by first generating a sample of additive errors on the latent states

and observed variables using a sequential optimization approach. The probability density of these errors is then estimated via nonparametric kernel conditional density estimation, thereby allowing for the characterization of complex error densities.

3 Data Assimilation Methods

3.1 Linear Data Assimilation

3.1.1 Kalman Filter

The Kalman filter (Kalman 1960) was the first true DA technique. Although the Kalman filter is rarely applied in hydrometeorology, due to its specific applicability to linear filtering problems, it is the basis of many generalized filters, making it a useful starting point for understanding many DA techniques. Kalman's solution to the filtering problem assumes the Gaussian distribution of errors, which greatly simplifies the state-updating process. Since the model is linear, and the form of uncertainty is known, the inversion of the model to estimate the optimal state value is analytical. The Kalman filter is applicable to models with linear state transition functions, of the form:

$$x_t^- = Ax_{t-1}^+ + Bu_t + \omega_t \tag{8}$$

where the model error is drawn from a Gaussian (normal) distribution of known covariance, Σ_m:

$$\omega_t \sim \mathcal{N}(0, \Sigma_m). \tag{9}$$

Within this linear model, A is a state transition matrix and B is an input transition matrix. Equation (8) allows direct propagation of the mean field of the state uncertainty distribution. When uncertainty in the initial states is also Gaussian, such that covariance of x_{t-1}^+ is notated P_{t-1}^+, then we can also directly propagate the state uncertainty variance:

$$P_t^- = AP_{t-1}^+A^T + \Sigma_o \tag{10}$$

In addition, the Kalman filter requires a linear observational operator:

$$y_t = Hx_t^- + \varepsilon_t \tag{11}$$

$$\varepsilon_t \sim \mathcal{N}(0, \Sigma_o) \tag{12}$$

In Eqs. (11) and (12), y_t is the observation; H is the observational operator, which is a function only of the modeled states; and ε_t is Gaussian observation error with covariance Σ_o. The Kalman filter only accounts for uncertainties due to model error and observation error. Based on these approximations, the states are linearly correlated with the observations and themselves have normally distributed uncertainty.

If all of these conditions are met, then we can solve the filtering problem (i.e., Eq. (4)) exactly:

$$x_t^+ = x_t^- + K_t\left(y_t - Hx_t^-\right) \tag{13}$$

$$K_t = P_t H^T \left(HP_t H^T + \Sigma_m\right)^{-1} \tag{14}$$

K_t is called the *Kalman gain* and x_t^+ are the updated model states. The updated state covariance is:

$$P_t^+ = (I - K_t H)P_{t-1}^+. \tag{15}$$

where I is the identity matrix.

3.2 Partially and Fully Nonlinear Deterministic Data Assimilation

3.2.1 Kalman Filter Extensions

Extended Kalman Filter

The extended Kalman filter (EKF) is a method developed in an effort to apply the Kalman filter to nonlinear dynamical systems models. Within the EKF, updates are performed on linearized approximations of the nonlinear model and work with nonadditive errors. Thus we will generalize Eqs. (1) and (2):

$$x_t = f\left(x_{t-1}, u_t, \theta^f, \omega_t\right) \tag{16}$$

$$y_t = h\left(x_t, \theta^h, \varepsilon_t\right) \tag{17}$$

As in the Kalman filter, the EKF estimates the prior states by progressing the model forward deterministically:

$$\hat{x}_t^- = f\left(\hat{x}_{t-1}^+, u_t, \theta^f\right) \tag{18}$$

where $\hat{x}_{t-1}^+$ and $\hat{x}_t^-$ are the updated and forecast states at time $t-1$ and t, respectively. The fact that the model is nonlinear makes estimating the Kalman gain significantly more difficult. Application to this system requires linearization of the model, allowing the Kalman update equation to effectively estimate the gradient of the state-observation relationship. To update the states, four partial derivatives are required. These partial derivatives will be taken from each model with respect to the states, as shown in Eqs. (19) and (20), and with respect to the model and observational errors, as shown in Eqs. (21) and (22):

$$A = \frac{df(.)}{dx} \tag{19}$$

$$H = \frac{dh(.)}{dx} \tag{20}$$

$$W = \frac{df(.)}{d\omega} \tag{21}$$

$$V = \frac{dh(.)}{d\epsilon} \tag{22}$$

The partial derivatives in Eqs. (19), (20), (21), and (22) linearize the model, which allows for application of the linear updating scheme of the Kalman filter. The error covariance is estimated as:

$$P_t = AP_{t-1}A^T + W\Sigma_m W^T, \tag{23}$$

where Σ_m is the state error covariance and based on that the Kalman gain may be estimated similarly to the standard Kalman filter, with a correction to the observation variance:

$$K_t = P_t H^T \left(HP_t H^T + V\Sigma_o V^T\right)^{-1} \tag{24}$$

where Σ_o is the observation error covariance. With this Kalman gain, the states may be updated in a way analogous to Eq. (13):

$$\hat{x}_t^+ = \hat{x}_t^- + K_t\left(y_t - h\left(\hat{x}_t^-, \theta^h, \varepsilon_t\right)\right) \tag{25}$$

The difference between the EKF update and the standard Kalman filter update is that the innovation, $y_t - h\left(\hat{x}_t, \theta^h, \varepsilon_t\right)$, is calculated from the nonlinear observational operator.

Unscented Kalman Filter

The unscented Kalman filter (UKF) is similar to the EKF, but linearizes around a set of state samples, instead of only one state estimate. The UKF can be thought of as a hybrid between deterministic and ensemble DA techniques. Since the method uses a sampling procedure to propagate uncertainty forward, error characterization is similar to ensemble techniques, but it retains a strictly Gaussian assumption by only updating the state expected value. This means that the posterior is a single deterministic value, representing the mean of the distribution, with a corresponding state error covariance. Due to the deterministic representation of the posterior, it cannot be considered a purely ensemble-based technique.

