International Journal of Sediment Research 31 (2016) 149–158

Contents lists available at ScienceDirect

International Journal of Sediment Research

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

Original Research

Coupling the *k*-nearest neighbor procedure with the Kalman filter for real-time updating of the hydraulic model in flood forecasting

Kailei Liu [a], Zhijia Li [a,b], Cheng Yao [a,*], Ji Chen [c], Ke Zhang [d,e], Muhammad Saifullah [a]

[a] *College of Hydrology and Water Resources, Hohai University, Xikang Road, Nanjing, Jiangsu Province 210098, China*
[b] *National Cooperative Innovation Center for Water Safety & Hydro-Science of Hohai University, Nanjing, Jiangsu Province 210098, China*
[c] *Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China*
[d] *The Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK, USA*
[e] *Hydrometeorology & Remote Sensing (HyDROS) Laboratory and School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA*

ARTICLE INFO

Article history:
Received 27 July 2015
Received in revised form
26 February 2016
Accepted 29 February 2016
Available online 28 March 2016

Keywords:
Kalman filter
K-nearest neighbor
Coupling method
State-transition matrix
Real-time updating

ABSTRACT

The Kalman filter (KF) updating method has been widely used as an efficient measure to assimilate real-time hydrological variables for reducing forecast uncertainty and providing improved forecasts. However, the accuracy of the KF relies much on the estimates of the state transition matrix and is limited due to the errors inherit from parameters and variables of the flood forecasting models. A new real-time updating approach (named KN^{2}K) is produced by coupling the *k*-nearest neighbor (KNN) procedure with the KF for flood forecasting models. The nonparametric KNN algorithm, which can be utilized to predict the response of a system on the basis of the *k* most representative predictors, is still efficient when the descriptions for input-output mapping are insufficient. In this study, the KNN procedure is used to provide more accurate estimates of the state transition matrix to extend the applicability of the KF. The updating performance of KN^{2}K is investigated in the middle reach of the Huai River based on a one-dimensional hydraulic model with the lead times ranging from 2 to 12 h. The forecasts from the KN^{2}K are compared with the observations, the original forecasts and the KF-updated forecasts. The results indicate that the KN^{2}K method, with the Nash-Sutcliffe efficiency larger than 0.85 in the 12-h-ahead forecasts, has a significant advantage in accuracy and robustness compared to the KF method. It is demonstrated that improved updating results can be obtained through the use of KNN procedure. The tests show that the KN^{2}K method can be used as an effective tool for real-time flood forecasting.

1. Introduction

Flood forecasting models are crucial for flood warning and flood management, while suffering from unavoidable uncertainties and assumptions. Many different mathematical methods have been coupled with flood forecasting models to account for these uncertainties (e.g. Bélanger & Vincent, 2005; Ren et al., 2010; Sil & Choudhury, 2015) and improve the forecasting accuracy. Real-time updating methods have been commonly used to improve forecast accuracy and may be classified into two basic categories wherein updates are made to the forecasted results or the model state. The combination of this two updating methods has also received many attentions (Lai et al., 2013; Madsen & Skotner, 2005).

In the first category of updating methods, aggressive regression (AR), neural networks (NN), *k*-nearest neighbor (KNN) nonparametric regression are commonly used. The AR method is generally used on the assumption that the variation in the time-series hydrological variables follows a gradual process and that these variables are sequentially related (Goswami et al., 2005). When utilized to solve problems near the peak of a flood or where the state of the water flow changes abruptly, the AR method may lead to the unreliable forecasts. The use of NN in real-time updating is restricted by its slow convergence, local optimization problems, and low generalization ability. Studies on applying the KNN method to machine learning and weather forecasting etc. have also yielded many beneficial findings (Karlsson & Yakowitz, 1987; Zheng & Su, 2014). The KNN method is largely distinct from the AR and NN methods as it does not need to prescribe detailed solutions to the input–output mapping. Liu et al. (2014) examined the validity of the KNN, AR, and Kalman filter (KF) methods for real-

* Corresponding author. Tel.: +86 139 1471 7604.
E-mail address: yaochhu@126.com (C. Yao).

http://dx.doi.org/10.1016/j.ijsrc.2016.02.002
1001-6279/

time updating of hydraulic model. Their results indicated that the KNN method is competent for real-time updating and performs more efficiently than the other two methods in peak flow forecasting. However, updating method in this category is generally carried out at a single, gauged station only and is not feasible for ungauged stations. This is especially the case for the reach of the Huai River in China where there are far more ungauged stations than gauged ones. Thus, the aforementioned state-updating methods are more effective as they are capable of providing updated forecasts for both gauged and ungauged stations.

