Journal of Hydrology 438–439 (2012) 223–233

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Influence of lag time on event-based rainfall–runoff modeling using the data driven approach

Amin Talei [a,*], Lloyd H.C. Chua [b,1]

[a] DHI-NTU Water & Environment Research Centre and Education Hub, School of Civil & Environmental Engineering, Nanyang Technological University, Blk N1.2 #B1-02, 50, Nanyang Avenue, Singapore 639798, Singapore

[b] School of Civil & Environmental Engineering, Nanyang Technological University, Blk N1 #01C-70, 50, Nanyang Avenue, Singapore 639798, Singapore

A R T I C L E I N F O

Article history:
Received 6 February 2011
Received in revised form 17 December 2011
Accepted 19 March 2012
Available online 28 March 2012
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Rainfall–runoff
Event-based
Lag time
Adaptive network-based fuzzy inference system
Artificial neural networks
Data driven models

S U M M A R Y

This study investigated the effect of lag time on the performance of data-driven models, specifically the adaptive network-based fuzzy inference system (ANFIS), in event-based rainfall–runoff modeling. Rainfall and runoff data for a catchment in Singapore were chosen for this study. For the purpose of this study, lag time was determined from cross-correlation analysis of the rainfall and runoff time series. Rainfall antecedents were the only inputs of the models and direct runoff was the desired output. An ANFIS model with three sub-models defined based on three different ranges of lag times was developed. The performance of the sub-models was compared with previously developed ANFIS models and the physically-based Storm Water Management Model (SWMM). The ANFIS sub-models gave significantly superior results in terms of the *RMSE*, r^2, *CE* and the prediction of the peak discharge, compared to other ANFIS models where the lag time was not considered. In addition, the ANFIS sub-models provided results that were comparable with results from SWMM. It is thus concluded that the lag time plays an important role in the selection of events for training and testing of data-driven models in event-based rainfall–runoff modeling.

1. Introduction

The use of data-driven techniques to model the rainfall–runoff (R–R) process has become popular over the last decade. Models such as the artificial neural network, commonly referred to as a black box model apply a direct mapping between rainfall and runoff, without the need for a detailed consideration of the physical processes. Recently, the research focus on the application of data driven modeling in hydrology has shifted from a black box approach to a semantic-based fuzzy neural architecture (Yu and Yang, 2000; Xiong et al., 2001; Nayak et al., 2004, 2005; Aqil et al., 2007; Talei et al., 2010a). A Neuro Fuzzy System (NFS) is a fuzzy system that takes advantage of the learning ability of neural networks and the reasoning ability of fuzzy systems (Cho et al., 2009). Fuzzy models that assume local model presentations with local function dynamics at the consequent or rule-layer of the models are known as Takagi–Sugeno–Kang (TSK) models. The adaptive network-based fuzzy inference system (ANFIS) is an example of TSK-models where global parameter tuning is implemented by minimizing the global error of the model (Jang, 1993). There have so far been a growing number of ANFIS applications in R–R modeling (Nayak et al., 2005; Vernieuwe et al., 2005; Chen et al., 2006; Aqil et al., 2007; Mukerji et al., 2009; Talei et al., 2010a, 2010b).

Rainfall–runoff modeling can be carried out either on a continuous basis or event-based. Most studies in the literature have focused on continuous modeling; however, event-based analysis is also important for realistic short term predictions for flood forecasting applications in small catchments (Tayfur and Singh, 2006). In event-based analysis, hydrograph characteristics are important considerations when choosing training events to be used for training. This is because data driven models function by associating the time series of outputs to inputs and hence factors such as hydrograph shape and time lag are expected to be important. Talei et al. (2010a) analyzed the same data used in the present study to investigate the effect of hydrograph shape in training event selection. The events were separated into three groups, according to the hydrograph shape (Fig. 1). The authors showed that events with one major peak in the hydrograph (Group1 and Group 2) were better suited as training events compared to events with multiple peaks (Group 3). This study is an extension of the

* Corresponding author. Tel.: +65 90276610.
E-mail addresses: atalei@ntu.edu.sg (A. Talei), chcchua@ntu.edu.sg (L.H.C. Chua).
[1] Tel.: +65 6790 5249.