To apply the sampling strategy, one will estimate multiple sigma points with the model, allowing for calculation of the error covariance from a sample. Each of these sigma points will be used for initialization of the model, as is shown in Eq. (26):

$$\hat{x}_{t,i}^{-} = f\left(\hat{x}_{t-1,i}^{+}, u_t, \theta^f\right) \tag{26}$$

In Eq. (26), $\hat{x}_{t-1,i}$ is the ith sigma point estimate of the initial states, which is described in Eqs. (27) and (28):

$$\hat{x}_{t-1,1}^{-} = \hat{x}_{t-1}^{-} \tag{27}$$

$$\hat{x}_{t-1,i}^{-} = \begin{cases} \hat{x}_{t-1}^{-} + \sqrt{n+\lambda}\sqrt{P_{t-1}^{-}} & \text{if} \quad 1 < i \le n \\ \hat{x}_{t-1}^{-} - \sqrt{n+\lambda}\sqrt{P_{t-1}^{-}} & \text{if} \quad n < i \end{cases} \tag{28}$$

Sigma points are generated to capture the mean and covariance of the state estimates. In Eq. (28), n is the length of the state vector, $\sqrt{P_{t-1}^{-}}$ is the ith column of the Cholesky decomposition of P_{t-1}^{-}, and λ is a scaling factor. After initialization of the model with each sigma point, the sigma point for the prior model states at the current time step is available. At this point, the state expected value is estimated according to Eq. (29), the expected value of the observation forecast is estimated in Eq. (30), and the covariances are calculated according to Eqs. (31) and (32):

$$\hat{x}_t^{-} = \frac{1}{n+\lambda}\left(\lambda\hat{x}_{t,1}^{-} + \frac{1}{2}\sum_{i=2}^{2n+1}\hat{x}_{t,i}^{-}\right) \tag{29}$$

$$\hat{y}_t^{-} = \frac{1}{n+\lambda}\left(\lambda h\left(\hat{x}_{t,1}^{-}, \theta^h\right) + \frac{1}{2}\sum_{i=2}^{2n+1} h\left(\hat{x}_{t,i}^{-}, \theta^h\right)\right) \tag{30}$$

$$C_{XX} = \frac{1}{n+\lambda}\left(\lambda\left(\hat{x}_{t,1}^{-} - \hat{x}_t^{-}\right)\left(\hat{x}_{t,1}^{-} - \hat{x}_t^{-}\right)^T + \frac{1}{2}\sum_{i=2}^{2n+1}\left(\hat{x}_{t,i}^{-} - \hat{x}_t^{-}\right)\left(\hat{x}_{t,i}^{-} - \hat{x}_t^{-}\right)^T\right) \tag{31}$$

$$C_{XY} = \frac{1}{n+\lambda}\left(\lambda\left(\hat{x}_{t,1}^{-} - \hat{x}_t^{-}\right)\left(h\left(\hat{x}_{t,1}^{-}, \theta^h\right) - \hat{y}_t^{-}\right)^T + \frac{1}{2}\sum_{i=2}^{2n+1}\left(h\left(\hat{x}_{t,i}^{-}, \theta^h\right) - \hat{x}_t^{-}\right)\left(h\left(\hat{x}_{t,i}^{-}, \theta^h\right) - \hat{y}_t^{-}\right)^T\right) \tag{32}$$

From the above equations, the covariance of the states ($C_{XX} \approx P_t^{-} H^T$) and the covariance between the states and observations ($C_{XY} \approx H P_t^{-} H^T$) are estimated, allowing approximation of the optimal linear update. Following the standard Kalman filter, the Kalman gain is estimated from the covariances, as shown in Eq. (33), and the updated state vector is estimated from Eq. (34):

$$K_t = C_{XY}\left(C_{YY} + \Sigma_o\right)^{-1} \tag{33}$$

$$\hat{x}_t = \hat{x}_t^{-} + K_t\left(y_t - \hat{y}_t^{-}\right) \tag{34}$$

Similar to the EKF, the UKF estimates the proper linear update for the model states, allowing for approximation of the posterior state value of the nonlinear model.

3.2.2 Variational Data Assimilation

The premise of variational DA surrounds the idea of a cost function. A cost function represents errors in the system, which we seek to minimize. Rather than requiring the linearization of nonlinear models, as is performed in the Kalman filter extensions, variational methods rely on optimization tools to find the optimal state values with respect to a predefined cost function. The general form of the cost function, for Gaussian error structures, is shown in Eq. (35):

$$C = \left(\hat{x}_t - \hat{x}_t^-\right) \sum_{m}^{-1} \left(\hat{x}_t - \hat{x}_t^-\right) + \left(y_t - h\left(\hat{x}_t, \theta^h\right)\right) \sum_{o}^{-1} \left(y_t - h\left(\hat{x}_t, \theta^h\right)_t^-\right) \tag{35}$$

In Eq. (35), C is the value of the cost function, and all other variables were defined in earlier sections. In this form, the cost function compares the state error and the forecast error, which may be minimized to find the optimal solution to the filtering problem. Since a solution to the cost function may not be derived analytically, inverse modeling must be performed.

One method for solving the cost function is through iterative optimization techniques. These methods will search the state space for the state values that optimize (minimize) the cost function. This optimal value is considered to be the expected value of the states and therefore the best estimate available for the true states. Although this strategy is effective, it requires multiple evaluations of the model itself, increasing the computational burden. Since this is a deterministic DA method, it is advantageous to avoid multiple model evaluations. In order to achieve this goal, the derivative of the cost function is required:

$$\nabla C = \sum_{m}^{-1} \left(\hat{x}_t - \hat{x}_t^-\right) + J(\hat{x}_t) \sum_{o}^{-1} \left(y_t - h\left(\hat{x}_t, \theta^h\right)\right) \tag{36}$$

$J(\hat{x}_t)$ is the Jacobian of the model, also referred to as the adjoint model. This requires finding the partial derivatives of the model with respect to each state. Once the adjoint model is available, Eq. (36) may be used to find the minimum of the cost function by finding the $\hat{x}_t$ vector that satisfies $\nabla C = 0$. Therefore, the primary challenge is developing the adjoint model. This is a separate topic of study, and the reader is referred to Errico (1997). There are also software tools for developing adjoint models, including the Tangent linear and Adjoint Model Compiler (TAMC) (Giering 1997).

Four-dimensional variational DA (4D-Var) is a generalization of the variational filter from Eq. (35) where the time dimension of the observations is taken into account. This creates a smoothing methodology to account for more observations in the cost function. By examining multiple observations simultaneously, more information is available to reduce the state uncertainty, that is, information from

observations can be projected backward in time. Through this reduction in uncertainty, more accurate and precise estimates of the model states are expected. The general form of the 4D-Var cost function is:

$$C = \left(\hat{x}_0 - \hat{x}_0^-\right) B^{-1} \left(\hat{x}_0 - \hat{x}_0^-\right) + \sum_{t=1}^{T} \left(y_t - h\left(\hat{x}_t, \theta^h\right)\right) \sum_{o}^{-1} \left(y_t - h\left(\hat{x}_t, \theta^h\right)\right) \quad (37)$$

This cost function is applied to all observations over some time period of length of T. Due to the increased information available to the technique, initial state estimates generally become more accurate, and therefore 4D-Var is often preferred to 3D-Var.

