Journal of Hydrology 530 (2015) 255–272

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models

Y.R. Fan [a], W.W. Huang [b], Y.P. Li [c], G.H. Huang [c,d,*], K. Huang [a]

[a] Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada
[b] Department of Civil Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada
[c] MOE Key Laboratory of Regional Energy and Environmental Systems Optimization, North China Electric Power University, Beijing 102206, China
[d] Institute for Energy, Environment and Sustainability Research, UR-NCEPU, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form 5 August 2015
Accepted 14 September 2015
Available online 30 September 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Niko Verhoest, Associate Editor

Keywords:
Uncertainty
Ensemble Kalman filter
Probabilistic collocation method
Gaussian anamorphosis
Hydrologic model
Monte Carlo

SUMMARY

In this study, a coupled ensemble filtering and probabilistic collocation (EFPC) approach is proposed for uncertainty quantification of hydrologic models. This approach combines the capabilities of the ensemble Kalman filter (EnKF) and the probabilistic collocation method (PCM) to provide a better treatment of uncertainties in hydrologic models. The EnKF method would be employed to approximate the posterior probabilities of model parameters and improve the forecasting accuracy based on the observed measurements; the PCM approach is proposed to construct a model response surface in terms of the posterior probabilities of model parameters to reveal uncertainty propagation from model parameters to model outputs. The proposed method is applied to the Xiangxi River, located in the Three Gorges Reservoir area of China. The results indicate that the proposed EFPC approach can effectively quantify the uncertainty of hydrologic models. Even for a simple conceptual hydrological model, the efficiency of EFPC approach is about 10 times faster than traditional Monte Carlo method without obvious decrease in prediction accuracy. Finally, the results can explicitly reveal the contributions of model parameters to the total variance of model predictions during the simulation period.

1. Introduction

Hydrologic models are simplified, conceptual representations of a part of the hydrologic cycle, which use relatively simple mathematical equations to conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed (Vrugt et al., 2005). Such conceptualization and aggregation lead to extensive uncertainties involved in both model parameters and structures, and consequently produce significant uncertainties in hydrologic predictions. Uncertainty in hydrologic predictions can originate from several major sources, including model structures, parameters, and measurement errors in model inputs (Ajami et al., 2007; Liu et al., 2012). Therefore, effective uncertainty quantification and reduction methods are required to produce reliable hydrologic forecasts for many real-world water resources applications, such as flooding control, drought management and reservoir operation (Fan et al., 2012, 2015a; Kong et al., 2015).

Previously, a number of probabilistic estimation methods have been proposed for quantifying uncertainty in hydrologic predictions. The probabilistic estimation methods approximate the posterior probability distributions of the hydrological parameters through the Bayesian theorem, conditioned on the streamflow observations. The generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992), Markov Chain Monte Carlo (Vrugt et al., 2009; Han et al., 2014), Bayesian model averaging (BMA) (Diks and Vrugt, 2010), and approximate Bayesian computation (Vrugt and Sadegh, 2013) methods are those extensively used probabilistic estimation methods. For instance, Madadgar and Moradkhani (2014) improved Bayesian Multi-modeling predictions through integration of copulas and Bayesian model averaging methods. DeChant and Moradkhani (2014b) proposed a full review of uncertainty quantification methods.

In a separate line of research, sequential data assimilation methods have been developed to explicitly handle various uncertainties and optimally merging observations into uncertain model predictions (Xie and Zhang, 2013; Zhang et al., 2012a,b; Zhang and Yang, 2013, 2014; Chang and Sayemuzzaman, 2014; Assumaning and Chang, 2014). In contrast to classical model calibration strategies, sequential data assimilation approaches continuously update

* Corresponding author at: Faculty of Engineering and Applied Sciences, University of Regina, Regina, Saskatchewan S4S 0A2, Canada. Tel.: +1 (306) 585 4095; fax: +1 (306) 585 4855.
E-mail addresses: huang@iseis.org, huangg@uregina.ca (G.H. Huang).

http://dx.doi.org/10.1016/j.jhydrol.2015.09.035
0022-1694/© 2015 Elsevier B.V. All rights reserved.

the states and parameters to improve model forecasts when new measurements become available (Vrugt et al., 2005). The prototype of sequential data assimilation techniques, the Kalman filter (KF) (Kalman, 1960) and the ensemble Kalman filter (EnKF) (Evensen, 1994), provide optimal frameworks for linear dynamic models with Gaussian uncertainties. The EnKF approach is one of the most frequently used data assimilation methods in hydrology due to its attractive features of real-time adjustment and easy implementation (Reichle et al., 2002). The EnKF method can provide a general framework for dynamic state, parameter, and joint state-parameter estimation in hydrologic models. For example, Moradkhani et al. (2005a) proposed a dual-state estimation approach based on EnKF for sequential estimation for both parameters and state variables of a hydrologic model. Weerts and El Serafy (2006) compared the capability of EnKF and particle filter (PF) methods in reducing uncertainty in the rainfall–runoff update and internal model state estimation for flooding forecasting purposes. Parrish et al. (2012) integrated Bayesian model averaging and data assimilation to reduce model uncertainty. DeChant and Moradkhani (2014a) combined ensemble data assimilation and sequential Bayesian methods to provide a reliable prediction of seasonal forecast uncertainty. Shi et al. (2014) conducted multiple parameter estimation using multivariate observations via the ensemble Kalman filter (EnKF) for a physically based land surface hydrologic model. However, due to the local complex characteristics of the watershed, some parameters in the hydrologic model may not be clearly identifiable and show slow convergence (Moradkhani et al., 2005b, 2012). Moreover, the same hydrologic model parameter may even show contrary convergence characteristics when different data assimilation methods are used. As shown by Moradkhani et al. (2005a,b), the C_{max} parameter for the Hymod was identifiable by using particle filter method but unidentifiable by using EnKF. Such unidentifiable parameters would lead to extensive uncertainties in hydrologic forecasts. Moreover, stochastic perturbations are usually added to the model inputs (e.g. precipitation, potential evapotranspiration, etc.) and observations (e.g. streamflow) to account for uncertainties in actual measurements. Such random noise would results in uncertainties in model parameters. Consequently, efficient forward uncertainty quantification methods (i.e. from model parameters to model predictions) are still desired for further analyzing the uncertainty in hydrologic predictions. Such methods can reveal the uncertainty evolution and propagation in hydrologic simulation.

