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., Zhongzheng 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 accuracy 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.

© 2016 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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.
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 Stutze (1981), Buja et al. (1989) proposed a back-fitting 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 hydroystem (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.
2. Algorithms

Numerous models, namely the OLS, pace, isotonic, and additive regressions and an MLP ANN, were used in this study. These algorithms are briefly reviewed in the subsequent sections.

2.1. OLS regression

The classic OLS linear regression, also known as least squared errors regression, is one of the most basic and most commonly used prediction techniques. In general, a linear regression model can be written as (Chen and Jackson, 2000)

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_K X_{iK} + \epsilon_i \]  

where \( Y_i \) is the \( i \)th observation in the response variable (i.e., dependent variable) in which \( i = 1, \ldots, N \) and \( N \) is the sample size; \( X_{i1}, \ldots, X_{iK} \) is the \( i \)th observation measured of \( K \) explanatory variables (independent variables); \( \beta_0, \beta_1, \ldots, \beta_K \) are the parameters to be estimated; and \( \epsilon_i \) is an error term. Model parameters are usually estimated using the least squares method by minimizing the residual sum of squares; that is

\[ \text{minimize} \sum_{i=1}^{N} (Y_i - b_0 - b_1 X_{i1} - \cdots - b_K X_{iK})^2 \]  

where \( b_0, b_1, \ldots, b_K \) are the least squares estimates of \( \beta_0, \beta_1, \ldots, \beta_K \), respectively. If the expected value of \( \epsilon_i \), \( \text{E}(\epsilon_i) = 0 \) and the independent variables are free of errors, \( b_0, b_1, \ldots, b_K \) are the unbiased least squares estimates of \( \beta_0, \beta_1, \ldots, \beta_K \). If errors follow the Gaussian-Markov conditions (Sen and Srivastava, 1990; i.e., \( \epsilon_i \sim N(0, \sigma^2) \) and \( \epsilon_i \epsilon_j = 0 \) when \( i \neq j \)), the variances for \( \beta_0, \beta_1, \ldots, \beta_K \) can be estimated and tested parametrically against the normal distribution.

2.2. Pace regression

Wang (2000) proposed the pace regression approach for fitting linear models. The concepts underlying pace regression are based on the empirical Bayes methodology of Robbins (1964). A maximum likelihood estimation (MLE) asymptotic normality property is used to transform the original parameters into dummy parameters. A nonparametric mixture estimate of the observed values of these dummy parameters is formed, and finally, an empirical Bayes analysis for minimizing the Kullback-Leibler distance is applied. The empirical Bayes methodology is briefly reviewed in the following. Given independent samples \( x_1, \ldots, x_k \) from distributions \( F(x; \theta_i) \), where values of \( \theta_i \) may be completely different from each other, it is known that the MLE obtained from the joint distribution \( F(x; \theta) \) is a vector, with each entry being a univariate MLE; for example, if \( F(x; \theta_i) \) is the normal distribution with the mean \( \theta_i \) then \( \theta = x \). However, the MLE estimator is inferior to the empirical Bayes estimator:

\[ \hat{\theta}_k^{EB} = \frac{\int f(x; \theta) \xi G_k(\theta) \theta^2}{\int f(x; \theta) \xi G_k(\theta)} \]  

where \( f(x; \theta_i) \) denoting the probability density function corresponding to \( F(x; \theta_i) \) is inferior in that it does not minimize the expected squared error \( \text{E}_f(\theta|X) \) with respect to the estimator \( \hat{\theta}(x) \), where \( \theta_1, \ldots, \theta_K \) are independent and identically distributed from \( G(\theta) \). Here, \( G \) is the mixing distribution of the mixture

\[ f_k(x) = \int f(x; \theta) \xi G_k(\theta) \], and \( G_k \) is a consistent estimator of \( G \) given the mixture sample \( x \). Robbins (1964) showed that \( \theta_k^{EB} \) minimizes the Bayes risk as \( k \to \infty \) and hence is asymptotically optimal. A proof demonstrating how to build models that predict probability optimally in the sense of minimizing the \( \Delta_k \) is provided in Wang (2000) and Wang and Witten (2002).

2.3. Isotonic regression

Isotonic regression is a prominent type of nonparametric regression. In the following, we briefly describe isotonic regression when there are simple order relations among the parameters. Consider \( p \) populations with \( \mu_i \) denoting a scalar parameter of interest for group \( i, i = 1, 2, \ldots, p \). It is assumed that there is simple order among \( \mu_i \) such as \( \mu_1 \geq \cdots \geq \mu_p \). Let \( \hat{\mu}_i \) denote an estimator of \( \mu_i \), for \( i = 1, 2, \ldots, p \) and \( \mu = (\mu_1, \ldots, \mu_p) \). To satisfy \( \mu_1 \geq \cdots \geq \mu_p \), the isotonic regression of \( \hat{\mu} \), denoted by \( \vec{\mu} \), is provided by (Nagatsuka et al., 2012)

\[ \vec{\mu} = \arg \min_{\mu} \sum_{i=1}^{p} \left( \hat{\mu}_i - \mu_i \right)^2 w_i \], subject to \( \mu_1 \geq \cdots \geq \mu_p \)
2.5. Multilayer perceptron neural network

ANNs are mathematical models of the human brain designed to exploit the massively parallel local processing and distributed storage properties believed to exist in the human brain. An ANN is a highly interconnected network of many simple processing units called neurons. The neurons in an input layer receive the input from an external source and transmit the input to a neuron in an adjacent layer, which can be either a hidden layer or an output layer (Coad et al., 2014; Jain and Srinivasulu, 2006; Wei, 2013, 2014). A feedforward backpropagation ANN, such as an MLP, uses processing units placed in input, hidden, and output layers. Each unit in a layer is connected to the units in adjacent layers with an associated weight (Cheng et al., 2005; Maier and Dandy, 2000; Wei et al., 2014). Mathematically, a three-layer ANN with \( N_1 \) input nodes, \( N_2 \) hidden nodes, and \( N_3 \) output nodes can be expressed as

\[
y_\zeta = f_2 \left( \sum_{j=0}^{N_2} w_{1j}^2 \cdot f_1 \left( \sum_{i=0}^{N_1} w_{ij}^1 \cdot x_i \right) \right) \quad \zeta \in \{1, N_3\} \tag{6}
\]

where \( \zeta \) is the index of the input nodes; \( \psi \) is the index of the hidden nodes; \( \zeta \) is the index of the output nodes; \( x_i \) is the input node in the input layer; \( w_{1j}^1 \) is the weight set connecting the input and hidden layers; \( w_{1j}^2 \) is the weight set connecting the hidden and output layers; \( y_\zeta \) is the network output; \( f_1(\cdot) \) is the activity function of the hidden layer; and \( f_2(\cdot) \) is the activity function of the output layer.