3.3 Ensemble Data Assimilation

3.3.1 Ensemble Filters

Ensemble Kalman Filter

Application of the ensemble Kalman filter (EnKF) has become highly popular within the hydrometeorology forecast community. This popularity is due to several factors including simplicity of application, efficiency of the method, and the explicit treatment of complex and interacting uncertainties in the form of an ensemble. The EnKF is relatively simple to apply, compared with other nonlinear DA techniques, resulting from the use of an ensemble to quantify the covariances required for the Kalman update equation (Evensen 2003). This removes the need to take model derivatives, which is very challenging due to the complexity of hydrologic models. With respect to model efficiency, the assumption of Gaussian error structure has been shown to be reasonable in some applications, which leads to efficient updates of model states. Finally, the ensemble nature of the EnKF explicitly quantifies the uncertainty with the ensemble, where each ensemble member is equally weighted.

Application of the EnKF begins with an ensemble simulation, as described in Eqs. (6) and (7). After performing these simulations, the covariances are estimated directly from the ensembles:

$$E\left[\hat{x}_t^-\right] = \frac{1}{N} \sum_{i=1}^{N} \hat{x}_{t,i}^- \quad (38)$$

$$E[\hat{y}_t] = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_{t,i} \quad (39)$$

$$C_{XY} = E\left[\left(\hat{x}_t^- - E\left[\hat{x}_t^-\right]\right)\left(\hat{y}_t - E[\hat{y}_t]\right)^T\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left(\left(\hat{x}_{t,i}^- - E\left[\hat{x}_t^-\right]\right)\left(\hat{y}_{t,i} - E[\hat{y}_t]\right)^T\right) \quad (40)$$

$$C_{YY} = E\left[\left(\hat{y}_t - E[\hat{y}_t]\right)\left(\hat{y}_t - E[\hat{y}_t]\right)^T\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\hat{y}_t - E[\hat{y}_t]\right)\left(\hat{y}_t - E[\hat{y}_t]\right)\right) \quad (41)$$

With these covariances, the Kalman gain may be estimated according to:

$$K_t = C_{XY}\left(C_{YY} + \Sigma_o\right)^{-1} \quad (42)$$

This formulation of the Kalman gain follows the original Kalman filter, except that all model covariances are approximated with the ensemble. By applying this approximation, there is no need to apply linearization of the model, greatly simplifying the update process. Once the Kalman gain is available, each ensemble member is updated as:

$$\hat{x}_{t,i} = \hat{x}_{t,i}^- + K\left(y_{t,i} - \hat{y}_{t,i}\right) \quad (43)$$

In Eq. (43), $y_{t,i}$ is the ith sample of the observation, which is estimated as follows:

$$y_{t,i} = y_t + \varepsilon_{t,i} \quad \varepsilon_{t,i} \sim N(0, \Sigma_o) \quad (44)$$

The additional error sampling in Eq. (44) is required to account for uncertainty in the observations.

Ensemble Square Root Filter

The ensemble square root filter (EnSRF) was developed to remove the need to perturb the observations in the updates of the EnKF (Whitaker and Hamill 2002). By removing the need to perturb the observation, the necessary ensemble size is reduced, as no sampling of the observation uncertainty is performed. This is achieved by formulating the observation error into the Kalman gain. When explicitly accounting for the observation uncertainty into Eq. (42), the Kalman gain formulation becomes:

$$K_t = C_{XY}\left[\sqrt{C_{YY} + \sum\nolimits_o}^{-1}\right]^T\left[\sqrt{C_{YY} + \sum\nolimits_o} + \sqrt{\sum\nolimits_o}\right] \quad (45)$$

With this formulation of the Kalman gain, each ensemble member may be updated following Eq. (43).

3.3.2 Particle Filters

PFs were developed to overcome the challenges associated with the assumptions required to apply the Kalman-based filters. Although Kalman-based filters have been

shown to be effective in many applications, the imposition of a Gaussian error structure can become problematic in some hydrometeorological applications. This scenario motivates the use of an increasingly generalized filter, which can effectively manage skewed, or even multimodal, distributions.

PFs are rooted more firmly in Bayes' theorem than any of the methods we have discussed so far.

According to Eq. (1), the model is known to be Markovian, and therefore the Chapman-Kolmogorov equation may be used to expand the prior probability:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \tag{46}$$

This makes the prior distribution the integration of the transition probability ($p(x_t|x_{t-1})$) and the posterior at the previous time step. Beyond the initial time step, $p(x_{t-1}|y_{1:t-1})$ will be available, and the transition probability may be approximated through the sequential Monte Carlo algorithms (described in following sections). The implied proportionality constant in Eq. (4) comes from the observation probability, which may be expanded as:

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t \tag{47}$$

Therefore, one may solve the observation probability through the integration of the numerator of Eq. (4). This leads to Eq. (48), where only the likelihood, transition probability, and posterior at the previous time step are required to solve sequential Bayes' theorem:

$$p(x_t|y_{1:t}) = p(x_t|y_t, y_{1:t-1}) = \frac{p(y_t|x_t)\int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}}{\int p(y_t|x_t)\left[\int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}\right]dx_t} \tag{48}$$

Sequential Importance Sampling

In general, it is often not possible to derive an analytical expression for the Bayesian posterior from Eq. (48), but it is possible to utilize importance sampling to estimate the probabilities sequentially. This is referred to as sequential importance sampling (SIS), which is the most general solution available for the filtering problem (Gordon et al. 1993). SIS relies on a weighted sample of "particles" to estimate the posterior distribution. Similar to Eq. (5), the posterior sample may be represented by Eq. (49):

$$p(x_t|y_t) \approx \sum_{i=1}^{N} w_{t,i}\delta\left(x_t - \hat{x}_{t,i}\right) \tag{49}$$

Sampling directly from the posterior is often not possible, which necessitates importance sampling. This is performed by sampling from a known distribution and

weighting based on our knowledge of the system. Posterior importance weights may be estimated according to Eq. (50), where $q(x_{t,i}|\, x_{t-1,i},\, y_t)$ is the importance density:

$$w_{t,i} \propto \frac{p\left(y_t|\,\hat{x}_{t,i}\right)p\left(\hat{x}_{t,i}|\,\hat{x}_{t-1,i}\right)}{q\left(\hat{x}_{t,i}|\,\hat{x}_{t-1,i},y_t\right)} \tag{50}$$

The most common and convenient method for developing the importance density is through Eq. (51), where the importance density is set equal to the transition probability:

$$q\left(\hat{x}_{t,i}|\,\hat{x}_{t-1,i},y_t\right) = p\left(\hat{x}_{t,i}|\,\hat{x}_{t-1,i}\right) \tag{51}$$

Through this choice of importance density, the posterior weights may be estimated as the normalized product of the prior weights and the likelihood, as is shown in Eq. (52).

$$w_{t,i} = \frac{p\left(y_t|\,\hat{x}_{t,i}\right)w_{t-1,i}}{\sum_{i=1}^{N} p\left(y_t|\,\hat{x}_{t,i}\right)w_{t-1,i}} \tag{52}$$

This provides a weighted sample at each model time step, representing the posterior distribution of the model states.

