Journal of Hydrology 512 (2014) 540–548

Contents lists available at ScienceDirect

Journal of Hydrology

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

Improving numerical forecast accuracy with ensemble Kalman filter and chaos theory: Case study on Ciliwung river model

Yabin Sun [a,*], Chi Dung Doan [a,b], Anh Tuan Dao [a], Jiandong Liu [a], Shie-Yui Liong [a,b]

[a] Tropical Marine Science Institute, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119227, Singapore
[b] Willis Research Network, Willis Re Inc., 51 Lime Street, London EC3M 7DQ, United Kingdom

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form 7 February 2014
Accepted 6 March 2014
Available online 25 March 2014
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Michael Brian Butts, Associate Editor

Keywords:
Data assimilation
Ensemble Kalman filter
Chaos theory
Ciliwung river

SUMMARY

The classic Kalman filter implementation uses the measurements up to the time of forecast to update the initial conditions of the numerical model, with the updating effect limited to a prediction horizon when the improved initial conditions are washed out. To further enhance the prediction capability, this study proposes a new hybrid data assimilation scheme, which adopts chaos theory to predict the measurements into the forecast phase, and then assimilates the predicted measurements into the numerical model using the ensemble Kalman filter.

The hybrid data assimilation scheme is applied in a simulated real-time forecast of the Ciliwung river model. It is revealed that the hybrid scheme can further improve the modelling accuracy up to a prediction horizon of 4 days as compared to the update based solely on the ensemble Kalman filter.

1. Introduction

Numerical modelling has proved to be an efficient tool to simulate hydrologic and hydraulic processes, providing important information for catchment planning, water resource utilization, real-time monitoring and flood warning systems. Nevertheless, the numerical model becomes less competent when it comes to questions such as 'how high will the water level rise?', 'when will the water level reach its peak?' and 'where will the flooding occur?' When applied as a flow forecasting system, the numerical model uses current information of the river system together with the forecasted model forcing (e.g. rainfall) to predict the future states of the river. Therefore, the forecast accuracy is hampered by the errors in the current system information as well as the errors in the forecasted model forcing.

Data assimilation was brought forward as a feedback process to improve the model forecast accuracy incorporating the system observations (Robinson et al., 1998). The techniques can be classified according to the modified variables, i.e., input variables, model states, model parameters and output variables (WMO, 1992). A widely applied and theoretically comprehensive class of data assimilation procedure is based on the Kalman filter. The Kalman filter was originally proposed to describe a recursive solution to the linear filtering problem (Kalman, 1960), whereas its application can be extended to the non-linear dynamic system through a statistical linearization procedure (Welch and Bishop, 2001). The merits of the Kalman filter are reflected in two respects: (i) it is able to update the entire state of a modelling system based on the information from few point measurements; and (ii) it explicitly takes model and measurement uncertainties into account in the updating process. The drawback of the Kalman filter, however, is that the measurements are only available till the time of forecast, which constrains the updating effect to a limited prediction horizon when the improved initial conditions are washed out.

Time series prediction emerged as an alternative data assimilation technique. The approaches range from the traditional statistical fitting methods, such as the autoregressive moving average model (Box and Jenkins, 1976), to state-of-art soft computing methods, such as artificial neural networks (Haykin, 1999) and support vector machines (Cristianini and Shawe-Taylor, 2000). Recent development in chaos theory shows that chaotic dynamics prevails in the non-linear systems despite their random behaviours (Ott, 1993; Williams, 1997; Sprott, 2003). Upon projecting into a multi-dimensional phase space, the underlying structure of the non-linear time series can be revealed and hence approximated by means of local approximation. The local model approach based

* Corresponding author. Tel.: +65 96379168; fax: +65 67761455.
E-mail address: tmssy@nus.edu.sg (Y. Sun).

http://dx.doi.org/10.1016/j.jhydrol.2014.03.016
0022-1694/

on chaos theory has gained popularity in time series prediction with superior performance over the statistical fitting methods and the soft computing methods in a variety of applications (Babovic et al., 2005; Liong et al., 2005; Sun et al., 2009, 2010).

Madsen and Skotner (2005) proposed a hybrid data assimilation procedure based on the steady-state Kalman filter combined with error forecasting at measurement points. The model errors were propagated into the forecast period with the error forecast models, i.e., a second-order autoregressive and a harmonic error forecast model; the forecasted model errors were then distributed to the entire modelling system using a predefined, time invariant Kalman gain vector. The data assimilation procedure was applied in an operational flood forecasting setup, where the model errors were generated with an artificial constant subtraction from the downstream tidal boundary. This hybrid procedure, however, may not be applicable in a real system with more complexities, where the model errors are primarily due to the lack of information on future forcing and exhibit highly non-linear characteristics.

This study develops a new hybrid data assimilation scheme that combines the ensemble Kalman filter with chaos theory: chaos theory is used to directly predict the measurements in the forecast phase, whereas the extended measurements are melded into the numerical model using the ensemble Kalman filter. This hybrid scheme is applied in a real-time forecast simulated by the Ciliwung river model – a 1D MIKE 11 river model, and the performance is assessed up to a prediction horizon of 4 days. The configurations of the numerical model and the data assimilation scheme are elaborated; the results are presented and discussed.