A gradient descent procedure known as generalized error-backpropagation is typically employed to train MLP networks; therefore, the MLP network is also known as a backpropagation network. To construct MLP networks, the parameters of the learning rate, momentum, and number of nodes in the hidden layer were chosen according to a sensitivity analysis in this study.

In neural networks, the activation and output functions of the input and output layers may be of different types. In particular, linear functions are frequently used for inputs and outputs, and nonlinear transfer functions are used for hidden layers (Govindaraju and Rao, 2000). Various activity functions such as linear, sigmoid, and hyperbolic tangent functions can be used. The sigmoid function is the most common form of activity function (Cybenko, 1989; Hornik et al., 1989). According to the survey by Duch and Jankowski (1999), neural networks with a single hidden layer using sigmoidal functions are universal approximators; that is, they can approximate an arbitrary continuous function on a compact domain with arbitrary precision given a sufficient number of neurons. Thus, the sigmoid and linear activity functions were adopted in the hidden layer and output layer, respectively, in the current study. Moreover, the parameters of hidden nodes, the learning rate, and momentum were calibrated. The parameters were subjected to sensitivity analysis by using a testing data set. During network training, the weights were iteratively altered among the neurons until the output signal matches the target output within a desired minimal error range.

3. Study site and materials

The Tsengwen Reservoir watershed is in the upstream section of the Tsengwen creek and covers an area of 481 km\(^2\), with a mean annual precipitation of approximately 2700 mm and a mean annual stream flow of 29.0 m\(^3\) s\(^{-1}\) (Pan et al., 2013). The Tsengwen Reservoir is located downstream from the watershed and situated at an altitude of 133 m. The topography and locations of hydrological and rain gauge stations in the Tsengwen Reservoir watershed are shown in Fig. 1. Average annual precipitation in this area is approximately 3000 mm, more than 80% of which occurs between June and September. The annual runoff is approximately 120 x 10\(^6\) m\(^3\) with 85% occurring between June and September (Lee, 2012).

3.1. Data

This study analyzed 20 typhoon events that affected the Tsengwen River watershed between 2004 and 2011 (Table 1). The Water Resource Agency (WRA) and Central Weather Bureau (CWB) in Taiwan supplied hydrometeorological hourly data on the study area. The data were divided into three types (Table 2), watershed hydrology (\(H\)), typhoon information (\(Y\)), and weather properties (\(W\)). Table 2 shows the mean, minimal, and maximal values of various attributes.

First, the watershed hydrological data were the reservoir inflow (denoted as \(H_1\)) and rainfall in the reservoir watershed (\(H_2\)) and were collected from the WRA and CWB, respectively. The five rainfall gauges in the watershed were the Leegia, Blauhu, Matoushan, Leye, and Lungmei gauges. Because the watershed is small, we averaged the amount of hourly rainfall at the five rainfall gauges to represent the hourly precipitation over the whole watershed. Second, typhoon information data were collected from the CWB. The data were the pressure at the typhoon center (denoted as \(Y_1\)), distance of the typhoon from the watershed (\(Y_2\)), direction angle of the typhoon relative to the watershed (\(Y_3\)), maximal wind speed near the typhoon center (\(Y_4\)), radius of winds

<table>
<thead>
<tr>
<th>Typhoon</th>
<th>Periods</th>
<th>Typhoon</th>
<th>Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talim</td>
<td>30 Aug–2 Sep 2005</td>
<td>Sinlaku</td>
<td>12–15 Sep 2008</td>
</tr>
<tr>
<td>Kaenzi</td>
<td>24–27 Jul 2006</td>
<td>Fanapi</td>
<td>18–20 Sep 2010</td>
</tr>
<tr>
<td>Bophia</td>
<td>8–10 Aug 2006</td>
<td>Lionrock</td>
<td>1–2 Sep 2010</td>
</tr>
</tbody>
</table>
over 15.5 m s⁻¹ (Y₅), and moving speed of the typhoon (Y₆). Third, the weather property data were collected from the CWB. The weather data comprised the air pressure on the ground (W₁), air pressure at sea level (W₂), temperature on the ground (W₃), relative humidity on the ground (W₄), surface wind velocity on the ground (W₅), surface wind direction on the ground (W₆), and global solar radiation on the ground (W₇).

### 4. Experiment

#### 4.1. Modeling process

To predict the streamflow during typhoons, this study used a procedure for conceptualizing the forecasting processes (Fig. 2). The procedure for forecasting streamflow during typhoon periods in a reservoir watershed comprised five steps. First, raw data including the watershed hydrology, typhoon information, and weather properties in the reservoir watershed were collected (see Section 3). A total of 1806 records were available and all variables in the data were measured on an hourly scale.

The statistical approaches to streamflow prediction were assumed to be a function of the potential attributes selected and were expressed as

\[
H_{t+k}^{HYW} = f \left( \left( H_{t-j} \right)_{j=1}^{N_h} N_{y} - 0.5 \cdot D_{n_y} \right) \cdot \left( Y_{t-j} \right)_{j=1}^{N_y} N_{y} - 1 \cdot D_{n_y} \cdot \left( W_{t-j} \right)_{j=1}^{N_w} N_{w} - 1 \cdot D_{n_w} \right)
\]

where \( t \) is the hourly time index; \( H_{t+k}^{HYW} \) is the streamflow prediction at the lead time; \( k \); \( (H_{t-j}, Y_{t-j}, W_{t-j}) \) are the \( i \)th attribute in subsets \( H, Y, W \) at lag time \( j \), respectively; \( (N_h, N_y, N_w) \) are the numbers of attributes in subsets \( H, Y, W \), respectively; and \( (D_n_h, D_n_y, D_n_w) \) are the lengths of lag times at the \( i \)th attribute in subsets \( H, Y, W \), respectively.

Because a river is a dynamic environment, lagged inputs of each of these variables should be included in the model ([Bowden et al., 2005a]). In Eq. (7), a forecast horizon of 1–12 h (i.e., \( k = 1,12 \)) was determined for the long-term predictions. In addition, the suitable time lags for each attribute were verified to facilitate selecting the model inputs.