Sampling Importance Resampling

Although SIS is the most generalized solution available to solving sequential Bayes' law, it is subject to failures over long simulations. Over a large number of time steps, it is common for many particles to drift from the observations, leading to weight degeneracy. Weight degeneracy refers to the scenario in which the weight of nearly all particles approach zero, with only a small portion of the sample having significant weight (Arulampalam et al. 2002). During this occurrence, the filter will fail, as it is unable to represent the posterior distribution. Weight degeneracy may be avoided with increasingly large sample size to represent all possible model trajectories, but this becomes limiting as the computational demand increases exponentially with dimensionality. In order to overcome this issue, the novel approach of resampling has been used. This is referred to as sampling importance resampling (SIR), where the sample of particles is resampled, based on the weights of the particles.

When applying SIR, it is common to estimate the effective sample size after each assimilation time step, to determine if resampling is required. The effective sample size may be estimated from Eq. (53) and compared to some predefined threshold:

$$N_{\text{eff}} = \sum_{i=1}^{N} \frac{1}{w_{t,i}^2} \tag{53}$$

where N_{eff} is the effective sample size. If this value drops below a specified value, it may be assumed that weight degeneracy is occurring, or about to occur, and resampling is required to avoid failure in the filter. At this stage, a representative posterior distribution may be developed by replicating higher probability particles and discarding low probability particles, which is a method referred to as resampling. The challenge when developing a resampling scheme is to create a representative sample of the posterior from current posterior sample. A common method is multinomial resampling.

Multinomial resampling is considered the simplest resampling technique for PFs (Douc and Cappe 2005). The first step in this resampling scheme is the development of the empirical cumulative density from the weights, as described in Eq. (54):

$$F_w(I) = \sum_{i=1}^{I} w_{t,i} \tag{54}$$

where I is the specified index of the current sample, which may be any integer on the range of [0,1]. With this cumulative density, it is possible to sample from the density with the uniformly sampled random variables (U), as described in Eq. (55):

$$I = F_w^{-1}(U) \quad U \sim U(0,1) \tag{55}$$

where $F_w^{-1}()$ is the inverse of the cumulative density. By putting a uniformly distributed random variable into the inverse cumulative density of the weights, it is possible to extract a corresponding index that should be sampled. This index may then be sampled according to Eq. (56), to develop the corresponding resampled value for ensemble member i ($x_{t,i}^{\text{res}}$):

$$x_{t,i}^{\text{res}} = x_{t,I} \tag{56}$$

A resampled state vector for each of the N particles will be sampled, at which point all weights are set uniformly, shown in Eq. (57), as the density of the particles represents the posterior:

$$w_{t,i} = \frac{1}{N} \tag{57}$$

According to Eq. (1), given that state variables are influenced by the forcing data uncertainty and system noise, the parameters are more susceptible to sample impoverishment as they are not dynamic quantities. To circumvent this issue, Moradkhani et al. (2005b) proposed a method to avoid sample impoverishment by perturbing the resampled parameters:

$$\theta_{t+1,i}^{-} = \theta_{t,i}^{+} + \gamma_{t,i} \qquad \gamma_{t,i} \sim N\left[0, s\text{Var}\left(\theta_{t,i}^{-}\right)\right] \tag{58}$$

where $\theta^{-}_{t+1,i}$ is the parameter at time step $t+1$, $\mathrm{Var}\left(\theta^{-}_{t,i}\right)$ is the variance of the prior parameters, and s is the variance multiplier, which should be tuned.

Despite the successful application of PF-SIR in many hydrologic practices, the convergence of parameters is dependent on the choice of tuning parameter s in the perturbation process. Moradkhani et al. (2012) developed a method to reduce the potential of sample impoverishment in PF-SIR by combining the strength of PF and Markov chain Monte Carlo (MCMC).

Particle Filter-Markov Chain Monte Carlo (PF-MCMC)

The PF-MCMC uses the PF-SIR algorithm to resample the state variables and parameters. Then, a proposal distribution is created to generate parameters $\theta^{p}_{t,i}$ allowing for larger move steps:

$$\theta^{p}_{t,i} = \theta^{+}_{t,i} + \gamma_{t,i} \qquad \gamma_{t,i} \sim N\left[0, s\mathrm{Var}\left(\theta^{-}_{t,i}\right)\right] \tag{59}$$

where s is the parameter variance tuning factor. To accept or reject the $\theta^{i,p}_{t}$ parameters, a metropolis acceptance ratio α is calculated:

$$\alpha = \min\left(1, \frac{p\left(x^{p}_{t,i}, \theta^{p}_{t,i} | y_{1:t}\right)}{p\left(x^{+}_{t,i}, \theta^{+}_{t,i} | y_{1:t}\right)}\right) \tag{60}$$

where $p\left(x^{p}_{t,i}, \theta^{p}_{t,i} | y_{1:t}\right)$ is the proposed joint probability distribution:

$$p\left(x^{p}_{t,i}, \theta^{p}_{t,i} | y_{1:t}\right) \propto p\left(y_{1:t} | x^{p}_{t,i}, \theta^{p}_{t,i}\right).p\left(x^{p}_{t,i} | \theta^{p}_{t,i}, y_{1:t-1}\right).p\left(\theta^{p}_{t,i} | y_{1:t-1}\right) \tag{61}$$

$$x^{p}_{t,i} = f\left(x^{+}_{t-1,i}, u_{t,i}, \theta^{p}_{t,i}\right) \tag{62}$$

where $x^{p}_{t,i}$ is a sample from the proposal state distribution at time step t.

Since the optimal tuning factor s is unknown in a sequential framework, it is beneficial to treat the s as a time-varying parameter and estimate it automatically. Moradkhani et al. (2012) modified the variable variance multiplier (VVM) method proposed by Leisenring and Moradkhani (2012) to automatically obtain the most fitting tuning factor s in Eq. 59.