Previously, Monte Carlo simulations are usually employed to quantify the uncertainty of hydrologic predictions resulting from uncertain model parameters (Knighton et al., 2014; Houska et al., 2014). In such a MC simulation process, model parameters would be sampled from known distributions, and each sample of model parameters would be entered into the hydrologic model to obtain statistics or density estimates of the model predictions. However, with complex hydrologic models such as distributed hydrologic models, this sampling approach is computationally intensive (Herman et al., 2013). The polynomial chaos expansions (PCEs) are effective for uncertainty propagation in stochastic processes, which represent the random variables through polynomial chaos basis and obtain the unknown expansion coefficients by the Galerkin technique or probabilistic collocation method (PCM) (Li and Zhang, 2007; Shi et al., 2009). The PCE-based methods have been widely used for uncertainty quantification of subsurface flow simulation in porous media (Li and Zhang, 2007; Shi et al., 2009), water quality modeling (Zheng et al., 2011), vehicle dynamics (Kewlani et al., 2012), mechanical systems (Blanchard, 2010), and so on. Fan et al. (2015b) integrated PCM into a hydrologic model for exploring the uncertainty propagation in hydrologic simulation, but it is only suitable for quantifying uncertainty of hydrologic models with specific distributions for model parameters (e.g. uniform, normal). However, in real-world hydrologic simulation, the posterior distributions of model parameters, after calibration through probabilistic estimation approaches, may be arbitrary.

In this study, a coupled ensemble filtering and probabilistic collocation (EFPC) method is proposed for uncertainty quantification of hydrologic models. In EFPC, the posterior distributions of model parameters will be approximated through EnKF; the obtained posterior distributions will be used as inputs for the probabilistic collocation method, in which PCEs will be constructed to connect the model parameters with the model responses. Such PCEs will reflect the uncertainty propagation between model parameters and its outputs. Therefore, the proposed EFPC will enable improved quantification of uncertainties existing in hydrologic predictions, model parameters, inputs and their interrelationships, and further reveal the uncertainty evolution through the obtained PCEs. Furthermore, a Gaussian anamorphosis (GA) approach will be presented to convert the obtained posterior distributions into standard normal random variables, which can be directly used as the inputs for PCM. The proposed approach will be applied to the Xiangxi River basin based on a conceptual rainfall–runoff model. The Xiangxi River basin, located in the Three Gorges Reservoir area of China, is one of the main tributaries in Hubei Province, with a draining area of about 3200 km^2. The Hymod, which has been used in many catchments, will be employed in this study (van Delft, 2007; Wang et al., 2009; Dechant and Moradkhani, 2012; Moradkhani et al., 2012). This application will help demonstrate the strength and applicability of the proposed methodology.

2. Methodology

2.1. Ensemble Kalman filter

The data assimilation methods have attracted increasing attention from hydrologists for exploring more accurate hydrological forecasts based on real-time observations (Moradkhani et al., 2005a; Weerts and El Serafy, 2006; Wang et al., 2009; DeChant and Moradkhani, 2011a,b; Plaza Guingla et al., 2013). Sequential data assimilation is a general framework where system states and parameters are recursively estimated/corrected when new observations are available. In a sequential data assimilation process, the evolution of the simulated system states can be represented as follows:

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

where f is a nonlinear function expressing the system transition from time $t-1$ to t, in response to model input vectors x_{t-1}^+ u_t and θ; x_{t-1}^+ is the analyzed (i.e. posteriori) estimation (after correction) of the state variable x at time step $t-1$; x_t^- is the forecasted (i.e. priori) estimation of the state variable x at time step t; θ represents time-invariant vectors, and ω_t is considered as process noise.

When new observations are available, the forecasted states can be corrected through assimilating the observations into the model, based on the output model responding to the state variables and parameters. The observation output model can be written as:

$$y_t^- = h(x_t^-, \theta) + v_t \tag{2}$$

where h is the nonlinear function producing forecasted observations; v_t is the observation noise.

The essential methods for states updating are based on Bayesian analysis, in which the probability density function of the current state given the observations is approximated by the recursive Bayesian law:

$$p(x_t, \theta_t | y_{1:t}) = \frac{p(y_t|x_t, \theta_t)p(x_t, \theta_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \tag{3}$$