#### 4.2. Model input selection

Among the numerous input selection methods available (Bennett et al., 2013; Maier et al., 2010), two primary approaches are typically adopted: the model-free approach and model-based approach. Model-free approaches (e.g., correlation-based criterion method, mutual information method, available data method and domain knowledge method) do not rely on the performance of trained statistical models for selecting appropriate inputs. The most commonly used measure of statistical dependence for input selection is a correlation measure, which has the disadvantage of measuring only the linear dependence between variables.

In contrast to model-free approaches, model-based approaches (e.g., stepwise selection method, ad-hoc method, sensitivity analysis method, global method) rely on the development (structure selection, calibration, and evaluation) of several statistical models with different inputs to determine which of the candidate inputs should be included. The primary disadvantage of this approach is that it is time consuming, because several statistical models must be developed. The most commonly used model-based approach is a stepwise approach, where inputs are systematically added (constructive) or removed (pruning). More details are provided in the reports by Maier et al. (2010) and Wu et al. (2013).

In this study, both the correlation-based method, categorized as a model-free approach, and the stepwise selection method, categorized as a model-based approach, were employed to select suitable model inputs.

#### 4.2.1. Correlation-based criterion method

The correlation coefficients (\( r \)) between the attributes and target (i.e., the streamflow) were evaluated to select the appropriate lag times (Fig. 3). In the figure, \( H_1, H_2, Y_3, Y_5, Y_6, W_4, W_6 \) and their lag times were positively correlated with streamflow, whereas \( Y_1, Y_2, Y_4, W_1, W_5, W_7 \) and their lag times were negatively correlated. As expected, the \( r \) values decreased as the lag times increased. In Fig. 3a, b, \( H_1 \) and \( H_2 \) with their lag times from 1 to 8 h exhibited strong correlations (\( r > 0.7 \)), indicating a high correlation with streamflow.

In general, an \( r \) value above 0.7 represents a strong correlation; an \( r \) value between 0.3 and 0.7 represents a median correlation; and an \( r \) value below 0.3 represents a weak correlation. Therefore, we designed three cases of model inputs, which involved distinct correlation coefficients for selecting the lag-time lengths of
attributes, that is, the parameters of \(D_{H1}, D_{H2}, D_{W1}\) in Eq. (7). For the first case (Case 1), a high correlation coefficient value of 0.7 was set (i.e., if the \(r\) values were greater than 0.7, the specific attributes were chosen). For Cases 2 and 3, the thresholds of \(r\) were determined to be 0.5 (median correlation) and 0.3 (weak correlation), respectively.

### 4.2.2. Stepwise selection method

In addition, we designed Case 4, which involved using a stepwise selection method. The stepwise procedure proposed by Efroymson (1960) is an automatic procedure for statistical model selection in cases involving a high number of potential explanatory variables and no underlying theory on which to base the model selection (Furundzic, 1998). This procedure entails three approaches (i.e., forward selection, backward elimination, and stepwise regression). In this study, we used a stepwise regression approach, which is briefly presented in the following. This approach is a modification of the forward selection technique, with the difference being that variables already entered in the model are not assumed to remain until the completion of the process. As in forward selection, variables are added sequentially to the model, and the F-statistic for a variable added to the model should be significant at a set level (we set the significance level at the 5% criterion). After the variable is added to the model, the stepwise regression approach is employed to inspect all variables included and delete any variable that does not produce an F-statistic that is significant at the selected level. After this examination is completed, another variable can be included in the subsequent election step.

Table 3 lists all the lag-time lengths obtained empirically. When the threshold of \(r\) value increased, the number of selected attributes decreased. The total numbers of selected lag-time attributes were 18, 113, 274, and 19 for Cases 1–4, respectively. In all cases, \(H_1\) and \(H_2\) were used as model inputs, indicating reservoir inflow and rainfall in the reservoir watershed; they are the vital factors in the rainfall-runoff process.

After determining the model inputs, we classified the typhoons into training and validation subsets. The training subset contained 12 typhoons (2004–2007), and the validation subset comprised eight typhoons (2008–2011). The training subset was used to build prediction model structures and train a preference model. The validation subset was then employed to evaluate the model performance and verify its suitability for generalization. To conduct model analysis fairly under a fixed computing environment, the various models in the study were used in the Waikato Environment for Knowledge Analysis (WEKA) environment, which is a software suite written in Java (Bouckaert et al., 2010).

### 4.3. Model performance and results

This section presents a comparison of the results from the model runs and an evaluation performed using numerical statistics. Three criteria were considered in assessing the performance levels of the predictions derived from these models. First, the sample Pearson correlation coefficient, denoted as \(r\) in the previous section, can be written as:

\[
r = \frac{\sum_{i=1}^{n} (O_i^{\text{obs}} - \bar{O}^{\text{obs}})(O_i^{\text{pre}} - \bar{O}^{\text{pre}})}{\sqrt{\sum_{i=1}^{n} (O_i^{\text{obs}} - \bar{O}^{\text{obs}})^2 \sum_{i=1}^{n} (O_i^{\text{pre}} - \bar{O}^{\text{pre}})^2}}
\]

where \(O_i^{\text{obs}}\) is the observed value at record \(i\), \(O_i^{\text{pre}}\) is the predicted value at record \(i\), \(\bar{O}^{\text{pre}}\) is the average of the predictions, and \(n\) is the number of records.

Second, the relative absolute error (RAE) is defined as:

\[
\text{RAE} = \frac{\sum_{i=1}^{n} |O_i^{\text{pre}} - O_i^{\text{obs}}|}{\sum_{i=1}^{n} |O_i^{\text{obs}} - \bar{O}^{\text{obs}}|}
\]

Third, the root relative squared error (RRSE) is:

\[
\text{RRSE} = \sqrt{\sum_{i=1}^{n} \left(\frac{O_i^{\text{pre}} - O_i^{\text{obs}}}{O_i^{\text{obs}}}\right)^2 / \sum_{i=1}^{n} \left(\frac{O_i^{\text{obs}} - \bar{O}^{\text{obs}}}{O_i^{\text{obs}}}\right)^2}
\]

According to these criteria, lower RAE and RRSE values and higher \(r\) values typically indicate favorable performance levels. A perfectly accurate model would have RAE and RRSE values approximating 0 and \(r\) values approximating 1.

