Journal of Hydrology 521 (2015) 302–313

Contents lists available at ScienceDirect

Journal of Hydrology

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

A real-time forecasting model for the spatial distribution of typhoon rainfall

Gwo-Fong Lin *, Bing-Chen Jhong

Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan

ARTICLE INFO

Article history:
Received 11 July 2014
Received in revised form 2 December 2014
Accepted 5 December 2014
Available online 13 December 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ashish Sharma, Associate Editor

Keywords:
Hourly typhoon rainfall forecasting
Spatial characteristics
Meteorological data
Multi-objective genetic algorithm
Support vector machine

SUMMARY

Accurate forecasts of hourly rainfall are necessary for early warning systems during typhoons. In this paper, a typhoon rainfall forecasting model is proposed to yield 1- to 6-h ahead forecasts of hourly rainfall. First, an input optimization step integrating multi-objective genetic algorithm (MOGA) with support vector machine (SVM) is developed to identify the optimal input combinations. Second, based on the results of the first step, the forecasted rainfall from each station is used to obtain the spatial characteristics of the rainfall process is presented. An actual application to Tsengwen river basin is conducted to demonstrate the advantage of the proposed model. The results clearly indicate that the proposed model effectively improves the forecasting performance and decreases the negative impact of increasing forecast lead time. The optimal input combinations can be obtained from the proposed model for different stations with different geographical conditions. In addition, the proposed model is capable of producing more acceptable the results of rainfall maps than other model. In conclusion, the proposed modeling technique is useful to improve the hourly typhoon rainfall forecasting and is expected to be helpful to support disaster warning systems.

1. Introduction

Rainfall forecasting is one of the most important issues in hydrologic research since early warnings of severe weather can help prevent damages and life-threatening casualties caused by serious natural disasters. Accurate and effective forecasts of hourly rainfall are crucial for hourly reservoir inflow forecasting, flooding prevention and making important reservoir operation. Many previous studies have been conducted on rainfall forecasting using a variety of techniques, such as numerical weather prediction models (Nunes and Cocke, 2004; Diomede et al., 2008; Boniface et al., 2009; He et al., 2013; Ushiyama et al., 2014), quantitative precipitation forecast (Grecu and Krajewski, 2000; Ganguly and Bras, 2003; Ramirez et al., 2005; Valverde et al., 2014), and the applications of radar and satellite precipitation products (Grecu and Krajewski, 2000; Sheng et al., 2006; Liu et al., 2008; Moreno et al., 2012; Chen et al., 2013b; Chang et al., 2014).

The heavy rainfall caused by typhoons frequently result in serious disasters. Obviously, the typhoon rainfall forecasting influences disaster mitigation and emergency operations. Effective forecasts of hourly rainfall during typhoons are needed for issuing flood warnings. Therefore, to improve the accuracy of typhoon rainfall forecasts is always an important task of flood management. However, typhoon rainfall is a highly nonlinear and extremely complex physical process. A physically-based mathematical model is difficult to be developed for typhoon rainfall forecasting because of its tremendous variability over a wide range of space and time scales. It involves many complicated variables which are interconnected, and the volume of rainfall calculation require sophisticated mathematical tool (Luk et al., 2001; Nasseri et al., 2008). Hence, it is not a feasible alternative in most cases.

An alternative to the physically-based models is artificial neural networks (ANNs), which is a kind of information processing system with good flexibility in modeling nonlinear processes. The ASCE Task Committee (2000a,b) and Maier and Dandy (2000) have presented comprehensive reviews of the applications of ANNs in hydrology. ANN-based models have been proposed for reservoir operation (e.g., Deka and Chandramouli, 2009; Wang et al., 2010), reservoir inflow forecasting (e.g., Lin et al., 2009a, 2010), flood forecasting (e.g., Chen et al., 2013a; Lin et al., 2013a; Pan et al., 2013; Lohani et al., 2014), and rainfall forecasting (e.g., Lin and Chen, 2005; Lin and Wu, 2009; Lin et al., 2009b; Srivastava et al., 2010; Mekanik et al., 2013; Chang et al., 2014). Hourly rainfall forecasting during typhoons has drawn attention in recent years. For example, Lin and Chen (2005) used an ANN model to forecast 1-h ahead forecasts of typhoon rainfall. Lin and Wu (2009) proposed a hybrid neural network model which combines

* Corresponding author. Tel.: +886 2 3366 4368; fax: +886 2 2363 1558.
E-mail address: gflin@ntu.edu.tw (G.-F. Lin).

http://dx.doi.org/10.1016/j.jhydrol.2014.12.009
0022-1694/

the self-organizing map (SOM) with the multilayer perceptron network (MLPN) to improve the accuracy of hourly typhoon-rainfall forecasts. In addition, Lin et al. (2009b) constructed a support vector machine (SVM)-based model with typhoon characteristics to yield 1- to 6-h ahead forecasts of hourly typhoon rainfall.