Evolutionary PF-MCMC (EPFM)

The EPFM was proposed by Abbaszadeh et al. (2018) to characterize a more accurate and reliable posterior distribution for state variables in data assimilation applications. What distinguishes the EFPM approach from the PF-MCMC is the utilization of hybrid genetic algorithm (GA) and MCMC (GA-MCMC) technique in the importance sampling step of the PF-MCMC model. In fact, the GA-MCMC expands the search space by implementing the crossover and mutation steps in the

GA, and subsequently the search space is refined via the MCMC technique resulting in more desirable prior distribution. This approach significantly minimizes the particle degeneracy and sample impoverishment problems that have been the main concerns in using the particle filters. The main structures of EPFM approach are summarized below:

1. Particles are selected from the initial ensemble pool for the crossover operation. To do this, one can use roulette wheel selection method, and a fitness value for each ensemble member is assigned. The value of weights, as an appropriate indication of ensemble member quality, can be directly used as the fitness value.
2. The arithmetic crossover is adopted for the crossover operation (Park et al. 2009; Yin and Zhu 2015). For this, a pair of new particles (offspring) is generated by a linear combination of a pair of selected particles in step 1:

$$x_{t-1}^{i'} = \xi . x_t^i + (1 - \xi) . x_{t-1}^j \tag{63}$$

$$x_{t-1}^{j'} = (1 - \xi) . x_{t-1}^i + \xi . x_{t-1}^j \tag{64}$$

where x_{t-1}^i and x_{t-1}^j are the parent particles, $x_{t-1}^{i'}$ and $x_{t-1}^{j'}$ are the pair of new offspring particle, and ξ is a uniform random value in the range of [0, 1].
3. To further promote diversity of the particles, a mutation strategy is designed. It is realized by Eq. 65 that x_{t-1}^k and $x_{t-1}^{k'}$ are the particles before and after mutation process, respectively:

$$x_{t-1}^{k'} = x_{t-1}^k + \eta \qquad x_{t-1}^k \in \left\{ x_{t-1}^{i'}, x_{t-1}^{j'} \right\} \qquad \eta \sim N\left(0, \psi \mathrm{Var}\left(x_{t-1}^{k-}\right)\right) \tag{65}$$

where η represents a random sample from a Gaussian distribution with mean zero and variance $\psi \mathrm{Var}\left(x_{t-1}^{k-}\right)$, where $\mathrm{Var}\left(x_{t-1}^{k-}\right)$ is the variance of the prior states at the time $t - 1$ and ψ is a small tuning parameter.
4. The MCMC algorithm is used to accept or reject the new particles generated by GA operators. This step is similar to the one used in the PF-MCMC model.

4 Applications

4.1 Variational

Variational methods have become popular for atmospheric DA, but are less popular in land surface applications. Although applicable to both, the atmospheric community has more readily developed the adjoint models necessary for variational methods, whereas the land surface community has generally relied on ensemble methods. Within the atmospheric DA community, several examples of variational DA applications are available (Barker et al. 2004; Dee et al. 2011; Hou et al. 2013; Županski and Mesinger 1995). Although uncommon, there are a few examples of

variational DA within the hydrologic community (Reichle et al. 2001; Seo et al. 2003). A few recent examples of land surface DA with variational methods include Hoppe et al. (2014), which assimilated soil moisture and temperature measurements into the Community Land Model (CLM); Meng et al. (2009), which assimilated land surface temperature into the CLM; and Lee et al. (2012), which assimilated streamflow into the Sacramento Soil Moisture Accounting model.

4.2 Kalman-Based Filters

Kalman-based filters are the most commonly used DA methods in hydrometeorology. This is a result of the combination of ease of application and effectiveness of the technique. Of particular focus have been ensemble applications of the Kalman filter. In the atmospheric and land surface DA communities alike, the EnKF has been widely applied. Although variational methods would be advantageous from an efficiency perspective, the development of an adjoint model can be challenging for highly nonlinear models, making ensemble methods attractive. Alternatively, PFs are highly robust estimators of the posterior distribution, but are subject to failure in small ensemble sizes. This makes the EnKF a useful tool in large dimensional problems. For atmospheric DA, applications typically involve observations of wind speed, wind direction, temperature, and humidity from radiosondes and satellites (Annan et al. 2005; Houtekamer and Mitchell 1998; Lorenc 2003). For land surface and hydrologic DA, the observations and applications are much more varied. Applications include soil moisture (e.g., Kumar et al. 2014; Reichle et al. 2002; De Lannoy et al. 2007; De Rosnay et al. 2013), passive microwave brightness temperature (Crow and Wood 2003; DeChant and Moradkhani 2011b; Durand and Margulis 2008), snow cover fraction (Andreadis and Lettenmaier 2006; Slater and Clark 2005), snow water equivalent (De Lannoy et al. 2012; Leisenring and Moradkhani 2011; Liu et al. 2012), streamflow (Clark et al. 2008; Moradkhani et al. 2005a; Noh et al. 2011; Samuel et al. 2014), and consideration of non-stationarity in dynamic catchments (Pathiraja et al. 2016a, b).

4.3 Particle Filters

The DA community has been slow to adopt the PF methods, primarily due to the understanding that PFs are overly demanding computationally (Snyder et al. 2008). The “Curse of Dimensionality” has been termed to designate the exponential scaling of necessary sample size for estimating the posterior with increasing degrees of freedom in the system (Bengtsson et al. 2008). Although this criticism has shown that certain PFs are subject to failure in large-scale systems, PFs are gaining popularity in many applications. Due to improvements in filter efficiency, and the identification of applicable problems, the PF has become a viable method for DA (Moradkhani et al. 2012). With the introduction of PF to hydrologic community, the PFs have gained a considerable attention in a variety of land surface applications

by assimilating variables including streamflow (e.g., Moradkhani et al. 2005b, 2012; Weerts and El Serafy 2006; DeChant and Moradkhani 2011a; Yan and Moradkhani 2016; Abbaszadeh et al. 2018), soil moisture (Montzka et al. 2011, 2013; Guingla et al. 2012; Yan et al. 2015, 2017), snow water equivalent (Leisenring and Moradkhani 2011; DeChant and Moradkhani 2011b, 2012), sediment load (Leisenring and Moradkhani 2012), flood inundation (Matgen et al. 2011; Plaza et al. 2012), and multi-modeling (Parrish et al. 2012; DeChant and Moradkhani 2014a).