We used the validation subset; Fig. 4 shows the model performance levels with corresponding \(r\), RAE, and RRSE values to enable a comparison of Cases 1–4 with the forecast horizons varying from 1 to 12 h. According to Fig. 4, the forecasts made 1 h in advance yielded superior \(r\), RAE, and RRSE values compared with those...
made 2–12 h in advance, indicating that increased forecasting horizons yielded increased errors in the prediction of streamflows. Regarding Case 3 in Fig. 4a,b,c, because the WEKA solver could not be executed using the pace and isotonic regressions, these regressions yielded no results. The reason might be that an increasing number of model inputs (i.e., 274) exceeded the limitations of the

Table 3
Attribute selection for four cases.

<table>
<thead>
<tr>
<th>Case</th>
<th>Selected attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>(H_{1,t}, H_{2,t}, H_{3,t}, H_{4,t}, H_{5,t} )</td>
</tr>
<tr>
<td>Case 2</td>
<td>(W_{1,t}, W_{2,t}, W_{3,t}, W_{4,t}, W_{5,t}, W_{6,t} )</td>
</tr>
<tr>
<td>Case 3</td>
<td>(Y_{1,t}, Y_{2,t}, Y_{3,t}, Y_{4,t}, Y_{5,t}, Y_{6,t} )</td>
</tr>
<tr>
<td>Case 4</td>
<td>(T_{1,t}, T_{2,t}, T_{3,t}, T_{4,t}, T_{5,t}, T_{6,t} )</td>
</tr>
</tbody>
</table>

Fig. 3. Correlation coefficients of various attributes versus streamflow.
Although the results could not be obtained, their absence did not influence the evaluations of these cases, because Case 3 did not involve optimal input combinations (this is discussed elsewhere).

To determine the prediction ability, we calculated the average performance levels for 1- to 12-h predictions. The results showed that the MLP ANN achieved more favorable performance (lower RAE and RRSE, and higher r values) than did the regression-based models (Table 4). For example, among the Case 4 prediction models, the order of RRSE values was as follows: MLP (0.513) < pace (0.514) < OLS (0.535) < additive (0.614) < isotonic (0.668). Moreover, we observed excellent results in Case 4 and the least favorable results in Case 3. For example, when the MLP was applied in the four cases, the order of the RRSE values was as follows: Case 4 (0.513) < Case 2 (0.521) < Case 1 (0.558) < Case 3 (0.565). Comprehensive evaluations of these models and cases are described in subsequent sections.

5. Evaluation and discussion

5.1. Comparison of regressions and ANNs

This study defined the improvement rate for comparison among five models (i.e., OLS, pace, isotonic, and additive regressions and MLP ANNs) in each case of selecting model inputs. The improvement rates of the criteria r, RAE, and RRSE (which refer to $I_{i,1}^r$, $I_{i,1}^{RAE}$, and $I_{i,1}^{RRSE}$, respectively) for model i (where i is referred to the models) in Case 1, for example, can be respectively expressed as

$$I_{i,1}^r = \left( r_{i,1} - r_{OLS,1} \right) / r_{OLS,1} \times 100$$  \hspace{1cm} (11)

$$I_{i,1}^{RAE} = \left( RAE_{OLS,1} - RAE_{i,1} \right) / RAE_{OLS,1} \times 100$$  \hspace{1cm} (12)

$$I_{i,1}^{RRSE} = \left( RRSE_{OLS,1} - RRSE_{i,1} \right) / RRSE_{OLS,1} \times 100$$  \hspace{1cm} (13)

where $r_{OLS,1}$, $RAE_{OLS,1}$, and $RRSE_{OLS,1}$ are the r, RAE, and RRSE values obtained using the OLS model, respectively, in Case 1, and $r_{i,1}$, $RAE_{i,1}$, and $RRSE_{i,1}$ are the r, RAE, and RRSE values obtained using model i, respectively, in Case 1. Here, we determine the OLS performance input sizes of the WEKA solver in the pace and isotonic regressions. Although the results could not be obtained, their absence did not influence the evaluations of these cases, because Case 3 did not involve optimal input combinations (this is discussed elsewhere).

To determine the prediction ability, we calculated the average performance levels for 1- to 12-h predictions. The results showed that the MLP ANN achieved more favorable performance (lower RAE and RRSE, and higher r values) than did the regression-based models (Table 4). For example, among the Case 4 prediction models, the order of RRSE values was as follows: MLP (0.513) < pace (0.514) < OLS (0.535) < additive (0.614) < isotonic (0.668). Moreover, we observed excellent results in Case 4 and the least favorable results in Case 3. For example, when the MLP was applied in the four cases, the order of the RRSE values was as follows: Case 4 (0.513) < Case 2 (0.521) < Case 1 (0.558) < Case 3 (0.565). Comprehensive evaluations of these models and cases are described in subsequent sections.

5. Evaluation and discussion

5.1. Comparison of regressions and ANNs

This study defined the improvement rate for comparison among five models (i.e., OLS, pace, isotonic, and additive regressions and MLP ANNs) in each case of selecting model inputs. The improvement rates of the criteria r, RAE, and RRSE (which refer to $I_{i,1}^r$, $I_{i,1}^{RAE}$, and $I_{i,1}^{RRSE}$, respectively) for model i (where i is referred to the models) in Case 1, for example, can be respectively expressed as

$$I_{i,1}^r = \left( r_{i,1} - r_{OLS,1} \right) / r_{OLS,1} \times 100$$  \hspace{1cm} (11)

$$I_{i,1}^{RAE} = \left( RAE_{OLS,1} - RAE_{i,1} \right) / RAE_{OLS,1} \times 100$$  \hspace{1cm} (12)

$$I_{i,1}^{RRSE} = \left( RRSE_{OLS,1} - RRSE_{i,1} \right) / RRSE_{OLS,1} \times 100$$  \hspace{1cm} (13)

where $r_{OLS,1}$, $RAE_{OLS,1}$, and $RRSE_{OLS,1}$ are the r, RAE, and RRSE values obtained using the OLS model, respectively, in Case 1, and $r_{i,1}$, $RAE_{i,1}$, and $RRSE_{i,1}$ are the r, RAE, and RRSE values obtained using model i, respectively, in Case 1. Here, we determine the OLS performance input sizes of the WEKA solver in the pace and isotonic regressions. Although the results could not be obtained, their absence did not influence the evaluations of these cases, because Case 3 did not involve optimal input combinations (this is discussed elsewhere).

