



Using an ensemble smoother to evaluate parameter uncertainty of an integrated hydrological model of Yanqi basin



Ning Li ^{a,b,*}, Dennis McLaughlin ^c, Wolfgang Kinzelbach ^a, WenPeng Li ^d, XinGuang Dong ^e

^a Institute of Environmental Engineering, ETH, 8093 Zurich, Switzerland

^b China Institute of Geo-Environmental Monitoring, Beijing, China

^c Department of Civil and Environmental Engineering, MIT, Cambridge, MA 02139, USA

^d Institute of Hydrogeology and Engineering Geology Techniques, China Geological Survey, Hebei, China

^e Xinjiang Department of Water Resources, Xinjiang, China

ARTICLE INFO

Article history:

Received 15 July 2014

Received in revised form 3 June 2015

Accepted 20 July 2015

Available online 26 July 2015

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:

Uncertainty analysis

Ensemble smoother with multiple data assimilation

Unbiased square root filter

Integrated hydrological model

Latin hypercube sampling

Yanqi basin

SUMMARY

Model uncertainty needs to be quantified to provide objective assessments of the reliability of model predictions and of the risk associated with management decisions that rely on these predictions. This is particularly true in water resource studies that depend on model-based assessments of alternative management strategies. In recent decades, Bayesian data assimilation methods have been widely used in hydrology to assess uncertain model parameters and predictions. In this case study, a particular data assimilation algorithm, the Ensemble Smoother with Multiple Data Assimilation (ESMDA) (Emerick and Reynolds, 2012), is used to derive posterior samples of uncertain model parameters and forecasts for a distributed hydrological model of Yanqi basin, China. This model is constructed using MIKESHE/MIKE11 software, which provides for coupling between surface and subsurface processes (DHI, 2011a–d). The random samples in the posterior parameter ensemble are obtained by using measurements to update 50 prior parameter samples generated with a Latin Hypercube Sampling (LHS) procedure. The posterior forecast samples are obtained from model runs that use the corresponding posterior parameter samples. Two iterative sample update methods are considered: one based on an a perturbed observation Kalman filter update and one based on a square root Kalman filter update. These alternatives give nearly the same results and converge in only two iterations. The uncertain parameters considered include hydraulic conductivities, drainage and river leakage factors, van Genuchten soil property parameters, and dispersion coefficients. The results show that the uncertainty in many of the parameters is reduced during the smoother updating process, reflecting information obtained from the observations. Some of the parameters are insensitive and do not benefit from measurement information. The correlation coefficients among certain parameters increase in each iteration, although they generally stay below 0.50.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Hydrological models can provide detailed information that improves our understanding of complex natural systems. They can also be used to forecast the effects of different water resource management scenarios (eg. Cartwright et al., 2006; Li et al., 2010; Matial and Johnes, 2012). The reliability of a model forecast depends primarily on uncertainties in model structure, in model inputs (initial conditions and forcing terms), and in model parameters. In this article we describe two efficient methods for

quantifying the effects of uncertainty, illustrating concepts with a field study that involves complex interactions among surface water, ground water, and salinity. We consider both input and parameter uncertainties and account for information provided by imperfect measurements of some of the forecast variables. Our approach illustrates how data assimilation techniques can be used to combine prior information, model predictions, and measurements to provide an integrated picture of uncertainty.

Bayesian methods provide a convenient framework for both data assimilation and uncertainty analysis. In practical applications that involve spatially distributed nonlinear models with non-Gaussian uncertainties the Bayesian approach is usually only feasible when implemented in an ensemble, or Monte Carlo, form. Ensemble uncertainty analysis works with a set (or ensemble) of

* Corresponding author at: Institute of Environmental Engineering, ETH, 8093 Zurich, Switzerland. Tel.: +41 44 6333074.

E-mail address: ning.li@ifu.baug.ethz.ch (N. Li).

parameter samples that describe the likely range of the unknown parameter values. The algorithm provides interval estimates (or probabilistic ranges) that explicitly provide information on uncertainty. This is in contrast to methods that provide a single point (or deterministic) parameter estimate without an accompanying assessment of uncertainty. In the Bayesian application considered here the end goal is to produce informative posterior probability distributions rather than “best estimates” of model parameters and forecast variables. In this context, point values such as posterior means and modes help to characterize the posterior density but are not ends in themselves.

Ensemble algorithms start with user-generated prior parameter samples that represent the best available information about the parameter prior distribution before considering field observations of forecast variables. The prior samples are modified by the Bayesian algorithm to give posterior (or updated) parameter samples that incorporate the field observations. The posterior statistics computed by an ensemble algorithm can be viewed as approximations to the corresponding distributional properties of the exact Bayesian posterior probability distribution, which is generally not possible to derive in closed form.

In the past few decades, ensemble methods have been widely used in uncertainty assessment and data assimilation applications, reflecting dramatic advances in computational hardware and methods (Kuczera and Parent, 1998; Helton and Davis, 2003; Ramirez et al., 2008). Examples include Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) methods, which implement the Bayesian approach with minimal simplifications and assumptions (Vrugt et al., 2009; Jin et al., 2010). In spatially distributed hydrologic applications such as the one considered here the number of uncertain model parameters and states can become very large. MCMC and other exact ensemble methods can require very large sample sizes to give accurate results for large problems, making them computationally expensive.