4.4 Parameter Inference and Model Structures

In addition to the methods described above, there are several DA methods in hydrology and from other branches of scientific literature that focus on more holistic treatments of the problem of reducing uncertainties in dynamical systems models by probabilistically conditioning model states on observations. Perhaps most notably, there have been several applications of various methods that simultaneously estimate model parameters and model states (e.g., Moradkhani et al. 2005a, b, 2012; DeChant and Moradkhani 2012; Smith et al. 2013; Ruiz and Pulido 2015; Gharamti et al. 2017; Abbaszadeh et al. 2018).

The sequential estimation of temporally varying model parameters through DA can also be used to improve hydrologic forecasting in systems with changing catchment properties (such as land use or land cover change) (Pathiraja et al. 2016a, b). Such a time-varying parameter framework can be useful whenever the catchment system is undergoing change in real time, that may be unknown to the modeler. Model parameters are sequentially updated in response to signals of change in observations, such that the model is improved as soon as an information about a change becomes available. This can be done through a joint state-parameter estimation DA framework with a careful choice on the parameter evolution model, i.e., the method for generating prior distributions of the parameters at each time (Pathiraja et al. 2016b). Additionally, the choice of the model structure itself is critical in ensuring that such a time-varying parameter framework can be useful under changing conditions. Specifically, the model structure must be sufficiently flexible so that it can represent the range of possible future changes to catchment conditions (Pathiraja et al. 2018b). In other words, the entire feasible parameter space and forcings must produce model states and outputs that capture all possible future outcomes.

Additionally, state-updating DA has been used to help understand complex model error distributions and to update model structures (Bulygina and Gupta 2009, 2010, 2011; Wilkinson et al. 2011; Nearing and Gupta 2015; Nearing et al. 2013). It is often difficult to use observation data to directly infer structural errors in complex systems models, because we often do not have observations related to all of the interacting biogeophysical processes in a watershed or other hydrologic system. Careful applications of DA can be used to update the internal states of the model

given whatever partial observations are available, so long as the uncertainty due to the model structure is appropriately quantified (Pathiraja et al. 2018a).

5 Conclusion

DA techniques are valuable tools for estimating initial conditions for hydrometeorological forecasts. Due to the uncertainties in initial conditions (DeChant and Moradkhani 2011a, 2014a), it is necessary to quantify and reduce these uncertainties, and DA is widely seen as the forefront of the science in performing this task. With developments in DA science in the last two decades, several assimilation techniques are becoming standard tools for quantifying and reducing model uncertainty. These tools have seen wide ranging applications, particularly in simulating atmospheric and land surface processes.

Although DA is becoming a standard set of tools, the variety of techniques requires significant thought in determining the proper technique for a given application. If the underlying system is linear, or nearly linear, the Kalman filter will likely be chosen as it will be an optimal filter. Alternatively, in highly nonlinear problems, which are the norm in hydrometeorology, the choice in technique becomes much more difficult. Choosing between deterministic and ensemble techniques is a challenge. Although generalized techniques (i.e., PFs) are preferred from a theoretical perspective, they require the ability to execute the model enough times to sample from the posterior. Assuming that a large number of simulations are possible, it is likely that the PF will be preferred. If the model is highly computationally demanding, this may be infeasible, and therefore it may be impossible to fully represent the posterior distribution. In this scenario, the EnKF, and similar methods, may be used to retrieve the expected value with smaller ensemble size. Another option is variational DA, which may also retrieve the expected value. Variational methods bring the challenge of requiring an adjoint model, but are very competitive with the EnKF. Overall, each DA technique will have benefits and drawbacks, which often makes the choice of technique situation specific.

References

P. Abbaszadeh, H. Moradkhani, H. Yan, Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo method. Adv. Water Resour. **111**, 192–204 (2018). https://doi.org/10.1016/j.advwatres.2017.11.011

K.M. Andreadis, D.P. Lettenmaier, Assimilating remotely sensed snow observations into a macroscale hydrology model. Adv. Water Resour. **29**, 872–886 (2006)

J.D. Annan, J.C. Hargreaves, N.R. Edwards, R. Marsh, Parameter estimation in an intermediate complexity Earth system model using an ensemble Kalman filter. Ocean Model. **8**(1), 135–154 (2005)

M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–188 (2002)

D.M. Barker, W. Huang, Y.-R. Guo, A.J. Bourgeois, Q.N. Xiao, A three-dimensional variational data assimilation system for MM5: implementation and initial results. Mon. Weather Rev. **132**(4), 897–914 (2004)

T. Bengtsson, P. Bickel, B. Li, Curse of dimensionality revisited: the collapse of importance sampling in very large scale systems, in IMS Collections: Probability and Statistics: Essays in Honor of David A. Freedman,, vol. 2, ed. by D. Nolan, T. Speed (Institute of Mathematical Statistics, Beachwood), pp. 316–334 (2008)

N. Bulygina, H. Gupta, Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. Water Resour. Res. **45**(12), W00B13 (2009). https://doi.org/10.1029/2007WR006749

N. Bulygina, H. Gupta, How Bayesian data assimilation can be used to estimate the mathematical structure of a model. Stoch. Environ. Res. Risk Assess. **24**(6), 925 (2010). https://doi.org/10.1007/s00477-00010-00387-y

N. Bulygina, H. Gupta, Correcting the mathematical structure of a hydrological model via Bayesian data assimilation. Water Resour. Res. **47**(5), W05514 (2011). https://doi.org/10.1029/2010WR009614

M.P. Clark, D.E. Rupp, R.A. Woods, X. Zheng, R.P. Ibbitt, A.G. Slater, J. Schmidt, M.J. Uddstrom, Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. Adv. Water Resour. **31**, 1309 (2008)

W.T. Crow, E.F. Wood, The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97. Adv. Water Resour. **26**(2), 137–149 (2003)

G.J.M. De Lannoy, R.H. Reichle, P.R. Houser, V.R.N. Pauwels, N.E.C. Verhoest, Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour. Res. **43**, W09410 (2007). https://doi.org/10.1029/2006WR00544

G.J.M. De Lannoy, R.H. Reichle, K.R. Arsenault, P.R. Houser, S. Kumar, N.E.C. Verhoest, V. Pauwels, Multiscale assimilation of advanced microwave scanning radiometer–EOS snow water equivalent and moderate resolution imaging spectroradiometer snow cover fraction observations in northern Colorado. Water Resour. Res. **48**, W01522 (2012). https://doi.org/10.1029/2011WR010588

P. De Rosnay, M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, L. Isaksen, A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. Q. J. R. Meteorol. Soc. **139**(674), 1199–1213 (2013). https://doi.org/10.1002/qj.2023

C. DeChant, H. Moradkhani, Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. Hydrol. Earth Syst. Sci. **15**, 3399–3410 (2011a). https://doi.org/10.5194/hess-15-3399.