To determine the prediction ability, we calculated the average performance levels for 1- to 12-h predictions. The results showed that the MLP ANN achieved more favorable performance (lower RAE and RRSE, and higher r values) than did the regression-based models (Table 4). For example, among the Case 4 prediction models, the order of RRSE values was as follows: MLP (0.513) < pace (0.514) < OLS (0.535) < additive (0.614) < isotonic (0.668). Moreover, we observed excellent results in Case 4 and the least favorable results in Case 3. For example, when the MLP was applied in the four cases, the order of the RRSE values was as follows: Case 4 (0.513) < Case 2 (0.521) < Case 1 (0.558) < Case 3 (0.565). Comprehensive evaluations of these models and cases are described in subsequent sections.
levels in Case 1 as the base values of Eqs. (11)–(13) to enable a comparison. Likewise, the improvement rates of various models in Cases 2–4 can be defined as Eqs. (11)–(13), and the performance levels of the OLS regression in Cases 2–4 are the base values for the improvement rate in Cases 2–4, respectively.

5.1.1. Comparisons among five models

Fig. 5 illustrates the measurements of the improvement rate for the r, RAE, and RRSE of five models in Cases 1–4. In each case, we observed that the pace regression and MLP ANN obtained positive scores, whereas the isotonic and additive regressions obtained negative scores, indicating that the prediction ability of the pace regression and MLP ANN is superior to that of the conventional OLS regression for streamflow predictions, whereas the isotonic and additive regressions are inferior to OLS regression.

Comparison between the MLP ANN and four regressions revealed that the MLP ANN exhibited an excellent prediction ability compared with the regression-based methods, according to the measures of the improvement rate in each case. This is because an ANN performs favorably in applications when the functional form is nonlinear (Tso and Yau, 2007). The ANN employs a network architecture that contains several hidden layers and hidden units. This feedforward ANN has been widely proven to be a universal approximator that can learn any continuous functions with arbitrary accuracy. For more detailed descriptions, see Li et al. (2015), Galelli et al. (2014), Maier and Dandy (2000), Maier et al. (2014), Ren and Zhao (2002), and Wu et al. (2013).

5.1.2. Comparison among regression-based models

A close inspection of the improvement rates of the regression-based models reveals that the improvement levels of the pace regression are superior to those of the OLS, pace, isotonic, and additive regressions (Fig. 5). The pace regression might represent an improvement over the OLS regression because it entails evaluating the effect of each variable and using a clustering analysis to improve the statistical basis. As indicated in Wang and Witten (1999), the pace regression outperforms OLS estimation, because it challenges the Gauss-Markov theorem, the OLS regression represents a linear unbiased estimator. However, the inferiority of the OLS is attributable to the unbiasedness constraint, which means that the OLS fails to use all information implicit in the data. Biased estimators such as subset selection and shrinkage can outperform the OLS estimator in particular situations. Therefore, the pace regression estimator, which is also biased, outperforms the OLS estimator.

5.2. Comparison among data selection methods

Regarding data selection, we adopted the correlation-based criterion (Cases 1–3) and stepwise selection methods (Case 4) for determining model inputs. For the development of statistical models such as the ANN and regressions, inputs must be determined on the basis of input independence (Bowden et al., 2005b; May et al., 2008; Castelletti et al., 2012). Similarly, we defined the improvement rate for comparison among various models in Cases 1–4. The improvement rates (\( \text{IR}_{\text{MLP}} \), \( \text{IR}_{\text{RAE}} \), and \( \text{IR}_{\text{RRSE}} \)) of criteria r, RAE, and RRSE for an MLP model in Case \( j \) (where \( j \) ranges from 1 to 4) I, for example, can be respectively expressed as

\[
\text{IR}_{\text{MLP}}(\%) = \left( \frac{\text{r}_{\text{MLP},j} - \text{r}_{\text{MLP},1}}{\text{r}_{\text{MLP},1}} \right) \times 100
\]

\[
\text{IR}_{\text{RAE}}(\%) = \left( \frac{\text{RAE}_{\text{MLP},j} - \text{RAE}_{\text{MLP},1}}{\text{RAE}_{\text{MLP},1}} \right) \times 100
\]

\[
\text{IR}_{\text{RRSE}}(\%) = \left( \frac{\text{RRSE}_{\text{MLP},j} - \text{RRSE}_{\text{MLP},1}}{\text{RRSE}_{\text{MLP},1}} \right) \times 100
\]

where \( \text{r}_{\text{MLP},j}, \text{RAE}_{\text{MLP},j}, \text{RRSE}_{\text{MLP},j} \) are the r, RAE, and RRSE values obtained using the MLP model in Case 1, and \( \text{r}_{\text{MLP},j}, \text{RAE}_{\text{MLP},j}, \text{RRSE}_{\text{MLP},j} \) are the r, RAE, and RRSE values obtained using the MLP model in Case \( j \). We defined the MLP performance levels as the base values. Likewise, the improvement rate of the other models (i.e., the OLS, pace, isotonic, and additive regressions) in Cases 1–4 can be defined as Eqs. (14)–(16), and the performance levels of the OLS, pace, isotonic, and additive regressions in Case 1 were the base values for the improvement rates of their corresponding regressions.

Fig. 6 illustrates the improvement rates of each model in Cases 1–4; these rates were subsequently used to identify a suitable set of attributes. Because of the limitations of the input sizes of the solver, the results of the pace and isotonic regressions in Case 3 in Fig. 6b,c are absent. The results and findings are presented in the following:

5.2.1. Comparing correlation-based methods among cases 1–3

As shown in Fig. 6, Case 3 exhibited negative improvement rates.
and the lowest performance levels of the three cases (excluding Fig. 6b,c), whereas Case 2 exhibited positive improvement rates and the highest performance levels. As shown in Table 4, the total numbers of selected lag-time attributes increased in Cases 1–3. Although Case 3 had more inputs than Cases 1 and 2 did, it exhibited lower performance levels than those cases. The reason might be the input redundancy, where certain selected inputs provide relevant information but are related to each other and therefore provide redundant information. This can cause several problems. First, redundant inputs increase the likelihood of overfitting for the ANN (Fig. 6e). This is because a high number of inputs generally increases the network size and, hence, the number of connection weights that must be calibrated (Maier et al., 2010; Wolfs and Willems, 2014). Second, a collinearity problem occurs, which involves the nonindependence of predictor variables, usually in a regression-type analysis (Fig. 6a,d). Collinearity is a problem where a solution is unobtainable if two highly collinear variables are both correlated with a dependent variable; without further information the true predictor cannot be identified (Dormann et al., 2013).

