Environmental Modelling & Software 85 (2016) 112–128

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Comparing single- and two-segment statistical models with a conceptual rainfall-runoff model for river streamflow prediction during typhoons

Chih-Chiang Wei

Department of Marine Environmental Informatics, National Taiwan Ocean University, No.2, Beining Rd., Jhongjheng District, Keelung City 20224, Taiwan

ARTICLE INFO

Article history:
Received 18 March 2015
Received in revised form
19 July 2016
Accepted 19 August 2016
Available online 27 August 2016

Keywords:
Streamflow
Statistical model
Conceptual model
Prediction
Typhoon

ABSTRACT

This study examined various regression-based techniques and an artificial neural network used for streamflow forecasting during typhoons. A flow hydrograph was decomposed into two segments, rising and falling limbs, and the individual segments were modeled using statistical techniques. In addition, a conceptual rainfall–runoff model, namely the Public Works Research Institute (PWRI)-distributed hydrological model, and statistical models were compared. The study area was the Tsengwen Reservoir watershed in Southern Taiwan. The data used in this study comprised the observed watershed rainfalls, reservoir inflows, typhoon characteristics, and ground weather data. The forecast horizons ranged from 1 to 12 h. A series of assessments, including statistical analyses and simulations, was conducted. According to the improvements in errors, among single-segment statistical models, the multilayer perceptron achieved superior prediction accurary compared with the regression-based methods. However, the pace regression was the most favorable according to an evaluation of model complexity and accuracy. To examine the robustness of the results for forecast horizons varying from 1 to 12 h, statistical significance tests were performed for the single- and two-segment models. The prediction ability of the two-segment models was superior to that of the single-segment models. In addition, Typhoon Sinlaku in 2008 was considered in a comparison between the conceptual PWRI model output and that of the developed statistical models. The results showed that the PWRI model yielded the least favorable results.

1. Introduction

Accurate streamflow forecasts are a crucial component of watershed planning and sustainable water resource management (Besaw et al., 2007). Flooding is the most frequent natural disaster and causes heavy losses of life and property worldwide. In Taiwan, tropical storms often result in disastrous floods because of steep terrain and heavy rainfall (Hsu et al., 2010). The water flow in streams in mountainous watersheds can be rapid, and the time of concentration is approximately 1–4 h. The short time poses severe challenges for flood forecasting and reservoir operation during typhoons (Pan et al., 2013). Simple, fast, and useful prediction methods enabling accurate streamflow estimation under the hydrological conditions of Taiwan are therefore necessary (Wei, 2012; Wei and Roan, 2012).

In past years, artificial neural networks (ANNs) have been applied in hydrological modeling and have exhibited high potential for application in rainfall-runoff modeling, flood forecasting, and precipitation estimation (Beh et al., 2014; Chen et al., 2014; Cheng et al., 2014; Hutton and Kapelan, 2015; Jakeman et al., 2006; Karri et al., 2014; Li et al., 2013, 2014a; Surridge et al., 2014; Wang et al., 2013, 2014). ANNs learn complex and nonlinear relationships that are difficult to model using conventional techniques. In most of the hydrological modeling applications, multilayer perceptrons (MLPs) have been used in the model architecture (Chau, 2007; Chen and Chau, 2006; Li et al., 2014b; Maier et al., 2010; Muttil and Chau, 2006; Taormina and Chau, 2015; Wu and Chau, 2013; Wu et al., 2008, 2014). However, ANNs exhibit several disadvantages. The network structures are difficult to determine and are usually determined using a trial and error approach (e.g., sensitivity analysis; Kisi, 2010).

Regression-based algorithms are commonly used methods in water resource management. The basic concept of regression analysis is to fit a linear model to a set of data. The most frequently used approach is the ordinary least squares (OLS) subset selection

E-mail address: ccwei@ntou.edu.tw.

http://dx.doi.org/10.1016/j.envsoft.2016.08.013

(Wang and Witten, 1999). The classical OLS estimator is simple, computationally inexpensive, and has a widely established theoretical justification (Wang, 2000). For example, Kisi (2004) compared ANN results with those of autoregressive models (ARs) and determined that ANNs performed more favorably than ARs in monthly streamflow forecasting. Chokmani et al. (2008) compared the performance of ANNs and regression models in estimating river streamflow affected by ice conditions. Wei (2012) compared the performance of support vector regressions with that of OLS regression in forecasting downstream water levels. Wei (2015) compared the performance of lazy learning, including locally weighted regression and k-nearest neighbor, and eager learning, including ANNs, support vector regression, and OLS regression, in river stage predictions. However, Wei (2012, 2015) constructed a streamflow prediction model by using an entire flow hydrograph, consequently neglecting the different physical processes (e.g., rising and falling limbs) occurring in a drainage system, which are usually represented by the runoff response.

