# Development of a real-time regional-inundation forecasting model for the inundation warning system

Gwo-Fong Lin, Hsuan-Yu Lin and Yang-Ching Chou

## ABSTRACT

Accurate forecasts of the inundation depth are necessary for inundation warning and mitigation. In this paper, a real-time regional forecasting model is proposed to yield 1- to 3-h lead time inundation maps. First, the *K*-means based cluster analysis is developed to group the inundation depths and to indentify the control points. Second, the support vector machine is used as the computational method to develop the point forecasting module to yield inundation forecasts for each control point. Third, based on the forecasted depths and the geographic information, the spatial expansion module is developed to expand the point forecasts to the spatial forecasts. An actual application to Siluo Township, Taiwan, is conducted to demonstrate the advantage of the proposed model. The results indicate that the proposed model can provide accurate inundation maps for 1- to 3-h lead times. The accurate long lead time forecasts can extend the lead time to allow sufficient time to take emergency measures. Furthermore, the proposed model is an efficient process that can be trained rapidly with real-time data and is more suitable to be integrated with the decision support system. In conclusion, the proposed modeling technique is expected to be useful to support the inundation warning systems.

**Key words** | inundation warning system, *K*-means clustering algorithm, rainfall-inundation forecasting, support vector machine

**Gwo-Fong Lin** (corresponding author)
**Hsuan-Yu Lin**
**Yang-Ching Chou**
Department of Civil Engineering,
National Taiwan University,
Taipei 10617,
Chinese Taiwan
E-mail: *gflin@ntu.edu.tw*

## INTRODUCTION

Typhoon rainfall produces valuable water resources, but the flood inundation resulting from excessive precipitation frequently causes loss of human life and serious economic damage. Taiwan is located on one of the main paths of the north-western Pacific typhoons. On average, about four typhoons have hit Taiwan each year during the past 100 years. Thus, in order to mitigate flood damage, the development of an early warning system has been recognized as an important task. Carsell *et al.* (2004) indicated that a warning system can increase the mitigation time, which is a consequence of a reduction in the time of several actions, such as data collection, flood evaluation, emergency notification, and decision making. With the increased mitigation time, the loss of life and the impact on the economy can be reduced. In the development of early warning systems, accurate and efficient forecasts of the inundation depth are always required as an important reference for making important emergency evacuation decisions. Hence, to improve inundation depth forecasting is an important task of disaster warning systems and is expected to be useful in disaster prevention and mitigation.

The prediction of flood inundation has always been a challenging task for hydraulic engineers. Based on the geographical information, the surface overland flow processes can be simulated using mathematical and numerical models (such as two-dimensional (2-D) overland-flow model). General introductions of hydraulic modeling and comprehensive reviews of their applications for geomorphological and hydrological purposes have been presented by Lane (1998). Examples of the numerical model can be found in the literature (Bates & De Roo 2000; Guo *et al.* 2007; Kuiry *et al.* 2010; Neal *et al.* 2012; Seyoum *et al.* 2012).

doi: 10.2166/hydro.2013.202

Conventionally, the kernel of these numerical models is based on non-inertia surface flow dynamics. Timely and accurate forecasts in flood-prone areas are essential prerequisites for the provision of reliable early warning systems. However, the numerical models suffer from the large computational time for the iterative process to yield the inundation depth, and hence the real-time inundation simulation (or prediction) cannot be achieved.

An attractive alternative to the numerical models is neural networks (NNs), which are a kind of information processing system with great flexibility in modeling nonlinear systems. The American Society of Civil Engineers (ASCE) Task Committee (2000a, b) and Maier & Dandy (2000) provide a general introduction and comprehensive reviews of the applications of NNs in hydrology. Due to their ability to model nonlinear systems, NNs have been extensively used in various aspects of hydrology (e.g., de Vos & Rientjes 2005; Giustolisi & Simeone 2006; Hu *et al.* 2007; Lin & Chen 2008; Chau & Wu 2010; Remesan *et al.* 2010; Chen *et al.* 2011; Pramanik *et al.* 2011; Wu & Chau 2011; Adamowski *et al.* 2012; Lin *et al.* 2012; Nguyen & Chua 2012; Wei *et al.* 2012; Lin *et al.* 2013a, b; Nourani *et al.* 2013). The major advantage of NNs is that they are capable of simulating the relationship between desired output and available inputs. Because of the flexibility in modeling nonlinear systems and the computational efficiency, NNs have received considerable attention.