5.2.2. Comparing the correlation-based method and stepwise selection method

As shown in Fig. 6, Case 4 exhibited positive improvement rates and the highest performance levels for all models in all cases. Thus, data selection using a stepwise selection method yielded more reliable results than those of selection using a correlation-based criterion, demonstrating that the selected inputs were critical input variables. For example, Fig. 6a shows a comparison of Cases 2 and 4; the number of selected lag-time attributes in Case 4 (i.e., 19) is considerably lower than that in Case 2 (i.e., 113). In addition, the results of Case 4 are superior to those of Case 2. This is because the stepwise approach, where inputs are systematically added (constructive) or removed (pruning), is used to select the combination of inputs that maximizes model performance (Maier et al., 2010). As indicated in Chokmani et al. (2008), the stepwise approach can be used to reduce the number of variables efficiently. In addition to enabling the selection of the most relevant explanatory variables, and consequently, the automatic yet rigorous selection of an optimal regressive model, the stepwise approach yields satisfactory results.

This study evaluated the correlation-based and stepwise selection methods and examined the influence of input independence. However, various techniques such as principal component analysis, clustering analysis, and self-organizing maps can be employed. Maier et al. (2010) and Wu et al. (2014) reviewed several techniques that can be used to assess the significance of the relationship between potential inputs and outputs.

5.3. Evaluating the tradeoff between model complexity and accuracy

To compare the various models equally, the AIC metric proposed by Akaike (1974) was introduced, facilitating the determination of the tradeoff between model complexity and accuracy. The AIC metric is a typical model selection method that enables determining the optimal model for minimizing an expected discrepancy (Bennett et al., 2013). The Kullback-Liebler distance was used as a fundamental basis for the model selection. The AIC comprises two terms: The first term depends only on the mean square error (MSE) of a model and is referred to as the error term. The second term, called the penalty term, depends on the number of parameters employed in the model and is used to penalize the parameters for only a small gain in predictive ability. Both the error and penalty terms ensure that the AIC favors simple models with low error values (Akpa and Unuabonah, 2011).

The AIC can be calculated as
\[ \text{AIC} = n \log(\sigma^2) + 2K \]  
(17)

where \( n \) is the number of observation data; \( \sigma^2 \) is the MSE between the target and actual outputs; and \( K \) is the number of model parameters in a regression.

Burnham and Anderson (2002) defined the bias adjustment or correction for the AIC when \( n/K < 40 \) as

\[ \text{AIC} = n \log(\sigma^2) + \frac{2nK}{n - K - 1} \]  
(18)

Fig. 7 shows the tradeoff between the error and penalty terms on AIC measures according to the forecasts made 1 h in advance. As shown, the MLP ANN can reduce the error term by increasing the number of parameters employed in the ANN. By contrast, the pace, OLS, isotonic, and additive regressions might increase the error term relative to that of the MLP ANN. The isotonic and additive regressions increased the error term particularly markedly. When we focused on the penalty term of the AIC, the MLP ANN yielded higher values than those of the four regression-based models because of the numerous parameters used in the ANN structures.

Fig. 8 shows the AIC metric curves for the forecasts made 1–12 h in advance. The results show that, first, the AIC values increase with the forecast horizon, indicating that an increase in the forecast

Fig. 7. Plot of error and penalty terms for forecasts made 1 h in advance.

Fig. 8. Plot of AIC metrics for forecasts made 1–12 h in advance.
horizons resulted in an increase in the error. This is consistent with Fig. 4b,c. Furthermore, the pace regression exhibits the lowest AIC curve in Fig. 8 (excluding Fig. 8c), and the OLS regression exhibits the second lowest AIC curve. Fig. 9 illustrates the average AIC metrics for all forecast horizons. The MLP ANN in Cases 1 and 4 (Figs. 8a,d and 9a,d) exhibits slightly higher AIC metrics than those of the pace and OLS regressions but lower AIC metrics than those of the isotonic and additive regressions. However, as shown in Figs. 8b,c and 9b,c, the MLP ANN exhibits the highest AIC metrics of all regressions, possibly because an increase in the number of parameters for modeling the ANNs (such as Cases 2 and 3) lowered the high AIC metrics.

6. Two-segment prediction modeling

6.1. Modeling process

The hydrograph observed at the drainage basin outlet for a given period is the sum of several flow components. As mentioned previously, because the rising and falling limbs have different physical meanings, the following two procedures were employed in the experiment: 1) the number of selected lag-time attributes was reidentified through a stepwise selection method; and 2) the pace and OLS regressions and MLP ANN were employed to formulate the two segments (i.e., rising and falling limbs) of the streamflow prediction models. According to the analysis of feature selection, the attributes $H_1, H_2, H_2/H_2, W_5, W_6, W_6, W_6, W_6, W_6, W_6, W_6, W_6, W_6, W_6, W_6$, were chosen for the rising limb, and $H_1, H_2, H_2/H_2, W_5, W_6, W_6, W_6, W_6, W_6, W_6, W_6, W_6$, were chosen for the falling limb. After the model inputs were determined, the two segments of the model were constructed individually. Fig. 10 depicts the model performance levels with corresponding r, RAE, and RRSE values for enabling a comparison between single- and two-segment model cases (i.e., pace and OLS regressions and MLP ANN) with the forecast horizons varying from 1 to 12 h. From the figure, we observed that the results for the two-segment models were superior to those of the single-segment models.

6.2. Statistical significance for single- and two-segment models

To examine the robustness of the results for forecast horizons varyin...
varying from 1 to 12 h, statistical significance tests were performed for single-segment and two-segment models. The tests were conducted to determine the population mean ($\mu$) and variance ($\sigma^2$). First, the difference between the two population means ($\mu_1 - \mu_2$) was tested:

Two-tailed test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$ (19)

Rejection region for $H_0$: $p$ value < $\alpha/2$ or $|Z| > z_{c,\alpha/2}$ (20)

where $H_0$ denotes the null hypothesis; $H_a$ denotes the alternative hypothesis; $\mu_1$ and $\mu_2$ denote the mean of the observed and predicted water level values, respectively; $\alpha/2$ is the level of significance (two-tailed); and $z_{c,\alpha/2}$ denotes the critical value of the pivotal statistic $z$ at $\alpha/2$.