C. DeChant, H. Moradkhani, Radiance data assimilation for operational snow and streamflow forecasting. Adv. Water Resour. **34**(3), 351–364 (2011b)

C.M. DeChant, H. Moradkhani, Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resour. Res. **48**(4), W04518 (2012)

C.M. DeChant, H. Moradkhani, Toward a reliable prediction of seasonal forecast uncertainty: addressing model and initial condition uncertainty with ensemble data assimilation and sequential Bayesian combination. J. Hydrol. **519**, 2967–2977 (2014a). https://doi.org/10.1016/j.jhydrol.2014.05.045. Special issue on Ensemble Forecasting and data assimilation

C.M. DeChant, H. Moradkhani, Hydrologic prediction and uncertainty quantification, in *Handbook of Engineering Hydrology, Modeling, Climate Change and Variability* (CRC Press, Taylor & Francis Group, Boca Raton, 2014b), pp. 387–414

D.P. Dee et al., The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. **137**(656), 553–597 (2011)

R. Douc, O. Cappe, Comparison of resampling schemes for particle filtering, paper presented at image and signal processing and analysis, 2005. ISPA 2005, in *Proceedings of the 4th International Symposium on, 15–17 Sept 2005* (2005)

M. Durand, S.A. Margulis, Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization. J. Geophys. Res. **113**(D2), D02105 (2008)
R.M. Errico, What is an adjoint model? Bull. Am. Meteorol. Soc. **78**(11), 2577–2591 (1997)
G. Evensen, The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. **53**(4), 343–367 (2003)
Z. Ghahramani, S.T. Roweis, Learning nonlinear dynamical systems using an EM algorithm. Adv. Neural Inf. Process. Syst. **11**, 431–437 (1999)
M.E. Gharamti, J. Tjiputra, I. Bethke, A. Samuelsen, I. Skjelvan, M. Bentsen, L. Bertino, Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. Ocean Model. **112**, 65–89 (2017)
R. Giering, *Tangent Linear and Adjoint Model Compiler, Users Manual* (Center for Global Change Sciences, Department of Earth, Atmospheric, and Planetary Science. MIT, Cambridge, 1997)
N. Gordon, D. Salmond, A. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Proc. Inst. Elect. Eng. F. **140**(2), 107–113 (1993)
P. Guingla, D. Antonio, R. De Keyser, G. De Lannoy, L. Giustarini, P. Matgen, V. Pauwels, The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. Hydrol. Earth Syst. Sci. **16**(2), 375–390 (2012)
C.M. Hoppe, H. Elbern, J. Schwinger, A variational data assimilation system for soil–atmosphere flux estimates for the Community Land Model (CLM3. 5). Geosci. Model Dev. **7**(3), 1025–1036 (2014)
T. Hou, F. Kong, X. Chen, H. Lei, Impact of 3DVAR data assimilation on the prediction of heavy rainfall over Southern China. Adv. Meteorol. **2013**, 1 (2013)
P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique. Mon. Weather Rev. **126**(3), 796–811 (1998)
R.E. Kalman, A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. **82**(Series D), 35–45 (1960). https://doi.org/10.1115/1111.3662552
S. Kumar, C. Peters-Lidard, D. Mocko, R. Reichle, Y. Liu, K. Arsenault, Y. Xia, M. Ek, G. Riggs, B. Livneh, M Cosh, Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. J. Hydrometeorol. **15**, 2446–2469 (2014). https://doi.org/10.1175/JHM-D-13-0132.1
H. Lee, D.J. Seo, Y. Liu, V. Koren, P. McKee, R. Corby, Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal scale of adjustment. Hydrol. Earth Syst. Sci. **16**(7), 2233–2251 (2012)
M. Leisenring, H. Moradkhani, Snow water equivalent prediction using Bayesian data assimilation methods. Stoch. Environ. Res. Risk Assess. **25**(2), 253–270 (2011)
M. Leisenring, H. Moradkhani, Analyzing the uncertainty of suspended sediment load prediction using sequential Monte Carlo methods. J. Hydrol. **468–469**, 268–282 (2012). https://doi.org/10.1016/j.jhydrol.2012.08.049
Y. Liu, A.H. Weerts, M. Clark, H.J. Hendricks Franssen, S. Kumar, H. Moradkhani, D.J. Seo, D. Schwanenberg, P. Smith, A.I.J.M. van Dijk, N. van Velzen, M. He, H. Lee, S.J. Noh, O. Rakovec, P. Restrepo, Toward advancing data assimilation in operational hydrologic forecasting and water resources management: current status, challenges, and emerging opportunities. Hydrol. Earth Syst. Sci. **16**, 3863–3887 (2012)
A.C. Lorenc, The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. Q. J. R. Meteorol. Soc. **129**(595), 3183–3203 (2003)
P. Matgen, R. Hostache, G. Schumann, L. Pfister, L. Hoffmann, H.H.G. Savenije, Towards an automated SAR-based flood monitoring system, Lessons learned from two case studies. Phys. Chem. Earth. **36**(7–8), 241–252 (2011). https://doi.org/10.1016/j.pce.2010.12.009
C.L. Meng, Z.L. Li, X. Zhan, J.C. Shi, C. Y. Liu, Land surface temperature data assimilation and its impact on evapotranspiration estimates from the Common Land Model. Water Resour. Res. **45**, W02421 (2009). https://doi.org/10.1029/2008WR006971
C. Montzka, H. Moradkhani, L. Weihermüller, H.J. Hendricks Franssen, M. Canty, H. Vereecken, Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the

particle filter. J. Hydrol. **399**(3–4), 410–421 (2011). https://doi.org/10.1016/j.jhydrol.2011.01.020

C. Montzka, J. Grant, H. Moradkhani, H.J. Hendricks Franssen, L. Weihermüller, M. Drusch, H. Vereecken, Estimation of radiative transfer parameters from L-Band passive microwave brightness temperatures using data assimilation. Vadose Zone Hydrol. Special Issue of Remote Sensing. (2013). https://doi.org/10.2136/vzj2012.0040

H. Moradkhani, S. Sorooshian, H.V. Gupta, P.R. Houser, Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. **28**(2), 135–147 (2005a)