For inundation forecasting, the potential of NNs has also been demonstrated in the literature (e.g., Chang *et al.* 2010; Pan *et al.* 2011; Kia *et al.* 2012). However, there are as yet few studies and further investigations are needed. Previous studies have shown that models yield acceptable forecasts for a lead time of 1-h only (Chang *et al.* 2010; Pan *et al.* 2011; Kia *et al.* 2012). Based on the definition from the Taiwan Water Resources Agency, the Level 2 alert is defined as 'if the rainfall continues over the flood warning area, it may be flooded at the flood-prone villages and roads in 3 h'. Thus, for inundation mitigation and disaster warnings, an inundation forecasting model will be more helpful if it can provide longer lead time forecasts. Additionally, the architecture and the weights of conventional NNs, such as back-propagation networks (BPNs) and radial basis function networks (RBFNs), are determined by a trial-and-error procedure which is a time-consuming iterative process. For the selection of a NN-based model, the influence of efficiency has been neglected in the past. An important question facing hydrologists in inundation forecasting is the efficiency of NN-based models. Therefore, to provide effective warnings for emergency evacuation decisions, there is justification to develop a well-performing model that can be trained rapidly and can extend the forecast lead time.

Support vector machines (SVMs) have been used for hydrologic time series forecasting (Liong & Sivapragasam 2002; Yu *et al.* 2004; Sivapragasam & Liong 2005; Lin *et al.* 2006; Yu & Liong 2007; Wu *et al.* 2008; Wang *et al.* 2009; Lin *et al.* 2010; Maity *et al.* 2010). For both rainfall and runoff forecasting, Lin *et al.* (2009a, b) found that SVM-based models are capable of producing accurate forecasts especially for longer lead time forecasts. Furthermore, the determination of the architecture and weights of SVM-based models are expressed in terms of a quadratic optimization problem which can be rapidly solved by a standard programming algorithm. Due to their high accuracy and efficiency, SVMs could be more rapidly retrained with real-time data and are more suitable to be integrated with disaster warning systems.

In this paper, a real-time regional inundation forecasting model based on clustering techniques and SVM is proposed to yield 1- to 3-h lead time inundation maps. In the first step, the classification module is developed to group the inundation depths of the inundated area into several clusters and to indentify the control points. Then, the rainfall and inundation data are used as input to develop the point forecasting module in the second step. Finally, based on the control point forecasts and the geographic information (such as elevation, coordinates, and rainfall) of the forecast grids, the spatial expansion module is developed to expand the point forecasts to the spatial forecasts. The proposed modeling technique can further improve the accuracy of the forecasted inundation map.

To assess the performance of the proposed model, an actual application to Siluo Township, which is an urban township in Yunlin County, Taiwan, is conducted to demonstrate the superiority of the proposed model. Siluo Township frequently suffers inundation damage due to torrential rain during the summer monsoon season (May–October) (Pan *et al.* 2011). Chen *et al.* (2006) have indicated that under

the same rainfall conditions, Yunlin County has larger inundated areas than other counties. Furthermore, Yunlin County is one of the demonstration sites of the Regulation Project of Flood-Prone Areas in Taiwan (Yang *et al.* 2011). Hence, Siluo Township has been selected as the study area in this paper. To reach just conclusions, cross validations are conducted to evaluate the overall performance of the model. The accuracy and the efficiency of the proposed model are discussed in depth to demonstrate its superiority more clearly. This paper is organized in the following manner. Following the Introduction, the details of the proposed model are described. In the Application section, the study area and data, cross validation and the performance indices are presented. Then, the performance of the proposed model is described in the Results and Discussion section. The last section presents the Summary and Conclusions.

## THE PROPOSED MODEL

A flowchart of the proposed model is shown in Figure 1. The proposed model is composed of three steps: classification, point forecasting, and spatial expansion. Details of these three steps are described as follows.

### The classification step

In order to determine the regional inundation map from a real-time inundation forecasting model, the inundation depths of the study area are clustered into several groups. For an efficiency model, the identification of different levels of inundation is an important issue. First, the inundated and non-inundated areas have to be identified. According to the Taiwan Government's Standards of Disaster Relief, an area with an inundation depth below 0.5 cm is regarded as non-inundated and no relief assistance is provided.