In addition, the ratio of two population variances was tested. The ratio of the two variances $\sigma_1^2/\sigma_2^2$ exhibited an $F$ distribution (Mendenhall and Sincich, 2007). The hypothesis tests of variance were formulated as

Two-tailed test $H_0: \sigma_1^2/\sigma_2^2 = 1$ vs. $H_a: \sigma_1^2/\sigma_2^2 \neq 1$ (21)

Rejection region for $H_0$: $p$ value < $\alpha/2$ or $F > F_{c,\alpha/2}$ (22)

where $\sigma_1^2$ and $\sigma_2^2$ denote the variance of the observed and predicted values of water levels, respectively. The significance of the hypothesis means and variances was determined according to the model results. The $\alpha$ value was set to 0.05. The critical values $z_{c,\alpha/2}$ and $F_{c,\alpha/2}$ were then used to determine whether the mean and variance hypothesis tests yielded two-tailed significance. If the $z$ and $F$ statistics are less than the $z_{c,\alpha/2}$ and $F_{c,\alpha/2}$ values, respectively, then the model results are nonsignificant and $H_0$ is not rejected; that is, the model results pass the test. The $z$ and $F$ statistics were obtained using the model predictions.

Fig. 11 illustrates the results of the mean and variance statistical significance tests. $H_0$ was not rejected when $|Z| \leq z_{c,\alpha/2}$ and $F \leq F_{c,\alpha/2}$. Regarding the single-segment models, $H_0$ was not rejected for 1- to 4-h-ahead predictions obtained using the OLS and pace regressions or for 1- to 3-h-ahead predictions obtained using the MLP. Regarding two-segment models, $H_0$ was not rejected for 1- to 6-h-ahead predictions obtained using the OLS and pace regressions or for 1- to 5-h-ahead predictions obtained using the MLP. According to the statistical analyses, the two-segment OLS and pace models were the most effective models.

7. Conceptual rainfall–runoff model and comparisons

In this section, the PWRI-distributed hydrological model results from Kimura et al. (2014) are compared with those obtained using the aforementioned statistical models. Typhoon Sinlaku in 2008 was used for the comparison. First, the conceptual PWRI model is briefly described.
7.1. Description of conceptual PWRI model

The PWRI-distributed hydrological model is used to calculate the conversion of rainfall into runoff (Yoshino et al., 1990). The model divides the entire watershed into uniform cells and computes the flow at each cell. The flow is computed for surface, aquifer, and river tanks through two or three vertical layers (Fig. 12). The relationship between rainfall and outflow is represented by a black box that contains several constants that determine the relationship. The constants can be set according to observed rainfall and outflow data or estimated using data from similar rivers. The distributed model can be divided into physical distributed and constant distributed models. The physical distributed model treats the outflow as the migration of rainfall in the river basin and represents the migration process by using infiltration and simple inequilateral flow equations. Generally, the physical distributed model requires considerable information such as soil, geology, and river shape for modeling; consequently, the time for calculation becomes long. However, the constant model uses the concept model for estimating outflow from the river basin (generally on mesh), greatly shortening the calculation time, and is considered an appropriate model for flood forecasting.

7.2. Simulated typhoon and parameter calibration

Typhoon Sinlaku stalled over Northern Taiwan in September 2008, and the associated precipitation caused severe landslide and flood damage throughout Taiwan. The measured accumulated rainfall and total discharge volume were approximately 750 mm and 290 million cubic meters, respectively, at the gauges near Tsengwen Reservoir (Water Resources Agency, 2008). According to Kimura et al. (2014), the maximum size of the watershed is 91 × 96 cells (horizontal and vertical), and each cell is uniformly 400 × 400 m in size. For determining all parameters in the PWRI-distributed hydrological model, the parameters were employed after manually tuning the value of each parameter to an acceptable range. Additional details on the model analysis procedure are provided by Kimura et al. (2014). The basin-averaged rainfall hyetographs and PWRI model-predicted runoff are depicted in Fig. 13.

7.3. Comparisons between conceptual and statistical models

Fig. 13 also illustrates hydrographs of six single- and two-segment statistical models for 1-h-ahead predictions of streamflow during Typhoon Sinlaku. Table 5 lists the performance indicators for the entire hydrograph and its corresponding rising and recession limbs according to the results of the conceptual PWRI model and various statistical models. Comparing the performance levels of the entire hydrograph revealed that the prediction ability of the two-segment statistical models was superior to that of the single-segment statistical models and conceptual PWRI model. When we separated the entire hydrograph into rising and recession limbs, the RAE and RRSE for the two-segment statistical models were superior to those for the single-segment statistical models and PWRI model in both limbs. Therefore, using two input–output mappings representing the two limbs of the hydrograph may be more effective than developing a single-segment model representing the mapping of the entire hydrograph.

7.4. Discussion

Rainfall–runoff transformation is among the most complex hydrological phenomena, usually involving a number of interconnected elements such as evapotranspiration, infiltration, surface and subsurface runoff generation, and routing (Chen and Adams, 2006). Approaches to flow prediction can be broadly divided into three categories: mechanistic modeling, statistical or “black box” modeling, and conceptual modeling (Leahy et al., 2008). A typical mechanistic approach combines precipitation observations or forecasts with detailed physical models of the river catchment. The statistical approach is based on the properties of observed data (e.g., time series of river stages or precipitation), rather than the physical properties of the catchment system itself. Conceptual modeling lies between these approaches, relying on a simplified representation of the physical system, which can be calibrated using past data.

In the previous section, we observed that the conceptual PWRI model yielded the least favorable results. This was possibly because, as indicated by Hingray et al. (2014), conceptual models
are based on a simplified representation of the factors influencing a hydrosystem, such as its geometry, its physical characteristics, and the physical processes that govern its behavior. The representation is conceptual in that it is dependent on how the hydrologist perceives the hydrological behavior of the basin. However, the physical characteristics of the natural environment exhibit high spatial variability, which is markedly difficult to describe (Shin et al., 2015). For example, the PWRI model is a spatially distributed hydrological model that is manually calibrated; limited information is available on the calibration methodology, and it is difficult to determine whether appropriate calibration has been applied. In other words, the hydrological processes that drive the hydrological behavior of the hydrosystem are so numerous and complex that they cannot all be described.

