Journal of Hydrology 519 (2014) 2707–2721

Contents lists available at ScienceDirect

Journal of Hydrology

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

On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models

Seong Jin Noh [a,*], Oldřich Rakovec [b,c], Albrecht H. Weerts [b,d], Yasuto Tachikawa [e]

[a] Water Resources Research Division, Department of Water Resources & Environmental Research, Korea Institute of Civil Engineering and Building Technology, Republic of Korea
[b] Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, The Netherlands
[c] UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany
[d] Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands
[e] Dept. of Civil and Earth Resources Eng., Kyoto Univ., Japan

ARTICLE INFO

Article history:
Available online 17 August 2014

Keywords:
Streamflow forecasting
Rainfall ensemble generator
Lagged particle filtering
Numerical weather prediction
Distributed hydrologic model

SUMMARY

We investigate the effects of noise specification on the quality of hydrological forecasts via an advanced data assimilation (DA) procedure using a distributed hydrological model driven by numerical weather predictions. The sequential DA procedure is based on (1) a multivariate rainfall ensemble generator, which provides spatial and temporal correlation error structures of input forcing, and (2) lagged particle filtering to update past and current state variables simultaneously in a lag-time window to consider the response times of internal hydrologic processes. The procedure is evaluated for streamflow forecasting of three flood events in two fast-responding catchments in Japan (Maruyama and Katsura). The rainfall ensembles are derived from ground-based rain gauge observations for the analysis step and numerical weather predictions for the forecast step. The ensemble simulation performs multi-site updating using information from the streamflow gauging network and considers the artificial effects of reservoir release. Sensitivity analysis is performed to assess the impacts of noise specification in DA, comparing a different setup of random state noise and input forcing with/without multivariate conditional simulation (MCS) of rainfall ensembles. The results show that lagged particle filtering (LPF) forced with MCS provides good performance with small and consistent random state noise, whereas LPF forced with Thiessen rainfall interpolation requires larger random state noise to yield performance comparable to that of LPF + MCS for short lead times.

1. Introduction

Flood is a natural hazard that occurs after extensive rainfall or snowmelt events. Accurate and reliable flood early warning systems can mitigate the number of casualties and economic damage related to flood, which causes particular problems in densely populated areas (e.g. Sene, 2008). Although important advances have been achieved in flood forecasting (e.g. Biondi and De Luca, 2013; Cloke and Pappenberger, 2009; Collier, 2007; Hapuarachchi et al., 2011; Kitanidis and Bras, 1980; and references therein), current knowledge is insufficient for accurate prediction with the required lead times owing to various uncertainties originating from simulation models, observations, and forcing data. The quality of hydrological forecasting systems is dependent on several factors: (1) the quality of the hydrological model (in terms of its structure and parameter estimates) and its suitability for a given catchment; (2) the initial conditions of the model states at the start of the forecast; (3) external forcing during the forecasted period (weather forecasts); and (4) human control (e.g. reservoir operations, irrigation) (e.g. Kumar, 2011; Moll, 1986; Weerts et al., 2013). To increase the certainty of the hydrological forecast, i.e. the forecast of the magnitude and timing of a flooding event, all of these sources of uncertainty must be considered and propagated through the hydrological modeling chain embedded in a flood forecasting system. This can be achieved using state updating or, more broadly speaking, data assimilation (DA). In this context, DA is a technique used to reduce uncertainty by combining the uncertain hydrological simulations of model states and fluxes with actual measurements/observations of river stages/streamflow, soil water content, or groundwater levels.

Numerous sophisticated DA algorithms have been proposed, from rules-based, direct-insertion methods to advanced smoothing and sequential techniques. These techniques and their many

* Corresponding author.
E-mail address: seongjin.noh@gmail.com (S.J. Noh).

http://dx.doi.org/10.1016/j.jhydrol.2014.07.049
0022-1694/

variants (Liu et al., 2012 and references therein) have facilitated recent progress in hydrologic DA for streamflow prediction (e.g. Aubert et al., 2003; Clark et al., 2008; DeChant and Moradkhani, 2012; Han et al., 2012; Komma et al., 2008; Lee et al., 2011; McMillan et al., 2013; Mendoza et al., 2012; Moradkhani et al., 2005; Noh et al., 2011, 2013a; Pauwels and De Lannoy, 2009; Pauwels et al., 2013; Rakovec et al., 2012b; Salamon and Feyen, 2010; Seo et al., 2009; Shiiba et al., 2000; Smith et al., 2012; Weerts and El Serafy, 2006; among others).