To group those inundation depths of inundated areas with the same statistical characteristics, the *K*-means clustering algorithm is used. The *K*-means clustering algorithm, which is one of the most frequently used clustering techniques, was proposed by MacQueen (1967). The symbol $K$ refers to the number of cluster centers used to classify the data sets. The inundation depths are the input to the *K*-means clustering algorithm, which is used to determine the different levels of the inundation depths (Chang *et al.* 2010). Hence, the inundation depths with specific different characteristics are identified, and the data points with minimum Euclidean distance with the cluster center $\boldsymbol{\mu}_t$ are recognized as the control points for each cluster. The function of the *K*-means clustering algorithm is written as:

$$\text{minimize} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \left\| \mathbf{x}_j - \boldsymbol{\mu}_i \right\| \tag{1}$$

where $k$ is the number of clusters, $\mathbf{x}_j$ is the $j$th data, and $\boldsymbol{\mu}_i$ is the mean of data in cluster $S_j$. Examples of the *K*-means clustering algorithm used in hydrologic regionalization studies have been reported in the literature (Lin *et al.* 2005; Satyanarayana & Srinivas 2008; Meshgi & Khalili 2009; Saf 2009; Nourani & Kalantari 2010; Nourani *et al.* 2012). For more details regarding the *K*-means clustering algorithm, refer to Johnson & Wichern (2002).

Determination of an optimal number of clusters is a difficult task. In order to obtain a satisfactory clustering result, various tests (Grover & Vriens 2006) are introduced to help determine the number of clusters which is one of the major issues in cluster analysis. Two indices are calculated in this paper and then are plotted against the number of clusters to determine the optimal number of clusters (Isik & Singh 2008). Therefore, by calculating these two indices, an optimal number of clusters can be objectively determined. These two indexes are the *R*-squared (RSQ) and semi-partial *R*-squared (SPRSQ):

1. RSQ:

$$\text{RSQ} = \frac{\text{SS}_b}{\text{SS}_t} \tag{2}$$

   where $\text{SS}_b$ refers to the sum of squares between different clusters, and $\text{SS}_t$ is the total sum of squares of the whole data.
2. SPRSQ:

$$\text{SPRSQ} = \frac{\left(\text{SS}_{w,k} - \text{SS}_{w,k+1}\right)}{\text{SS}_t} \tag{3}$$

