Journal of Hydrology 519 (2014) 2722–2736

Contents lists available at ScienceDirect

Journal of Hydrology

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

An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting

Yuan Li [a,b], Dongryeol Ryu [a,*], Andrew W. Western [a], Q.J. Wang [b], David E. Robertson [b], Wade T. Crow [c]

[a] Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia
[b] CSIRO Land and Water Flagship, P.O. Box 56, Highett, Victoria 3190, Australia
[c] USDA-ARS Hydrology and Remote Sensing Laboratory, Rm. 104, Bldg. 007, BARC-W, Beltsville, MD 20705-2350, USA

ARTICLE INFO

Article history:
Available online 23 August 2014

Keywords:
Ensemble Kalman filter
Ensemble Kalman smoother
Maximum *a posteriori* estimation
Real-time flood forecasting

SUMMARY

For operational flood forecasting, discharge observations may be assimilated into a hydrologic model to improve forecasts. However, the performance of conventional filtering schemes can be degraded by ignoring the time lag between soil moisture and discharge responses. This has led to ongoing development of more appropriate ways to implement sequential data assimilation. In this paper, an ensemble Kalman smoother (EnKS) with fixed time window is implemented for the GR4H hydrologic model (modèle du Génie Rural à 4 paramètres Horaire) to update current and antecedent model states. Model and observation error parameters are estimated through the maximum *a posteriori* method constrained by prior information drawn from flow gauging data. When evaluated in a hypothetical forecasting mode using observed rainfall, the EnKS is found to be more stable and produce more accurate discharge forecasts than a standard ensemble Kalman filter (EnKF) by reducing the mean of the ensemble root mean squared error (*MRMSE*) by 13–17%. The latter tends to over-correct current model states and leads to spurious peaks and oscillations in discharge forecasts. When evaluated in a real-time forecasting mode using rainfall forecasts from a numerical weather prediction model, the benefit of the EnKS is reduced as uncertainty in rainfall forecasts becomes dominant, especially at large forecast lead time.

1. Introduction

Flood forecasting is important for timely flood warning and emergency responses but is also subject to uncertainties in input data, initial states, model structure and parameters (Sene, 2008). Data assimilation provides an effective way to integrate observation information, such as gauged discharge data and remotely sensed data, into hydrologic models to reduce these uncertainties (Liu et al., 2012). Past research has demonstrated that assimilating remotely sensed data, such as soil moisture (Crow and Ryu, 2009; Flores et al., 2012), evapotranspiration (Zhang et al., 2009), snow cover and snow water equivalent (Andreadis and Lettenmaier, 2006; He et al., 2012; Slater and Clark, 2006) can reduce the uncertainty in model outputs. However, assimilating discharge observations is still a more popular choice for real-time flood forecasting (Clark et al., 2008a; He et al., 2012; Komma et al., 2008; Lee et al., 2011; Lee et al., 2012; Ricci et al., 2011; Seo et al., 2009; Thirel et al., 2010), as discharge observations are more directly related to streamflow forecasts and, in many basins, are readily available for operational application.

Various data assimilation methods have been proposed for integrating discharge data into hydrologic models, including both sequential and variational techniques (Liu et al., 2012). Variational data assimilation has been used for operational forecasting systems (Lee et al., 2012; Seo et al., 2009); however, as a deterministic approach, it is less compatible with probabilistic real-time forecasting than sequential techniques. As a sequential and stochastic approach, the ensemble Kalman filter (EnKF) can improve streamflow forecasts by integrating observations to sequentially update model states, e.g., soil moisture (Komma et al., 2008; Li et al., 2013; McMillan et al., 2013; Thirel et al., 2010), and snow water equivalent (DeChant and Moradkhani, 2011b; He et al., 2012). In addition, the EnKF is sometimes applied to simultaneously update model states and parameters (Moradkhani et al., 2005b; Nie et al., 2011; Wang et al., 2009a). Despite the existence of more complex competing approaches, such as the particle filter (PF) (DeChant and Moradkhani, 2011a; Salamon and Feyen, 2009) and particle Markov chain Monte Carlo (PMCMC) methods (Andrieu et al., 2010; Moradkhani et al., 2012; Vrugt et al., 2013), the EnKF still remains

* Corresponding author. Tel.: +61 3 8344 7115.
E-mail address: dryu@unimelb.edu.au (D. Ryu).

http://dx.doi.org/10.1016/j.jhydrol.2014.08.009
0022-1694/

a popular choice for operational streamflow forecasting due to its computational efficiency and ability to directly generate ensemble forecasts (Clark et al., 2008a; He et al., 2012; Komma et al., 2008; McMillan et al., 2013).

One issue with the standard EnKF is that it is not designed to account for the time lag between errors in soil moisture and in discharge originating from water travel time between the hillslope and basin outlet. For hydrologic models whose routing processes are fully simulated through conceptual storage states, e.g., the probability distributed model (PDM) (Li et al., 2011; Li et al., 2013) and the conceptual hydrologic model (HyMOD) (Moradkhani et al., 2005a), the lagged errors accumulate in routing storages and can be effectively corrected by real-time filtering methods. In these cases, there is no need to account for time lags in the data assimilation procedure (Li et al., 2013; Moradkhani et al., 2005b). However, for other models whose routing processes are fully or partially simulated by unit hydrographs, e.g., the Hydrologiska Byrns Vattenbalansavdelning (HBV-96) model (Pauwels and De Lannoy, 2009) and the modèle du Génie Rural à 4 paramètres Journalier (GR4J model) (Li et al., 2013), the errors are lagged via unit hydrographs, which are difficult to directly update via stream discharge assimilation. For models employing unit hydrographs, the standard EnKF is not an ideal tool to correct errors in the antecedent states (Li et al., 2013; McMillan et al., 2013; Pauwels and De Lannoy, 2009).

