Journal of Hydrology 519 (2014) 2663–2675

Contents lists available at ScienceDirect

Journal of Hydrology

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

Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models

Arezoo Rafieeinasab [a,*], Dong-Jun Seo [a], Haksu Lee [b], Sunghee Kim [a]

[a] Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX, USA
[b] LEN Technologies, Oak Hill, VA, USA

ARTICLE INFO

Article history:
Available online 7 July 2014

Keywords:
Data assimilation
Maximum likelihood ensemble filter
Ensemble Kalman filter
Initial conditions
Nonlinear observation equations

SUMMARY

Various data assimilation (DA) methods have been used and are being explored for use in operational streamflow forecasting. For ensemble forecasting, ensemble Kalman filter (EnKF) is an appealing candidate for familiarity and relative simplicity. EnKF, however, is optimal in the second-order sense, only if the observation equation is linear. As such, without an iterative approach, EnKF may not be appropriate for assimilating streamflow data for updating soil moisture states due to the strong nonlinear relationships between the two. Maximum likelihood ensemble filter (MLEF), on the other hand, is not subject to the above limitation. Being an ensemble extension of variational assimilation (VAR), MLEF also offers a strong connection with the traditional single-valued forecast process through the control, or the maximum likelihood, solution. In this work, we apply MLEF and EnKF as a fixed lag smoother to the Sacramento (SAC) soil moisture accounting model and unit hydrograph (UH) for assimilation of streamflow, mean areal precipitation (MAP) and potential evaporation (MAPE) data for updating soil moisture states. For comparative evaluation, three experiments were carried out. Comparison between homoscedastic vs. heteroscedastic modeling of selected statistical parameters for DA indicates that heteroscedastic modeling does not improve over homoscedastic modeling, and that homoscedastic error modeling with sensitivity analysis may suffice for application of MLEF for soil moisture updating using streamflow data. Comparative evaluation with respect to the model errors associated with soil moisture dynamics, the ensemble size and the number of streamflow observations assimilated per cycle showed that, in general, MLEF outperformed EnKF under varying conditions of observation and model errors, and ensemble size, and that MLEF performed well with an ensemble size as small as 5 while EnKF required a much larger ensemble size to perform closely to MLEF. Also, MLEF was not very sensitive to the uncertainty parameters and performed reasonably well over relatively wide ranges of parameter settings, an attribute desirable for operational applications where accurate estimation of such parameters is often difficult.

1. Introduction

Uncertainties in the initial conditions (IC) of soil moisture and observed boundary conditions (BC) of precipitation and potential evaporation (PE) introduce considerable errors in hydrologic forecasts. In recent years, data assimilation (DA) has been gaining great attention to reduce these uncertainties (Liu et al., 2013; Brocca et al., 2010; De Lannoy et al., 2007; Ibbitt et al., 2007; Liu and Gupta, 2007; Seo et al., 2003; Reichle et al., 2002). DA makes inference on the model states by bringing together all available observations from often disparate sources, quantifying the uncertainties in the model and observation errors, and updating the state variables by optimally combining model predictions with observations. In addition to DA in the single-valued sense, ensemble DA is also necessary to allow state updating in operational ensemble forecasting systems (Seo et al., 2006; Demargne et al., 2014; Cloke and Pappenberger, 2009; Schaake et al., 2007a,b; Schellekens et al., 2011; Thielen et al., 2008; Werner et al., 2005, 2009).

In hydrologic forecasting, one would ideally like to assimilate soil moisture observations to update the model ICs of soil moisture, in which case the observation equation would be linear. In reality, however, soil moisture states are seldom observed in-situ and,

* Corresponding author. Address: Department of Civil Engineering, The University of Texas at Arlington, Box 19308, 416 Yates St. Suite 425, Arlington, TX 76019, USA. Tel.: +1 817 272 9130, fax: +1 817 272 263.

E-mail address: arezoo.rafieeinasab@mavs.uta.edu (A. Rafieeinasab).

http://dx.doi.org/10.1016/j.jhydrol.2014.06.052
0022-1694/

even if such measurements are available, they are generally not representative of the conditions at the scale where the hydrologic models operate. Streamflow observations, on the other hand, are much more widely available and reflect the catchment-wide conditions, albeit only in some spatiotemporally integrated sense. For the use of streamflow data for updating of soil moisture states, the observation equation involved is generally highly nonlinear, which poses an additional challenge in DA.

Various DA techniques such as Kalman filtering (Kalman, 1960), variational assimilation (VAR, Jazwinski, 1970; Li and Navon, 2001; Seo et al., 2003, 2009), particle filtering (Weerts and Serafy, 2006), etc., have their own merits and demerits (Liu and Gupta, 2007; Liu et al., 2013). Extensions of Kalman filter have been developed to deal with nonlinear systems. Extended Kalman filter (EKF), e.g., involves linearizing the model dynamics using the first-order Taylor series approximation (Maybeck, 1979). To overcome the limitations of EKF, a Monte Carlo-based Kalman filter, or EnKF, was proposed by Evensen (1994). The novelty of EnKF is in its ability to deal with nonlinear model dynamics naturally without linearizing model equations (Moradkhani et al., 2005). Unlike VAR, EnKF does not assume temporally constant model error covariance or requires a separate adjoint model. For the above reasons and algorithmic simplicity, EnKF has gained great popularity in various applications recently (Evensen, 2003; Chen et al., 2011; Xie and Zheng, 2010).