Although it is generally desirable that all sources of uncertainty in a data assimilation scheme should be constrained as much as possible, this cannot always be achieved. As discussed by Seo et al. (2009), the correction of model inputs, states, initial conditions, and parameters is often conducted in a rather empirical and subjective way, which may reduce the credibility and transparency of operational forecasts. For example, input uncertainty such as that associated with rainfall and weather is assumed to be governed by different distributions: the lognormal distribution (DeChant and Moradkhani, 2012), the normal distribution (Weerts and El Serafy, 2006), or a mixture of uniform distributions (Clark et al., 2008) with different error variances using multiplicative, additive, or mixture noises. State variables such as soil moisture are usually perturbed by random normal noises with or without consideration of temporal correlations using hyper-parameters (e.g. Clark et al., 2008; Noh et al., 2013a). Observation uncertainty of streamflow is often assumed to follow a normal or lognormal distribution based on different formulations (e.g. McMillan et al., 2013; Moradkhani et al., 2005; Weerts and El Serafy, 2006). Leisenring and Moradkhani (2012) used a variable variance multiplier to tune the noise level. However, the impact of different noise specifications on performance has not yet been treated explicitly in the operational setup of models. Crow and Van Loon (2006) already demonstrated that incorrect assumptions can affect the final outcome of an assimilation scheme. Therefore, advances in DA not only imply improvement of accuracy but also embrace an approach to investigate, obtain, and adopt the best possible error description. As an example, Rakovec et al. (2012a) proposed a multivariate rainfall generator that was used to generate an ensemble of model forcings that considered spatial and temporal correlations based on rain gauge observations and simulations; this ensemble was used as an input into a data assimilation scheme using a distributed hydrological model (Rakovec et al., 2012b).

Lagged filtering methods, which were first suggested for particle filtering (PF) by Noh et al. (2011a) and later extended to ensemble Kalman filtering (EnKF) (McMillan et al., 2013; Noh et al., 2013a), are DA algorithms that consider both time lags in the routing process and the different timescales of hydrologic processes. In the present study, we focus on primarily on PF because it does not require strict assumptions such as linearity or a Gaussian distribution which is not feasible to be kept in most of hydrological forecasting. Unlike Kalman filter-based methods, PF performs updating on particle weights instead of state variables (Liu and Gupta, 2007), thus reducing numerical instability, especially in physically based or process-based models. For example, for updating state variables of soil moisture simulated by the Green–Ampt equation (Green and Ampt, 1911), Kalman filter-based methods use linear updating rules to innovate the depth of wetting front at each grid without considering numerical stability and physical conditions (e.g. ponding conditions at soil surface), while PF selects and duplicates particles (model ensembles) with high weights including the total states (and parameters) of the simulation system.

Meanwhile, owing to the rapid development of computing power, applications of numerical weather prediction (NWP) and distributed hydrologic modeling (DHM) have provided alternative directions in flood forecasting. Although the quality of NWP is still considered to be limited by uncertainties regarding the localization, timing, and intensity of events (Habets et al., 2004), advances in flood forecasting and reservoir operation using NWP have been reported (e.g. Jasper et al., 2002; Pappenberger et al., 2008; Saavedra Valeriano et al., 2010; Schellekens et al., 2011; Smiatek et al., 2012; Verbunt et al., 2006; Wang et al., 2012). In addition, despite its inherent uncertainty (Beven, 1993), applications of DHM have been increasing with increases in the accessibility of spatial and temporal information about the earth system from various sources (e.g. remote sensing). However, few attempts have been made to incorporate NWP and DHM into a DA framework to quantify uncertainties in an operational flood forecasting setting.

The present study aims primarily to gain insight in the effects of noise specification in a DA scheme on the quality of hydrological forecasts. Because it is difficult to disentangle the effects of noise specification for distributed models, we limit the scope of this study to comparison of two methods of noise specification in a DA scheme: (1) soil moisture state perturbation based on white noise with error assumptions based on expert knowledge (e.g. Noh et al., 2011a), and (2) input forcing using a multivariate rainfall generator (Rakovec et al., 2012a,b). The intention behind this comparison is not to advocate that uncertainty in the boundary conditions may substitute epistemic uncertainty regarding model structure model error, but rather that proper specification of the forcing uncertainty may lead to additional specification of model structure uncertainty. The effects of the noise specification are investigated by conducting several DA experiments in two fast-responding catchments in Japan (Maruyama and Katsura) using a DHM. Finally, the operational setup of the framework is tested by combining both approaches with NWP for three flood events.

The remainder of this paper is organized as follows. Section 2 describes the study catchments, the hydrologic model, the lagged PF method, noise models (including state perturbation and rainfall ensemble generator), the numerical weather prediction data, and the procedures involved in the DA experiments. In Section 3, we present our results with reference to the rainfall ensembles, assess the impacts of uncertainties of DA through sensitivity analysis of the magnitude of random noise and different combinations of DA methods, and discuss the performance in forecasting and hindcasting modes. Finally, in Section 4 we summarize our results and discuss the main findings.

2. Materials and methods

2.1. Study catchments

The present study was conducted within two Japanese catchments (Fig. 1), the Maruyama River catchment (909 km^2) (Fig. 2) and the Katsura River catchment (887 km^2) (Fig. 3), in which rainfall occurs primarily in the summer season from May to September. Both catchments can be characterized as fast-responding, with a time of concentration of less than 10 h. There are 10 rain gauges and 4 streamflow gauges in the Maruyama catchment and 13 rain gauges and 1 streamflow gauge in the Katsura catchment. All rain and streamflow gauges are operated by the Ministry of Land, Infrastructure, Transport and Tourism in Japan (http://www1.river.go.jp/) and are available at hourly time steps. The hourly time series of air temperature, relative humidity, wind speed, and duration of sunlight were obtained from two meteorological stations operated by the Japan Meteorological Agency (http://www.jma.go.jp).

The dominant land use types of the Maruyama River catchment include forest (85%), agricultural areas (7%), residential areas (2%), inland water and other (5%) (Hunukumbura et al., 2012), whereas