2. Numerical model

2.1. MIKE 11 modelling system

MIKE 11 is a one-dimensional modelling system for the simulation of flows, water quality and sediment transport in rivers, channels, estuaries and other water bodies. Different computational modules are integrated in MIKE 11 to describe different physical processes (DHI, 2011).

A river model setup in MIKE 11 normally includes the rainfall–runoff module and the hydrodynamic module. The river catchment is divided into sub-basins based on its topography and land use, etc.; the rainfall–runoff process in each sub-basin is described by a conceptual rainfall–runoff model (Nielsen and Hansen, 1973). The produced runoff is then applied in terms of lateral inflow to the hydrodynamic module which defines how water flows in the river system.

The hydrodynamic module in MIKE 11 solves the vertically integrated equations for the conservation of continuity and momentum, i.e., the Saint Venant equations

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q \tag{1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial\left(\alpha \frac{Q^2}{A}\right)}{\partial x} + gA\frac{\partial h}{\partial x} + \frac{gQ|Q|}{C^2AR} = 0 \tag{2}$$

where Q is the discharge, x is the direction index, A is the flow area, t is the time, q indicates the lateral inflow, g is the acceleration due to gravity, h is the water level, C is the Chezy resistance coefficient, R is the hydraulic radius, and α represents the momentum distribution coefficient. The solution of the differential equations is based on a six-point implicit finite difference scheme defined in a staggered grid consisting of alternating discharge and water level points (Abbott and Minns, 1998).

The ensemble Kalman filter implementation in this study is based on the data assimilation module in MIKE 11, which can be used for assimilation of measurements in the hydrodynamic model. The detailed description can be referred to in Madsen and Skotner (2005) and DHI (2011).

2.2. Ciliwung river model

Jakarta, as the capital and largest city of Indonesia, is located on the northwest coast of Java (Fig. 1). The land area of Jakarta city is 662 km^2, whereas the total Jakarta watershed area is about 1500 km^2. Jakarta lies in a low flat basin, averaging 7 m above mean sea level. In the northern area 40% of Jakarta is below mean sea level, while the southern parts are comparatively hilly. Rivers flow from the Puncak highland to the south of the city, across the city northwards to the Java Sea. Ciliwung river is the most important among these rivers, which divides the city into western and eastern principalities. The rivers flowing across the city, combined with its low topography and abundant rainfall, make Jakarta fairly prone to flooding.

Jakarta has a hot and humid climate on the boundary between tropical monsoon (Am) and savanna (Aw) according to the Köppen climate classification system. The wet season in Jakarta covers the majority of the year, running from November through June. The remaining 4 months forms the city's dry season. As summarized in Table 1, Jakarta's annual precipitation is about 1700 mm, whereas its wet season rainfall peak is January with average monthly rainfall of 389 mm and its dry season low point is September with a monthly average of 30 mm (WMO, 2013).

Fig. 2 shows the locations of Ciliwung river and its catchment, whereas Table 2 summarizes the hydro-meteorological data used for the Ciliwung river model (Dao et al., 2012). The Ciliwung river consists of one main branch and separates into two tributaries at Manggarai with a total length of 136 km. The slope at upstream is fairly steep (3.32%), while it becomes rather flat at downstream (0.05%). The Ciliwung river catchment covers an area of 352 km^2, which is divided into 9 sub-basins for rainfall–runoff modelling. A combination of the NAM and Urban rainfall–runoff models is used to describe the hydrological process, driven by the weighted average rainfall from the 13 rainfall stations across the region. Tidal levels in the Jakarta Bay are prescribed as open boundary conditions at the river downstream. Daily discharge observations are recorded at three measurement stations: Katulampa, Ratujaya and Sugutamu.

3. Data assimilation scheme

3.1. Ensemble Kalman filter

Kalman filter is named after Rudolph E. Kalman, who in 1960 published the paper describing a recursive solution to the linear filtering problem (Kalman, 1960). Kalman filter is essentially a set of mathematical equations implementing a 'forecast' and 'analysis' estimate that produces an optimal state vector with minimum analysis error covariance (Welch and Bishop, 2001).

Ensemble Kalman filter was developed to ease the computational burdens involved in the propagation of error covariance matrix in the original Kalman filter algorithm (Evensen, 1994). The statistical properties of the state vector are represented by an ensemble of possible state vectors; the error covariance is propagated based on the Monte Carlo simulation. The algorithm of ensemble Kalman filter can be summarized as follows (Madsen and Cañizares, 1999; Evensen, 2003):

(i) An ensemble of M state vectors is propagated forward in a state-space form

$$\boldsymbol{x}^f_{i,k} = \boldsymbol{\Phi}\left(\boldsymbol{x}^a_{i,k-1}, \boldsymbol{u}_k + \boldsymbol{w}_{i,k}\right), \qquad i = 1, 2, \ldots, M, \tag{3}$$