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2012.03.027

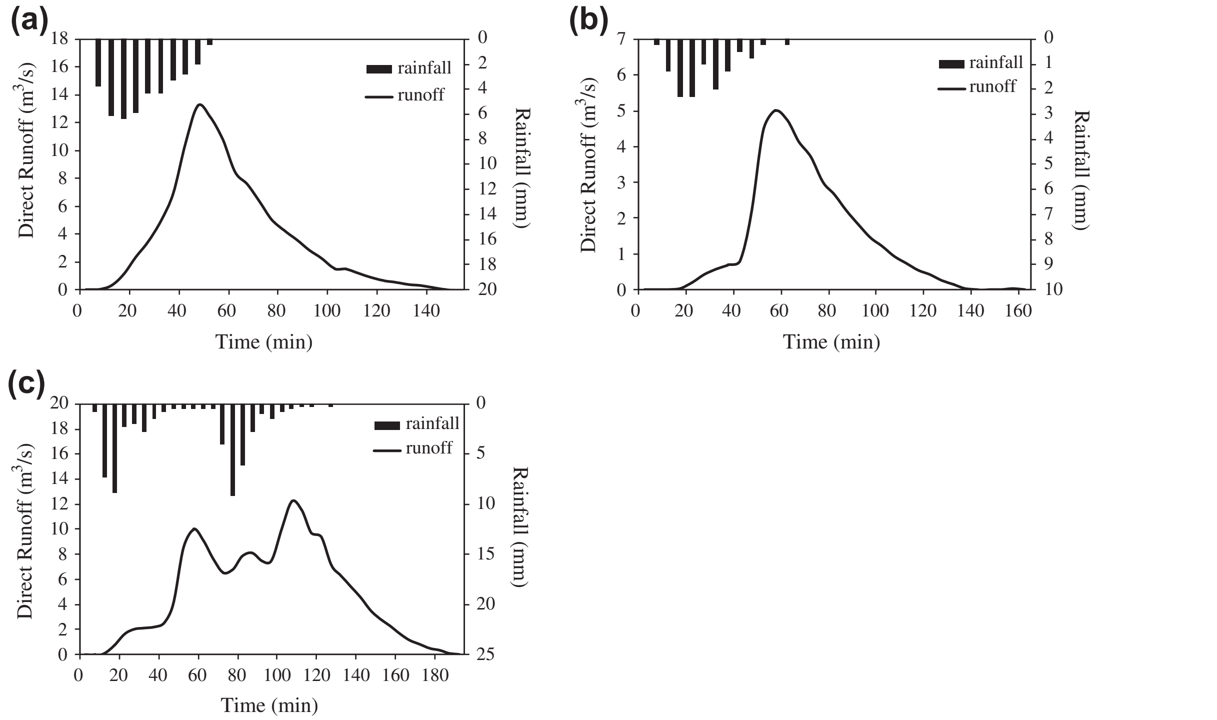

Fig. 1. Typical time series of rainfall and discharge for (a) Group 1, (b) Group 2, and (c) Group 3 events discussed in Talei et al. (2010a).

paper by Talei et al. (2010a) as it considers the lag time as an additional measure in the selection of training events. Talei et al. (2010b) also used an ANFIS model to forecast the runoff from an artificial catchment comprised of a 25 m^2 rectangular plane. It was not obvious that lag time would have a significant impact on event selection in Talei et al.'s (2010b) study, since the area of the rectangular plane was very small and there was very little variability of the lag times between the recorded events. However, it is expected that for a real catchment, lag time would be a significant factor in event selection. It should be added that the influence of hydrograph shape and lag time are masked out when the continuous modeling approach is adopted, since these factors are integrated when many events are considered at the same time, as is the case with modeling on a continuous basis. Therefore, the objective of this study was to investigate the influence of lag time on the performance of data-driven models in event-based R–R modeling. The ANFIS model was chosen for this study because it is fast gaining popularity among researchers involved in hydrological modeling with the data driven approach. However, the findings are expected to be equally applicable to other data driven models, such as the neural network. To the best of our knowledge, our study is the first to demonstrate the lag effect for event-based modeling using data driven models.

2. Review

A brief review of the works where the ANN and ANFIS models have been employed in event-based analysis, with an emphasis on the methodologies adopted in selection of training events is summarized here. The methodology for the selection of training events is highlighted since despite its potential importance on the performance of data driven models, the lag time has never been considered. Halff et al. (1993) proposed a three layered feed forward neural network for event-based R–R modeling in Bellevue, Washington and adopted a cross validation approach to swap the testing event among the available events to find the best set of events for training. The neural network was found to be successful in modeling the general hydrograph shape. Cheng and Noguchi (1996) developed a Back-Propagation Neural Network (BPNN) to model R–R process in a catchment in Ziyang, Sichuan province, China. Rainfall and initial soil moisture deficit were taken as inputs and discharge as output. In this study events with the first and second highest peak discharge were chosen for training and the remaining events for testing. BPNN was found to give better results comparing to the physically-based Xinanjiang model. Jain and Prasad Indurthy (2003) employed a BPNN model for an event-based R–R modeling in Salado Creek, Bitters Road, San Antonio and compared it with two unit hydrograph and four regression models. The authors chose the events with highest peak discharge for training and the remaining for testing. The 4-layered BPNN was found to give the best performance compared to the other models. Tayfur and Singh (2006) compared the ability of BPNN, fuzzy logic (FL) model, and a Kinematic Wave Approximation (KWA) model. Events with different values of rainfall intensities and peak flow rates were chosen for training the models. BPNN and FL models were found to be comparable to KWA; however, they required longer historical data comparing to KWA. Sohail et al. (2008) analyzed 14 years of data for two small sub-basins in Tono, central Japan and developed a BPNN for event-based R–R modeling. Events within the first 9 (1990–2000) years were chosen as training events while events from 1999 to 2000 and 2001 to 2003 were chosen as verification and testing datasets, respectively. In general, the BPNN model was found to be superior compared to an Auto-Regressive Moving Average (ARMA) model. Chua et al. (2008) developed an ANN model and compared it with kinematic wave model (KWM) to simulate event-based R–R process on an experimental overland plane surface. Both peak discharge (low, medium, and large) and hydrograph shape (single, double, and multiple peak events) were considered in the selection of training events. It was concluded that ANN models which contain both rainfall and discharge in their inputs outperformed the KWM while ANN