Variations of EnKF and different types of ensemble filter have also been developed. Anderson (2001) proposed an ensemble-based filter called ensemble adjustment Kalman filter (EAKF) in which both the mean and covariance of updated ensembles are preserved. He concluded that EAKF is superior to EnKF especially for small ensemble sizes. Another variant of EnKF was introduced by Whitaker and Hamill (2002) called ensemble square root filter (ESRF) in which the perturbation of observation is avoided. Sakov and Oke (2008) presented a linear approximation of ESRF with comparable performance. Bocquet (2011) proposed a deterministic variant of EnKF named finite-size ensemble transform Kalman filter (ETKF-N) which is less sensitive to sampling errors. Van Leeuwen and Evensen (1996) introduced ensemble smoother (ES), and Evensen and van Leeuwen (2000) developed ensemble Kalman smoother (EnKS). Cohn et al., (1994) used a fixed-lag smoother to incorporate all available observation at current time as well as a fixed amount of time past each analysis time. In this work, we use a fixed-lag smoother formulation of EnKF.

Unlike VAR, however, EnKF and its variants assume linear observation equation. As such, if the observation equation is nonlinear as in assimilating streamflow data for updating soil moisture states, EnKF may not be expected to perform well. To address the above limitation in EnKF, Zupanski (2005) developed maximum likelihood ensemble filer (MLEF) which combines the strength of EnKF and VAR. MLEF may be considered as an ensemble extension of VAR in which, once the VAR solution-like maximum likelihood or control solution is obtained, ensemble members are generated by perturbing the control states and propagating them forward as in EnKF. The purpose of this work is to compare EnKF with MLEF for assimilation of streamflow data in soil moisture updating.

In MLEF, the analysis solution is obtained as a model state that maximizes the posterior conditional probability distribution. The maximum likelihood solution, in the single-valued sense, is superior to ensemble mean if the normality assumption is not met. In operational forecasting, provision of such a "most likely" solution, in addition to the ensemble members, is very important to the human forecasters as the former provides a reference solution for the existing manual DA process, referred to as run-time modifications (MOD) in NWS (Seo et al., 2009). The maximum likelihood state is estimated via iterative minimization, thus making the MLEF approach closely related to iterated Kalman filter (Jazwinski, 1970; Cohn, 1997; Zupanski, 2005). As with other ensemble data assimilation algorithms, MLEF produces an estimate of the uncertainty in the analysis solution (e.g., analysis error covariance). Unlike VAR, however, MLEF does not require an adjoint code and solves a reduced-rank problem in ensemble subspace with superior preconditioning (Zupanski, 2005).

MLEF has been used in various studies such as carbon transport (Lokupitiya et al., 2008; Zupanski et al., 2007), aerosol retrieval (Carrio et al., 2008), wind power forecasting (Zupanski et al., 2010) and targeting additional observations for forecasting of tropical cyclones (Kim et al., 2010). To the best of the authors' knowledge, however, MLEF has never been used in hydrologic applications or objectively compared with EnKF for streamflow assimilation until this paper. Additional significant new contributions of this paper include systematic sensitivity analysis of DA performance with respect to the ensemble size, the number of streamflow observations assimilated per cycle and the magnitude of model and observational errors, and comparative evaluation of performance of DA under homoscedastic and heteroscedastic modeling of observation errors.

It is noted here that the evaluation carried out in this work is in the single-valued sense only. That is, we only consider the DA techniques as minimization tools for single-valued analysis. By "single-valued", we mean analysis or prediction expressed by a single representative value, such as the maximum likelihood solution in MLEF or the ensemble mean in EnKF, rather than by multiple values such as an ensemble. The term single-valued forecast was introduced recently in the hydrologic literature (Schaake et al., 2007a,b; Wu et al., 2011; Regonda et al., 2013; Demargne et al., 2014) to distinguish from deterministic forecast. The paper is organized as follow. Section 2 describes the formulation of the assimilation problem. Section 3 describes the EnKF and MLEF methodologies. Error modeling is described in Section 4. Section 5 describes the study basins, data used and experiment design. We present the results in Section 6. Section 7 provides conclusion and future research recommendations.

2. Formulation of the assimilation problem

Assume a headwater basin with a stream gauge at the outlet with hourly observations of streamflow, mean areal precipitation (MAP) and mean areal potential evaporation (MAPE) available in real time. Assume also that lumped rainfall–runoff and routing models operate for continuous simulation and prediction of streamflow at the catchment outlet. The rainfall–runoff and routing models used in this work are the Sacramento soil moisture accounting model (SAC) (Burnash et al., 1973) and unit hydrograph (UH) (Chow et al., 1988), respectively. The SAC model has six state variables which are updated by DA: the upper zone tension water content (UZTWC), the upper zone free water content (UZFWC), the lower zone tension water content (LZTWC), the lower zone supplemental free water content (LZFSC), the lower zone primary free water content (LZFPC) and tension water content in the additional impervious area (ADIMC) (Burnash et al., 1973).

Our problem is then to assimilate the observations of MAP, MAPE and streamflow for real-time updating of the soil moisture states of the rainfall–runoff model. To account for the time lag between the generation of runoff and its arrival at the catchment outlet, we formulate the DA problem as fixed-lag smoothing (Schweppe, 1973; Li and Navon, 2001) following Seo et al. (2003, 2009) and Lee et al. (2011, 2012). The size of the fixed lag, or the assimilation window, should be comparable to the response time of the basin. The experience thus far indicates that the size of the window should be about the length of the unit hydrograph, or the basin response time of fast runoff (Seo et al., 2003, 2009; Lee