Furthermore, one of the most important steps in the ANN modeling of the rainfall forecasting is the determination of significant input variables. In general, not all potential variables are used as input to the model because some may be noisy or have no significant relationship with the output variable. Hence, to determine significant input variables plays an important role in the ANN modeling process. More recently, an optimization method called multi-objective genetic algorithm (MOGA) has increasingly been applied. The capabilities of MOGAs to explore and discover Pareto-optimal fronts on multi-objective optimization problems have been well recognized (Deb et al., 2002; Liu, 2009). Prasad and Park (2004) presented a multi-objective genetic algorithm approach to the design of a water distribution network. Reddy and Kumar (2006) derived a set of optimal operation policies for a multipurpose reservoir system using multi-objective evolutionary algorithm. Reed and Minsker (2004) demonstrated the use of high-order Pareto optimization on a long-term groundwater monitoring application. Several hybrid methods, especially a combination between a MOGA and a SVM, have been implemented to optimize parameters in many fields (e.g. Giustolisi, 2006; Wu et al., 2009; Zhang et al., 2010). Giustolisi (2006) employed a MOGA to construct an optimal SVM. A support vector machine's performance depends on the kernel parameter and input selection, and these are used as decision variables for the evolutionary strategy based on a MOGA. Lin et al. (2013b) proposed a model integrating MOGAs and SVMs to improve hourly typhoon rainfall forecasting on a reservoir scale for only two point-scale ground rain gauges. In this paper, 20 meteorological and rainfall stations are used on a catchment scale. Moreover, the main difference between the previous work and this paper is that the spatial characteristics of the rainfall process are also discussed to demonstrate the performance of the proposed model.

The objective of this paper is to develop a typhoon rainfall forecasting model to yield 1- to 6-h ahead forecasts of hourly rainfall. The spatial characteristics of the rainfall process are also presented. An actual application to Tsengwen river basin is conducted to demonstrate the superiority of the proposed model, and the accuracy of the proposed model is discussed in depth to demonstrate its superiority. The paper is organized as follows. The details of the proposed forecasting technique are presented in Section 2. Section 3 provides the description of the study area and data. Section 4 shows the results, including the performance of the proposed model in the upstream and downstream regions, the optimal input combinations, and the spatial characteristics of the rainfall process for hourly typhoon rainfall forecasting. Finally, the summary and conclusions are given in Section 5.

2. The proposed forecasting technique

The input optimization that SVM is integrated with MOGA is to identify the optimal input combinations for decreasing the relatively irrelevant input information. Then, according to the results derived from models, the forecasted rainfall from each station is used to obtain the spatial characteristics of the rainfall process is presented. Details of the proposed forecasting technique are described as follows.

2.1. Input optimization

Because rainfall and some meteorological data are used as input to models, how to determine the lag lengths of input variables is very important. Thus, the training procedure by using several types of optimization methods is considered. In multi-objective genetic algorithms (MOGAs), the domination concept is applied to determine the Pareto-optimal solutions. In the domination concept, a comparison is made to determine whether one solution dominates the other or not. A set of Pareto-optimal solutions in a single run of the algorithm can be captured by MOGAs. Hence, MOGAs are especially appropriate to solve the multi-objective nonlinear optimization problems.

2.1.1. Objective functions and the fitness values

The evaluation of the fitness value of a chromosome is based on the objective function. The fitness value of a chromosome means the level of the goodness of the chromosome with respect to the training. Nonetheless, the better chromosome should be expressed by a larger fitness value, and the fitness values should be nonnegative (Goldberg, 1989). To follow these two conditions, the values of the objective functions should be transformed properly into the fitness values. In this paper, two frequently used objective functions are adopted for the training of the proposed model. One is the mean absolute error (MAE):

$$\mathrm{MAE}(\varphi) = \frac{1}{N_e}\sum_{i=1}^{N_e}\left[\frac{1}{n}\sum_{t=1}^{n}\left|\widehat{R}_t(\varphi) - R_t\right|\right]_i \tag{1}$$

where R_t and $\widehat{R}_t(\varphi)$ are respectively the observed and forecasted rainfall at time t, n is the number of forecasts, φ is the set of model inputs to be trained, and N_e is the number of typhoon events. The MAE is nonnegative. For the MAE objective function, the set of model inputs that minimizes the MAE is the optimal solution.

The other objective function employed to train the proposed model is the efficiency coefficient (EC):

$$\mathrm{EC}(\varphi) = \frac{1}{N_e}\sum_{i=1}^{N_e}\left\{1 - \frac{\sum_{t=1}^{n}\left[R_t - \widehat{R}_t(\varphi)\right]^2}{\sum_{t=1}^{n}\left(R_t - \overline{R}\right)^2}\right\}_i \tag{2}$$

where $\overline{R}$ is the average of the observed rainfall. A larger EC represents the corresponding inputs set is a better one and otherwise the inputs set is not appropriate. The model inputs that maximize the EC are the optimal solution when the EC is employed as the objective function.

The domination concept can be applied to determine the better solution between any two solutions. Most MOGAs use the domination concept to find the Pareto-optimal solutions. A solution Y_1 is said to dominate the other solution Y_2, if the following conditions (a) and (b) are satisfied (Deb, 2001): (a) the solution Y_1 is no worse than Y_2 in all objective functions, and (b) the solution Y_1 is strictly better than Y_2 for at least one objective. If one of the above conditions is not satisfied, the solution Y_1 does not dominate the solution Y_2. Most MOGAs can find all of the non-dominated solutions that are also called the Pareto-optimal solution by using the domination concept. In this paper, the objective 1 is MAE, and the objective 2 is EC.

Generalized Pareto-based scale-independent fitness function (GPSIFF), which is a simple way to determine the Pareto-optimal solution, is used herein. GPSIFF evaluates the domination of each solution by a score function. The score value of a solution Y is calculated according to the following score function:

$$\text{score }(Y) = l - m + c \tag{3}$$

where l is the number of solutions dominated by Y, and m is the number of solutions which dominate Y. The scaling constant c is used to obtain a positive fitness value. The GPSIFF concept is employed to determine the Pareto-optimal solution for the determination of model inputs. The values of the objective functions