The computational limitations of these methods have prompted the development of alternative ensemble approaches that reduce computational effort by making application-specific approximations and simplifications. It should be kept in mind that some ensemble updating procedures used to derive posterior distributions from sampled prior distributions may be suboptimal for nonlinear problems. Nevertheless, these approximate methods are typically much more computationally feasible than other sampling methods (such as MCMC) that make fewer approximations (but are still suboptimal due to finite sample sizes and slow convergence). This is especially true when the model of interest is computationally demanding.

Examples of efficient but approximate alternatives to MCMC include the ensemble Kalman filter and related ensemble smoothing algorithms (Evensen, 1992, 1994). These algorithms have been used for data assimilation in oceanic models (Evensen and van Leeuwen, 1996; Houtekamer and Mitchell, 1998), meteorological forecasting (Houtekamer and Mitchell, 2001), soil moisture investigations (Reichle et al., 2002), snow data analysis (Slater and Clark, 2004), and parameter updating in hydrological models (Moradkhani et al., 2005; Chen and Zhang, 2006).

The ensemble Kalman filter (EnKF) is a sequential algorithm that derives an ensemble of updated parameter values at the current time from a linear combination of the current measurement and an ensemble of model forecasts initialized at the previous update time. The weighting between the measurement and forecast ensemble depends on covariances between these quantities. In the EnKF the forecast ensemble implicitly defines a Bayesian prior distribution while the updated ensemble implicitly defines a Bayesian posterior distribution. The original form of the EnKF requires the current measurement to be perturbed with a random term that has the statistical properties assumed to apply to the

actual (but unknown) measurement error. Square Root Filters (SRF) were introduced, in part, to avoid the need for such perturbed observations (Whitaker and Hamill, 2002; Tippett et al., 2003; Evensen, 2004; Sakov and Oke, 2008). The SRF method updates the posterior mean and square root of posterior covariance of the parameters separately. Livings et al. (2008) provides a generic set of necessary and sufficient conditions for the SRF to yield an unbiased state estimate.

When the EnKF or SRF are used to assess the uncertainty of time-invariant parameters, such as aquifer hydraulic conductivities, the updated parameter and state ensembles change at each measurement update. The complete set of measurements is incorporated only at the final time of the historical period. In this case, the samples of the states obtained by inserting the posterior parameter samples into the model are different from the corresponding samples obtained from the Kalman filter. This is because the Kalman filter posterior state samples computed at a given time depend only on measurements obtained at or before this time while the smoother samples depend on all the measurements. Also, the sequential filter update requires that the forecasting model be stopped, reinitialized, and restarted at each measurement update. These difficulties can be resolved by using a version of the ensemble smoother (ES) proposed by van Leeuwen and Evensen (1996). The smoother updates the parameters and states simultaneously using all observations, without stopping and restarting the numerical model. The ES typically runs faster than the EnKF but can give less accurate posterior distributions when the states are nonlinear functions of the parameters (Emerick and Reynolds, 2012; van Leeuwen and Evensen, 1996). The different computational time required in the two methods occurs mainly because the EnKF requires additional time to restart model simulations. Examples of computation time differences observed for the two methods can be found in Emerick and Reynolds (2012). Bailey et al. (2012) applied ES to a more complicated agricultural aquifer-canal-stream example. However, in hydrology the ES has been applied primarily to synthetic cases or relatively simple examples (Dunne and Entekhabi, 2005; Bailey and Bau, 2010, 2012; Emerick and Reynolds, 2012). Generally speaking, the EnKF has been more popular for real-world applications, despite the conceptual limitations and inconsistencies that arise when it is used to assess uncertainty in time-invariant parameters.

Emerick and Reynolds (2012) developed an improved version of the ES, which they call an Ensemble Smoother with Multiple Data Assimilation (ESMDA). The ESMDA is an iterative smoothing algorithm that takes an approach that differs from either the ensemble Kalman filter or the traditional ensemble smoother. On each iteration of the algorithm the parameter ensemble is updated with all measurements at once, rather than sequentially. The state prediction ensemble for the current iteration is derived by running the forecast model for the entire data period with the current parameter values. This gives a predicted state ensemble that is consistent with the parameter ensemble, overcoming the EnKF inconsistency mentioned above. In this respect ESMDA is similar to the ES. However, in the ESMDA the entire smoothing process is repeated for several iterations until the parameter samples and associated state forecasts converge. The measurement noise covariance is adjusted between iterations to compensate for the reuse of the same data. Emerick and Reynolds (2012) show that this adjustment process is optimal for linear smoothing problems and a good approximation for some nonlinear problems.

Since the ESMDA is a relatively new method it has not yet been applied to complex hydrological models. This paper investigates the application of ESMDA in an uncertainty assessment of a computationally intensive distributed flow and transport model of the Yanqi Basin of Northwest China. We consider both the original version of ESMDA, which uses perturbed observations similar to

Download English Version:

<https://daneshyari.com/en/article/6410981>

Download Persian Version:

<https://daneshyari.com/article/6410981>

[Daneshyari.com](https://daneshyari.com)