Journal of Hydrology 475 (2012) 428–440

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Integration of evolutionary based assimilation into Kalman-type methods for streamflow simulations in ungauged watersheds

Gift Dumedah [a,*], Paulin Coulibaly [b]

[a] Department of Civil Engineering, Monash University, Building 60, Melbourne, Victoria 3800, Australia
[b] School of Geography and Earth Sciences, Department of Civil Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L8

ARTICLE INFO

Article history:
Received 27 June 2012
Received in revised form 9 September 2012
Accepted 24 October 2012
Available online 2 November 2012
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Attilio Castellarin, Associate Editor

Keywords:
Data assimilation
Ungauged basins
Regionalization
Ensemble Kalman filter
Particle filter
Multi-objective evolutionary algorithms

SUMMARY

Data assimilation (DA) has emerged as a valuable tool for the design and application of streamflow forecasting systems. But DA applications for streamflow simulations in ungauged basins are still very limited primarily because most updated ensemble members are not usually associated with converged state and model parameterizations. Other limitations include the evaluation of massive number of ensemble members, weak/unknown relationships between parameter values and predictors, and the transfer of several members from gauged watersheds to ungauged ones is computationally expensive. But the inherent dynamics of DA to account for uncertainties in model, forcing data, and imperfect observation provide an appealing approach to simulate watershed response in ungauged basins. This study proposes a DA method namely the Pareto-Particle-Ensemble Kalman Filter (ParetoParticleEnKF) to generate and archive a small number of continuously evolved members using multi-objective evolutionary strategy where these members are updated using particle and ensemble Kalman filtering methods. The archived members for gauged watersheds are combined using inverse distance weighting where they are applied to simulate watershed response in ungauged basins.

The proposed method is demonstrated by assimilating daily streamflow into the Sacramento Soil Moisture Accounting (SAC-SMA) model for 10 watersheds in southern Ontario, Canada. After successfully transferring ensemble members from gauged watersheds to ungauged ones, the updated ensembles were applied to simulate streamflow for up to 10-days ahead to determine how long into the future would the quality/accuracy of simulations persist before they begin to deteriorate in the ungauged basin. The results show that the designed method can facilitate simulation of accurate streamflows for any time step, and generate accurate simulations for up to 10 days ahead in the ungauged basins. A unique evaluation procedure is the transfer of updated members in the form of forcing data uncertainties, and state and model parameterizations from gauged to ungauged watersheds and their subsequent assessment for multiple time step ahead simulations. The overall decline in accuracy from gauged to ungauged watersheds for the entire 10-day lead time across all 10 watersheds is 20% for Nash–Sutcliffe efficiency and 12% for percent bias.

1. Introduction

Data assimilation (DA) continue to facilitate and improve hydrological forecasts (Dumedah et al., 2011; Thirel et al., 2010; Xie and Zhang, 2010; Van Leeuwen, 2009; Clark et al., 2008). The forecasting improvements are usually illustrated for a single catchment but there is a growing demand to transfer forecast skill and its associated state and model parameter values from gauged watersheds to ungauged ones – a process usually called regionalization (Zhang and Chiew, 2009). Regionalization methods link gauged watersheds to ungauged ones through the transfer of model parameter values and their associated catchment attributes (Bastola et al., 2008; Oudin et al., 2008b; Zhang et al., 2008; Gotzinger and Bardossy, 2007; Parajka et al., 2005). A persistent challenge to applying DA in regionalization studies remains the evaluation of massive number of model ensembles for determining the updated ensemble members and subsequent evaluation of these members for streamflow forecasting (Parada and Liang, 2010; Sivapalan et al., 2003). The transfer of hundreds of ensemble members for each simulation time step is computationally expensive. Additionally, weak relationships between model parameters and predictors of watershed response may limit the performance of regionalization methods (Parada and Liang, 2010; Yadav et al., 2007).

Moreover, in most DA methods the model state and parameterizations are not often adequately adjusted to re-generate the updated ensemble members. This means that the ensemble simulations are not necessarily generated from converged state and

* Corresponding author. Tel.: +61 3 9905 5934; fax: +61 3 9905 4944.
E-mail address: dgiftman@hotmail.com (G. Dumedah).

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2012.10.033

model parameterizations. For example, in Kalman-type and particle filter methods the updated ensemble is a statistical approximation (or a weighted compromise) between simulation and observation (usually perturbed). As a result, the updated members may have streamflows that may not be physically possible within the context of the hydrological model due to the dynamics between states and model parameters. That is, the updated states and model parameterizations from which the ensemble simulations are generated may not necessarily reproduce the updated ensemble members. These have important implications on any forecasting procedure where future model estimates are made from a model whose state has not been adequately parameterized or converged to determine the updated members. But the theoretical foundation of DA provides an appealing concept to account for uncertainties in model, forcing data, and imperfect observations when estimating watershed response in ungauged basins.