Numerous advanced regression-based models have been developed. For example, the pace approach proposed by Wang (2000), which is based on a methodology that resembles empirical Bayes estimator. Gupta (2012) reported the evaluation results of a proposed approach for predicting the number of zombies in distributed denial-of-service by using the pace regression model. In computing, distributed denial-of-service attack is an attempt to make a machine or network resource unavailable to its intended users. Pace regression is a type of linear regression analysis that has been shown to outperform other types of linear model-fitting method, particularly when the number of features is high and several of them are mutually dependent (Wang and Witten, 1999). Pace regression contains a type of feature selection; therefore, not all features are used in the resulting models. Additional regression-based models, such as isotonic regression and additive regression, have been developed (Section 2). Isotonic regression is a simple and useful tool and enables estimating parameters for any distributions, incorporating information about order relationships among the parameters (Nagatsuka et al., 2012). Isotonic regression is most frequently used in making inferences regarding ordered parameters. Recently, isotonic regression has received renewed attention (Guyader et al., 2014; Keshvari and Kuosmanen, 2013; and Piegorsch et al., 2014). Additive regression was suggested by Friedman and Stutzle (1981). Buja et al. (1989) proposed a backfitting algorithm for estimating an additive model and studied its properties. Stone (1985), Burman (1988), and Mallows (1986) provided more details on the additive model. However, we determined that the use of these regression techniques in streamflow prediction has not been investigated.

The purpose of this study was to examine software-based computing techniques, which refer to various regression predictors. We investigated the OLS, pace, isotonic, and additive regression techniques and compared them by using MLP ANNs. First, this study examined single-segment statistical models, and the objectives are summarized as follows:

- To assess the prediction ability of various regressions and ANNs, the effects of multisource data with long lag times on streamflow predictions were investigated. For a river basin system, streamflow prediction can comprise a complex combination of various hydrometeorological factors. To achieve accurate streamflow predictions, this study collected data comprising hydrometeorological attributes, namely observed watershed rainfalls, reservoir inflows, typhoon characteristics, and ground weather data.
- To determine the appropriate number of time-lagged input data, the dimensionality determination problem (formally known as the model selection or subset selection problem) was addressed. As indicated by Wang (2000), numerous researchers have investigated methods for subset selection to determine the number of parameters that should be used in a final estimated model. This study adopted the conventional correlation-based criterion and stepwise selection methods to evaluate the inputs of various models.
- To evaluate the complexity of various models, the Akaike information criterion (AIC) was used to calibrate the tradeoff between the goodness of fit and the complexity of the models. This study determined whether the use of various regressions and the MLP ANN can be justified in river streamflow predictions.

Moreover, this study decomposed a flow hydrograph into two segments (i.e., rising and falling limbs), because the runoff response of a drainage system, represented in the different segments of a flow hydrograph, is produced by different physical processes occurring in the system. As indicated by Jain and Srinivasulu (2006), the rising limb of a flow hydrograph represents the gradual release of water from various catchment storage elements caused by gradual repletion of the storage elements when the drainage system receives rainfall input. The characteristics of the rising limb of a flow hydrograph, such as the size, shape, and slope, are influenced by varying infiltration capacities, drainage storage characteristics, and the nature of the input, namely the intensity and duration of the rainfall. However, the falling limb (or recession limb) of a flow hydrograph is the result of the gradual release of water from the drainage system after the rainfall input has stopped and is influenced more by the storage characteristics of the drainage system and climatic characteristics. In this study, to examine the robustness of the results regarding forecast horizons, statistical significance tests were performed for single- and two-segment statistical models.

This study also compared the aforementioned single- and two-segment statistical models with a conceptual rainfall–runoff model. The conceptual physical approach entails using the fundamental laws of physics to represent and explain the hydrological processes governing the behavior of the studied hydrosystem (Hingray et al., 2014). To simulate typhoon river floods by using the conceptual rainfall–runoff model, we employed an integrated hydrological simulation system, namely Integrated Flood Analysis System (IFAS), which was developed by the International Centre for Water Hazard and Risk Management (Fukami et al., 2009). The IFAS has been practically applied to past flood events in Asian countries such as Japan (Sugiura et al., 2008) and Pakistan (Aziz and Tanaka, 2011). A conceptual, distributed rainfall–runoff analysis engine, the Public Works Research Institute (PWRI)-distributed hydrological model (Yoshino et al., 1990), is employed in the IFAS. The performance of the aforementioned statistical models and PWRI model in predicting typhoon floods (Typhoon Sinlaku in 2008) at the Tsengwen Reservoir watershed in Southern Taiwan was compared.

The remainder of this paper is organized as follows: Section 2 introduces the theorem for the four regression-based models and the MLP ANN. Section 3 describes the experimental area and recorded typhoon events. Section 4 presents the proposed methodology for streamflow prediction modeling, the input parameters for the studied case, and the model performance levels. Section 5 provides an evaluation of single-segment statistical models. Section 6 presents the advanced two-segment statistical models and an examination of the statistical significance for single- and two-segment models. Section 7 describes the conceptual PWRI rainfall–runoff model and comparisons with statistical models. Finally, Section 8 presents the conclusion.