In contrast to conceptual models, statistical models are based on the observed relationships between the inputs and outputs of the considered hydrosystem; they represent relationships between system input and output variables, such as the rainfall—runoff relationship, through a set of equations developed and adjusted on the basis of data obtained regarding the system. In this type of model, the hydrosystem is considered a black box model. The resulting representation can account for various components of the hydrological cycle. Advantageous statistical approaches have been developed for easy use in an operational mode. As stated in Dibike and Solomatine (2001), while conceptually and physically based models crucial in understanding hydrological processes, there are numerous practical situations where the main concern is generating accurate predictions at specific locations. In these situations it is preferable to implement a simple black box statistical, data-driven, or machine learning, model to identify a direct mapping between the inputs and outputs without detailed information on the internal structures involved in the physical process. However, these models have several limitations; for example, they can omit one or more critical factors affecting hydrological behavior (Hingray et al., 2014). Although they are often capable of suitably reproducing observations, it is difficult to use these models in hydrometeorological contexts that are different from those for which they were developed.

Notably, O’Connor (1997) stated that conceptual models can be employed to evaluate the effect of land use on hydrological processes on the basis of relationships between model parameters and measurable physical characteristics; furthermore, compared with black box models, conceptual models have greater potential for further structural development. Because of the physical basis of conceptual models, it is reasonable to expect that conceptual models are more accurate in simulating the rainfall—runoff process (Chen and Adams, 2006). Generally speaking, statistical methods are often used in short-range flood forecasting, whereas conceptual hydrological models are employed in medium-range forecasts in small catchments.

### 8. Summary and conclusions

In water resources and environmental management, the prediction of river streamflow during typhoons is a critical research topic that has attracted considerable interest. This study examined regression-based techniques that involve using ordinary least squares, pace, isotonic, and additive regressions for streamflow forecasting during typhoons. To evaluate the effectiveness of these traditional regression approaches, the prediction results were compared with those from an artificial neural network. Moreover, this study decomposed a flow hydrograph into two segments, rising and falling limbs, and modeled individual segments by using statistical techniques. In addition, the conceptual PWRI rainfall—runoff model and statistical models were compared. The forecast horizons ranged from 1 to 12 h at the Tsengwen Reservoir watershed in Southern Taiwan.

A series of assessments, including statistical analyses and simulations, was conducted. First, we examined the single-segment statistical models and obtained the following findings:

- Evaluating the appropriate number of time-lagged input data according to an appropriate subset selection method. This study involved designing four cases of attribute combinations with various lag times of the forecast target. The conventional correlation-based criterion and stepwise selection method were used in selecting the inputs of various models. The results show that Case 4 and using a stepwise selection method yielded positive improvement rates and exhibited the highest performance levels of all cases. Thus, data selection using a stepwise selection method yielded more reliable results than those of selection using a correlation-based criterion, demonstrating that the selected inputs were critical input variables.
- Assessing multisource data with long lag times by using four regressions and the MLP ANN to formulate streamflow predictions. The data examined 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. The results demonstrate that the MLP ANN provides superior prediction compared with the regression-based methods, according to the improved levels. In addition, a comparison among regression-based models showed that the pace regression is superior to the OLS, pace, isotonic, and additive regressions.
- Determining the optimal model according to the tradeoff between model complexity and accuracy. The AIC criterion was introduced to evaluate the various models equally. The results show that the MLP ANN reduced the error term by increasing the number of parameters employed in the ANN. By contrast, the pace, OLS, isotonic and additive regressions increased the error term relative to that of the MLP ANN. Regarding the penalty term of the AIC, the MLP ANN yielded higher values than the

### Table 5

<table>
<thead>
<tr>
<th>Model</th>
<th>Entire hydrograph</th>
<th>Rising limb</th>
<th>Recession limb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f</td>
<td>RAE</td>
<td>RSSE</td>
</tr>
<tr>
<td>PWRI</td>
<td>0.984</td>
<td>0.286</td>
<td>0.127</td>
</tr>
<tr>
<td>OLS</td>
<td>0.964</td>
<td>0.230</td>
<td>0.075</td>
</tr>
<tr>
<td>Pace</td>
<td>0.964</td>
<td>0.231</td>
<td>0.075</td>
</tr>
<tr>
<td>MLP</td>
<td>0.971</td>
<td>0.222</td>
<td>0.078</td>
</tr>
<tr>
<td>Two-Seg.-OLS</td>
<td>0.983</td>
<td>0.166</td>
<td>0.038</td>
</tr>
<tr>
<td>Two-Seg.-Pace</td>
<td>0.983</td>
<td>0.166</td>
<td>0.038</td>
</tr>
<tr>
<td>Two-Seg.-MLP</td>
<td>0.988</td>
<td>0.191</td>
<td>0.053</td>
</tr>
</tbody>
</table>
four regression-based models did because of the numerous parameters used in constructing the network structures. Therefore, we conclude that the pace regression can be used as a prediction model according to the evaluation of the tradeoff between model complexity and accuracy.

Second, we compared the single- and two-segment statistical models and examined the robustness of the results with forecast horizons varying from 1 to 12 h. Statistical significance tests were performed for the single- and two-segment models. Regarding the single-segment models, $H_0$ was not rejected for 1- to 4-h-ahead predictions obtained using the OLS and pace regressions or for 1- to 3-h-ahead predictions obtained using the MLP. Regarding two-segment models, $H_0$ was not rejected for 1- to 6-h-ahead predictions obtained using the OLS and pace regressions or for 1- to 5-h-ahead predictions obtained using the MLP. According to the statistical analyses, the two-segment OLS and pace models were favorable models. Overall, 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 comparisons between the PWRI model output and that of the developed statistical models. The results showed that the PWRI model yielded the least favorable results. We determined the prediction ability of the two-segment statistical models to be superior to that of the single-segment statistical models and conceptual PWRI model. Therefore, using two input—output mappings representing the two limbs of the hydrograph may be more effective than developing a single-segment model representing the mapping of the entire hydrograph.

Acknowledgements

We are grateful to the Ministry of Science and Technology, Taiwan for supporting this study under Grant No. MOST105-2221-E-019-041. The author also acknowledges the constructive comments of the referees.

References

Leahy, P., Kiey, G., Corcoran, G., 2008. Structural optimisation and input selection of