The updating methods in the second category aiming at correcting model state generally involve the KF method and its derivatives. By using state updating methods, the parameters/variables in the forecasting model are to be corrected to obtain improved forecast accuracy (Xu et al., 2015). The diverse contributions made by the KF method in respect of rainfall-runoff forecasting and water quality modeling are presented by Todini (2005). Jin and Fread (1993) introduced KF into their applications for updating real-time flood forecasting models. Their results indicate that significant improvement in flood forecasting can be achieved using the KF method. The extended Kalman filter (EKF) is always used in nonlinear systems, in which the noise terms are modeled as Gaussian profiled (Anderson & Moore, 1979). When the nonlinearity is quite large, it does not likely provide satisfactory descriptions on the actual distribution of the model state and diverges easily. Ever since Evensen (1994) first described the ensemble Kalman filter (EnKF) method, the technique has been widely applied to solving problems in ocean circulation and meteorology (Evensen, 1994; Lai et al., 2013, 2014; van Loon et al., 2000). The method starts with a cluster of state vectors and propagates them through the nonlinear model to obtain the probability distribution of the state. However, there has not been a universally accepted theory on generating the initial state vectors. In case of coupling with a flood forecasting model, which is always susceptible to the variation in their model state, the state vectors randomly generated in the EnKF method can easily lead to the divergence of the model. Han and Li (2008) examined the performance of four Kalman type filtering technologies, including the unscented Kalman filter (UKF), the sampling importance resampling particle filter (SIR-PF), the unscented particle filter (UPF) and the EnKF, in predicting one-dimensional soil moisture content. In their tests, the computational cost of the EnKF was comparable to the SIR-PF, 4 times heavier than the UKF and 1.5 times heavier than the UPF. It implicates that the KF appears to be more efficient and is worth in-depth study for real-time flood forecasting.

However, in real-time updating applications, there are still some problems that may affect performance of the KF. Except for the problem in estimation of noise terms in KF, inaccurate state transition matrix, the elements in which are inherited from parameters and variables in flood forecasting models, contributes a lot to the errors of the updated forecasts. Thus, the updated model state predicted on the basis of the inaccurate state transition matrix may differ greatly from its true value (Tsuda, 2010; Weeden & Cefola, 2010). Tsuda (2011) and Ursini et al. (2014) pointed out that an accurate state-transition matrix, which is hard to be obtained in real-time applications, is the prerequisite and a major restriction of the Kalman-based prediction. When using a one-dimensional hydraulic model for flood forecasting, the estimation of the model parameters and variables is affected by many factors. Both the initial condition and the upper/lower boundary conditions are likely to be issued with unavoidable errors. Worse still, in multi-step forecasts, errors from these inaccurate components are propagated forward and amplifies significantly with the increase of the forecast lead time. This study seeks to mitigate the problems caused by inaccurate parameters/variables in the forecasting model by combining the KF method with another mathematical algorithm.

Generally, the forecasting accuracy can be improved by the combination of forecasts updating and state updating methods (Kuncheva, 2004; Moore et al., 2005). Many mathematic algorithms have been employed coupling with Kalman type or other categories of data assimilation methods to improve the capability of these methods (Lahoz et al., 2010; Zhang et al., 2009). A simple algorithm is also in need to improve the applicability of KF in real-time flood forecasting. Earlier in this section, we introduced the efficient nonparametric KNN algorithm. Sikorska et al. (2014) examined the validity of using the KNN method to evaluate the uncertainty in hydrological predictions. Their findings demonstrated that the KNN resampling technique can provide simplicity in application and lead to an efficient uncertainty assessment. The KNN method, which has the advantage of addressing the similarity matching issue, can be used to evaluate the correlation between each of the historical samples and the present one and carries out subsequent estimations utilizing the k most correlated samples. This way of dealing with non-stationary problems in flood forecasting (e.g. using strongly correlated subsets) has been demonstrated in many operational applications (Chinnarasri et al., 2003; Karlsson & Yakowitz, 1987; Sandri & Marzocchi, 2004).

Coupling the KNN and KF methods has already been attempted in the fields of statistics and traffic engineering (Wang et al., 2013; Zheng & Su, 2014). This coupled approach has been demonstrated to be accurate and efficient for statistical analysis and short-term traffic volume forecasting. Here, we consider coupling the KNN and KF methods (hereafter referred to as KN^2K) to improve the performance of the KF method when applied to real-time updating of flood forecasting models. The KN^2K updating method is characterized by the features that the state transition matrix of the KF method is recalculated using the KNN method. The validity of the KN^2K algorithm for real-time flood forecasting is tested in the Huai River basin in China. Additionally, the performance of the KN^2K is investigated and compared to that of the original KF to determine whether improved forecasts can be obtained when coupling the KNN procedure in the updating method.

The rest of this paper is organized as follows. Section 2 presents the methodologies of real-time updating (using either the KF method or the KN^2K method) to a one-dimensional hydraulic model. Section 3 provides a brief introduction to the study area and datasets involved. Section 4 summarizes performance studies of the KN^2K method compared to the performance of the original KF method with the emphasis on accuracy, robustness and computing cost. The conclusions are presented in Section 5.

2. Methodology

Two methods, including the original KF method and the proposed KN^2K method, are utilized in the real-time updating of the hydraulic model. The following is a brief description of these two methods.

2.1. The KF method used in the updating of the hydraulic model

2.1.1. The linearized hydraulic model

The one-dimensional St.-Venant equations are commonly used to describe the transformation of the water flow from upstream to downstream cross-sections. Linearized approximations of these equations can be computed using the four-point Preissmann scheme (Ettema, 2000; Novák, 2010). Two extra equations (one for the upstream inflow discharge, as the upper boundary condition, and one for the downstream stage-discharge rating curve, as the lower boundary condition) are also employed in order to produce