To address this time lag issue, various new approaches have been investigated, which include shifting observations backwards by a fixed number of steps (Weerts and El Serafy, 2006), updating antecedent model states with a fixed lag (Meier et al., 2011), mapping the observational information to states within a time window through a retrospective EnKF (Pauwels and De Lannoy, 2006; Pauwels and De Lannoy, 2009), a recursive EnKF (McMillan et al., 2013), and an EnKS (Li et al., 2013). Even though the recursive EnKF (McMillan et al., 2013) modifies model states within a past time window, it updates antecedent and present model states *iteratively*. Due to its 'iterative' update-prediction process which uses the same observation multiple times, there is a potential risk of over-correcting the state error. The retrospective EnKF updates past and present states within a time window simultaneously then, however; it reruns the whole model from the beginning of the past analysis time window to obtain the model prediction. This means that corrected model states within the time window, except for the initial states, are not used for prediction. Consequently, error accumulates throughout the analysis time window and the benefits of considering a time lag can become marginal (Pauwels and De Lannoy, 2009). The EnKS with a fixed time window addresses the time lag issue in a theoretically robust and efficient manner and can produce more accurate forecasts than the standard EnKF; however, to date, the EnKS has only been assessed in synthetic studies (Li et al., 2013). As a result, its reliability in real data assimilation scenarios is uncertain.

Sequential data assimilation updates model state estimates using a weighted average of the background model prediction and the observation. The weights used in the averaging are largely determined by a comparison of model and observation uncertainties. Therefore, besides the time lag issue, quantifying these uncertainties remains a significant challenge in hydrologic data assimilation (Liu et al., 2012). The uncertainty of discharge observations is normally represented by applying either additive (Pauwels and De Lannoy, 2009) or multiplicative Gaussian noise (Clark et al., 2008a; DeChant and Moradkhani, 2012; Komma et al., 2008) or a combination of both (Noh et al., 2011). Model uncertainty is usually conceptually separated into several different sources including forcing error (e.g., precipitation, potential evapotranspiration, and temperature), model state error, parameter error, and model structure error. A large number of approaches have been applied to represent modeling errors in uncertainty analyses and/or data assimilation studies. For instance, precipitation error can be assumed to be additive (Weerts and El Serafy, 2006) or multiplicative random noises (DeChant and Moradkhani, 2012) or to follow a more sophisticated error model based on the consideration of spatial-temporal correlation (Clark et al., 2008a; McMillan et al., 2013). Parameter error can be represented by direct perturbation (DeChant and Moradkhani, 2012; Nie et al., 2011) or simulated through Markov Chain Monte Carlo (MCMC) based approaches (Moradkhani et al., 2012; Vrugt et al., 2013). Model structural error can be characterized by multi-model predictions and addressed through Bayesian model averaging (Duan et al., 2007; Leisenring and Moradkhani, 2012). Finally, model state error can be characterized directly through absolute or relative noises (Li et al., 2013; Noh et al., 2011; Ryu et al., 2009) or propagated from forcing and other uncertainties (Komma et al., 2008; Weerts and El Serafy, 2006).

While various error models can be applied, determining parameter values for these models is generally challenging. For discharge observations, it is possible to approximately quantify the uncertainty from the observation itself, for example, by analyzing the uncertainty of rating curves used to calculate discharge (Clark et al., 2008a). But it is hard to directly quantify forcing and model uncertainties. To address this challenge, two types of approaches have been suggested. One type is to use adaptive filtering techniques, which quantify observation and model errors during the online cycling of data assimilation (Crow and Reichle, 2008; Crow and van den Berg, 2010). However, these approaches rely on a number of assumptions, including the serial independence of observation errors. More sophisticated adaptive filtering approach, which require a reduced set of assumptions, have been developed (Crow and Yilmaz, 2014) but not widely applied or tested. A second approach for estimating error parameter are likelihood-based/Bayesian uncertainty estimation techniques which disaggregate the total differences between model predictions and observations into various sources of uncertainties. Specific examples include the generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992), the framework for understanding structural errors (FUSE) (Clark et al., 2008b), the Bayesian total error analysis (BATEA) (Kavetski et al., 2006; Renard et al., 2011), and the differential evolution adaptive metropolis (DREAM) algorithm (Vrugt et al., 2008). However, unlike the particle filtering techniques, these uncertainty estimation approaches have not been widely applied to quantify error parameters for data assimilation. One exception is the integrated uncertainty and ensemble-based data assimilation (ICEA) system (He et al., 2012), which use an MCMC-based uncertainty analysis tool to generate ensemble spread for the EnKF. In addition to the adaptive filtering and likelihood-based/Bayesian uncertainty estimation techniques, Leisenring and Moradkhani (2012) suggested a variable variance multiplier approach, which was further improved by Moradkhani et al. (2012). It rescales variance multipliers (error parameters) according to the accuracy of the mean predictions relative to a certain confidence interval (CI) of ensemble predictions. Therefore, it can complement the uncertainty estimation methods as an error variance corrector.

The aim of this paper is to address the time lag issue and error estimation issue through an integrated data assimilation scheme. Specifically, the focus is twofold: (1) testing the EnKS in real data assimilation scenarios and (2) examining the potential of likelihood-based/Bayesian approaches to inform error parameters required for hydrologic data assimilation. To achieve the objective, an ensemble-based maximum *a posteriori* (MAP) estimation method, which is a relatively simple likelihood-based approach, is incorporated to quantify the model and observation error parameters. The observational error information drawn from the analysis of gauging data is used as an *a prior*i constraint for the