H. Moradkhani, K.L. Hsu, H. Gupta, S. Sorooshian, Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. **41**, W05012 (2005b)

H. Moradkhani, C.M. DeChant, S. Sorooshian, Evolution of ensemble data assimilation for uncertainty quantification using the Particle Filter-Markov Chain Monte Carlo method. Water Resour. Res. **48**, W12520 (2012). https://doi.org/10.1029/2012WR012144

G.S. Nearing, H.V. Gupta, The quantity and quality of information in hydrologic models. Water Resour. Res. **51**(1), 524–538 (2015)

G.S. Nearing, H.V. Gupta, W.T. Crow, Information loss in approximately bayesian estimation techniques: a comparison of generative and discriminative approaches to estimating agricultural productivity. J. Hydrol. **507**, 163–173 (2013)

S.J. Noh, Y. Tachikawa, M. Shiiba, S. Kim, Applying sequential Monte Carlo methods into a distributed hydrologic model: lagged particle filtering approach with regularization. Hydrol. Earth Syst. Sci. **15**(10), 3237 (2011)

S. Park, J.P. Hwang, E. Kim, H. Kang, A new evolutionary particle filter for the prevention of sample impoverishment. IEEE Trans. Signal Process. **13**(4), 801–809 (2009)

M. Parrish, H. Moradkhani, C.M. DeChant, Towards reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. Water Resour. Res. **48**, W03519 (2012). https://doi.org/10.1029/2011WR011116

S. Pathiraja, L. Marshall, A. Sharma, H. Moradkhani, Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. Adv. Water Resour. **94**, 103–119 (2016a). https://doi.org/10.1016/j.advwatres.2016.04.021

S. Pathiraja, L. Marshall, A. Sharma, H. Moradkhani, Hydrologic modeling in dynamic catchments: a data assimilation approach. Water Resour. Res. (2016b). https://doi.org/10.1002/2015WR017192

S. Pathiraja, D. Anghileri, P. Burlando, A. Sharma, L. Marshall, H. Moradkhani, Time varying parameter models for catchments with land use change: the importance of model structure. Hydrol. Earth Syst. Sci. Discuss. (2017). https://doi.org/10.5194/hess-2017-382

S. Pathiraja, H. Moradkhani, L. Marshall, A. Sharma, G. Geenens, Data driven model uncertainty estimation in data assimilation. Water Resour. Res. (2018a). https://doi.org/10.1002/2018WR022627

S. Pathiraja, D. Anghileri, P. Burlando, A. Sharma, L. Marshall, H. Moradkhani, Insights on the impact of systematic model errors on data assimilation performance in changing catchments. Adv. Water Resour. (2018b). https://doi.org/10.1016/j.advwatres.2017.12.006

D.A. Plaza, R. De Keyser, G.J.M. De Lannoy, L. Giustarini, P. Matgen, V.R.N. Pauwels, The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. Hydrol. Earth Syst. Sci. **16**(2), 375–390 (2012)

R.H. Reichle, D. Entekhabi, D.B. McLaughlin, Downscaling of radio brightness measurements for soil moisture estimation: a four-dimensional variational data assimilation approach. Water Resour. Res. **37**(9), 2353–2364 (2001)

R.H. Reichle, D.B. McLaughlin, D. Entekhabi, Hydrologic data assimilation with the ensemble Kalman filter. Mon. Weather Rev. **130**(1), 103–114 (2002)

J. Ruiz, M. Pulido, Parameter estimation using ensemble-based data assimilation in the presence of model error. Mon. Weather Rev. **143**(5), 1568–1582 (2015)

P. Salamon, L. Feyen, Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. J. Hydrol. **376**(3), 428–442 (2009)

J. Samuel, P. Coulibaly, G. Dumedah, H. Moradkhani, Assessing model state variation in hydrologic data assimilation. J. Hydrol. **513**, 127–141 (2014). https://doi.org/10.1016/j.jhydrol.2014.03.048

D.-J. Seo, V. Koren, N. Cajina, Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. J. Hydrometeorol. **4**(3), 627–641 (2003)

D.J. Seo, Y. Liu, H. Moradkhani, A. Weerts, Ensemble prediction and data assimilation for operational hydrology. J. Hydrol. **519**, 2661–2662 (2014). https://doi.org/10.1016/j.jhydrol.2014.11.035

A.G. Slater, M.P. Clark, Snow data assimilation via an ensemble Kalman filter. J. Hydrometeorol. **7**, 478 (2005)

P.J. Smith, G.D. Thornhill, S.L. Dance, A.S. Lawless, D.C. Mason, N.K. Nichols, Data assimilation for state and parameter estimation: application to morphodynamic modelling. Q. J. R. Meteorol. Soc. **139**(671), 314–327 (2013)

C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering. Mon. Weather Rev. **136**(12), 4629 (2008)

J.A. Vrugt, C.G.H. Diks, H.V. Gupta, W. Bouten, J.M. Verstraten, Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resour. Res. **41**(1), W01017 (2005). https://doi.org/10.1029/2004WR003059

A.H. Weerts, G.Y.H. El Serafy, Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour. Res. **42**, W09403 (2006). https://doi.org/10.1029/2005WR004093

J.S. Whitaker, T.M. Hamill, Ensemble data assimilation without perturbed observations. Monthly Weather Rev. **130**(7), 1913–1924 (2002). https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

R.D. Wilkinson, M. Vrettas, D. Cornford, J.E. Oakley, Quantifying simulator discrepancy in discrete-time dynamical simulators. J. Agric. Biol. Environ. Stat. **16**(4), 554–570 (2011)

H. Yan, H. Moradkhani, Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. Adv. Water Resour. **94**, 364–378 (2016). https://doi.org/10.1016/j.advwatres.2016.06.002

H. Yan, C.M. DeChant, H. Moradkhani, Improving soil moisture profile prediction with the Particle Filter-Markov Chain Monte Carlo method. IEEE Trans. Geosci. Remote Sens. (2015). https://doi.org/10.1109/TGRS.2015.2432067

H. Yan, H. Moradkhani, M. Zarekarizi, A probabilistic drought forecasting framework: a combined dynamical and statistical approach. J. Hydrol. **548**, 291–304 (2017). https://doi.org/10.1016/j.jhydrol.2017.03.004

S. Yin, X. Zhu, Intelligent particle filter and its application to fault detection of nonlinear systems. IEEE Trans. Ind. Electron. **62**(6), 3852–3861 (2015)

D.a. Županski, F. Mesinger, Four-dimensional variational assimilation of precipitation data. Mon. Weather Rev. **123**(4), 1112–1127 (1995)