Recent studies using evolutionary based data assimilation have shown that the dynamics of large number of members could be approximated by a small number of continuously evolved members (Dumedah et al., 2011; Dumedah, 2012c; Dumedah and Coulibaly, 2012b). In particular, the updated members are tightly linked to these evolved members which represent specific values for states and model parameterizations. These members have been illustrated in Dumedah et al. (2011) and Dumedah (2012c) to efficiently merge ensemble simulations with new observational information, and to improve the assimilation and forecast accuracy when applied in a single watershed. However, the forecasting performance of these members (and DA methods in general) has not been thoroughly investigated to simulate watershed response for neighboring (or geographically close) watersheds in the literature.

As a result, this study will demonstrate a designed DA method to generate and archive updated ensemble members for regionalization studies, and to subsequently apply these members (or solutions) to simulate streamflow in neighboring ungauged watersheds. The designed DA method uses evolutionary based assimilation to generate continuously evolved members (including the Pareto-optimal set) where they are separately updated using the ensemble Kalman filter (EnKF) and particle filtering methods. That is, the designed DA is an integration of Pareto-optimal members into particle filter and EnKF (denoted ParetoParticleEnKF) to produce a small number of updated members. The updated members for gauged watersheds consisting of state and model parameterizations, and forcing data uncertainties are archived for different assimilation time steps before they are applied to simulate streamflow in ungauged basins. Instead of evaluating hundreds of members, the designed method transfers a small number of ensemble members from the assimilation in gauged watersheds to ungauged ones.

The remaining sections of this paper are organized as follows. This current section provides additional background information on the proposed DA method. The next section describes the study area comprising 10 watersheds and the SAC-SMA model. This section also provides an outline of the designed DA method, ParetoParticleEnKF. The results and discussions section presents the assimilation outputs for various watersheds and the evaluation of their archived members in neighboring ungauged watersheds. The findings of this study and utility of the ParetoParticleEnKF are summarized in the conclusions section.

1.1. Brief background on the proposed approach

Evolutionary algorithms use the concept of evolution and natural selection by allowing several members in a population to compete among themselves based on evaluation objectives. The fitter (or competitive) members are naturally selected and varied to reproduce a new population through crossover and mutation–variation operators which modify and maintain diversity between members. The continuous selection of fitter members and their subsequent variation (or reproduction) usually leads to changes in the evaluation conditions. That is, for each population the fitness of its members usually increase with every cycle of evolution which consequently increases/changes the evaluation conditions. Each cycle of evolution of a population to reproduce a new population of members is called a generation. As a result, members evolve within a population whereas the population evolves through generations to increase/change the evaluation conditions and to generate fitter population of members. This continuous evolution and variation of population of members results into the Pareto-optimal set – a set of equally competitive members which are not dominated when compared to other members using the evaluation objectives.

The evolutionary based assimilation uses the variational DA approach (Caparrini et al., 2004; Caparrini et al., 2003; Reichle et al., 2001) to minimize the cost (or penalty) function by finding the optimal least squares estimate based on the model estimate, the observation, and their associated uncertainties. The cost function and other accuracy measures such as the root mean square error are used as evaluation objectives to evolve a population of competitive members through several generations at each assimilation time step and between time steps. For each time step, several members are evaluated but only the final fitter population which comprises the non-dominated members is chosen to represent the updated ensembles to determine the ensemble mean and its variance. Note that a solution (which is the same as an ensemble member) may represent a vector comprising values for states, model parameters, and forcing data uncertainty which are applied in the measurement (i.e., the hydrological) model to generate streamflow. Further information on the computational procedure for evolutionary based assimilation is outlined in Section 2.2.

The EnKF and particle filter are advanced DA methods with clear and standard computational procedures, and they have been applied in several studies including: Xie and Zhang (2010), Van Leeuwen (2009), Komma et al. (2008a), Clark et al. (2008), Vossepoel and van Leeuwen (2007), Clark et al. (2006b), Weerts and El Serafy (2006), Moradkhani et al. (2005), and Moradkhani and Hsu (2005). The EnKF was developed by Evensen (1994) where it uses Monte Carlo integration to estimate the posterior probability density function (pdf) through the ensemble mean and its covariance. The ensembles for states and model parameters are propagated to make predictions for future time. The ensemble of simulated outputs is combined with perturbed observation to determine the Kalman gain (denoted, K), where it is applied to update state and model parameter components. The procedure is repeated to evolve state and model parameter components through subsequent assimilation time steps.

The particle filter was designed by Gordon et al. (1993) where it also estimates the posterior pdf but through the use of weighted ensemble members. As in the EnKF, the ensembles for states and model parameters are propagated forward in time to generate ensemble predictions. These predictions are combined with perturbed observation to determine the ensemble weight – the weights are re-sampled so that low weighted members are replaced with normalized weighted members. The re-sampled weights are used to update state and model parameter components. The procedure is repeated to evolve state and model parameter components through subsequent assimilation time steps.

The designed DA method, ParetoParticleEnKF is illustrated by assimilating daily streamflow into the Sacramento Soil Moisture Accounting (SAC-SMA) for 10 watersheds in southern Ontario, Canada. Using leave-one-out approach, the updated members which are archived for the gauged watersheds are combined using inverse distance weighting to simulate streamflow in the ungauged watersheds. Additionally, the archived members for different