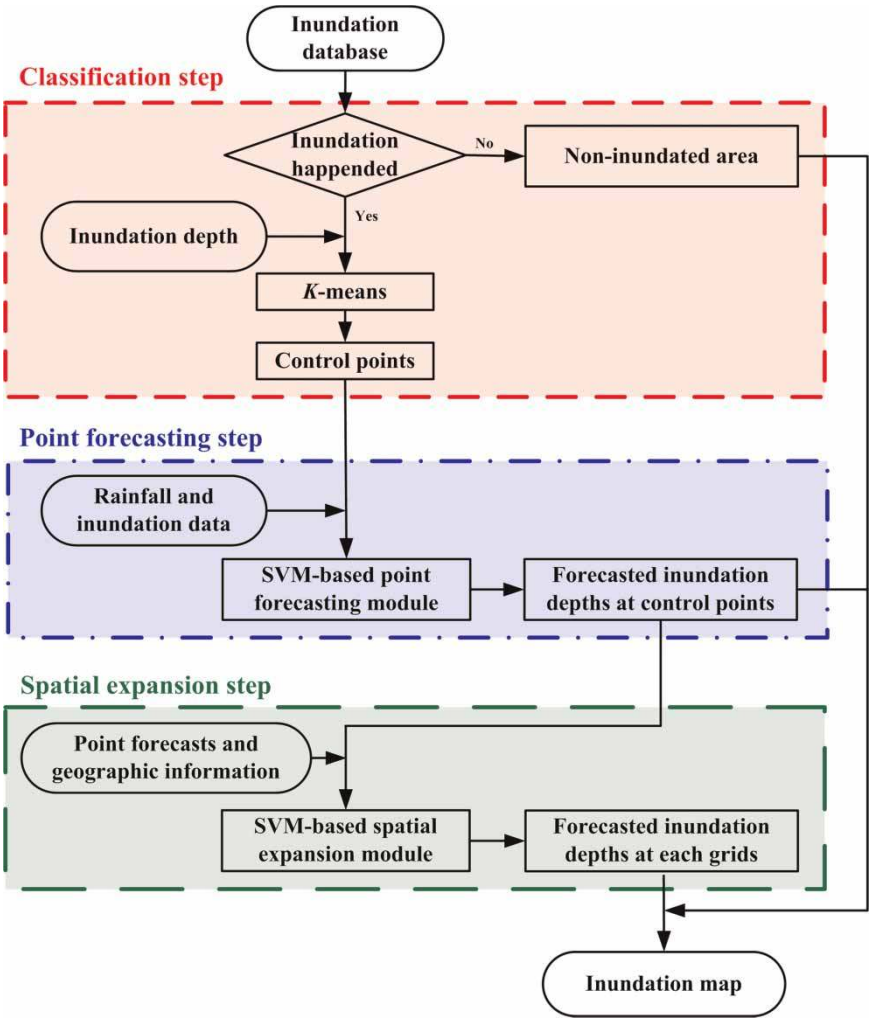


**Figure 1** | Flowchart of the proposed model.

where $SS_{w,k}$ and $SS_{w,k+1}$ are sums of squares within clusters $k$ and $k+1$, respectively.

## The point forecasting step

The second step (the point forecasting step) is to construct the rainfall-inundation point forecasting module for each control point. According to the results from the classification step, the control points of each cluster are identified. For each control point, the rainfall and the antecedent inundation depth are used as inputs to develop the rainfall-inundation point forecasting module. The point forecasting modules can be written in a general form as:

$$D_{\mathrm{CP},t+\Delta t} = f_{\mathrm{point}}(D_{\mathrm{CP},t}, D_{\mathrm{CP},t-1} \cdots, D_{\mathrm{CP},t-(L_D-1)}, R_{\mathrm{CP},t}, R_{\mathrm{CP},t-1}, \cdots, R_{\mathrm{CP},t-(L_R-1)}) \quad (4)$$

where $t$ is the current time, $\Delta t$ is the lead time period (from 1- to 3-h), $D_{\mathrm{CP},t}$ is the inundation depth of the control point at time $t$, $L_D$ denotes the lag length of inundation, $R_{\mathrm{CP},t}$ is the rainfall depth of the control point at time $t$, $L_R$ denotes the lag length of rainfall, and $D_{\mathrm{CP},t+\Delta t}$ is the point forecasted inundation depth at time $t+\Delta t$.

In the point forecasting module's construction, determination of the appropriate lag lengths of input is an important

work. Different lag lengths of input will influence the capability of NN-based models. Lin *et al.* (2009b) shows the suitability of the relative percentage error (RPE) as a criterion for selecting the lag lengths in hydrological time series. The RPE is expressed as:

$$\mathrm{RPE} = \frac{E(L) - E(L+1)}{E(L)} \times 100 \tag{5}$$

where $E(L)$ and $E(L+1)$ are the root mean square errors (RMSEs) for modules with $L$ and $L+1$ lag lengths, respectively. In general, the RMSE decreases with increasing lag term. When the RPE is less than 5%, the increase of lag lengths is stopped and the best inputs of forecasting modules are selected.

The development of point forecasting modules is to construct a nonlinear function approximator which can map the inputs into the desired output. In this paper, the SVM is used as the computational method. SVMs are firstly developed for classification and then expanded for nonlinear regression. The methodology of the support vector regression (SVR) used in this paper is briefly reviewed in this subsection. More mathematical details about SVR can be found in several textbooks (e.g., Vapnik 1995, 1998; Cristianini & Shaw-Taylor 2000).

The objective of the SVR is to find a regression function to yield the output $\hat{y}$, which is the best approximate of the desired output $y$ with an error tolerance of $\varepsilon$. The input vector $\mathbf{x}$ is mapped to a higher dimensional feature space by a nonlinear function $\phi(\mathbf{x})$. The regression function can then be written as:

$$\hat{y} = \mathbf{w}^T \phi(\mathbf{x}) + b \tag{6}$$

where $\mathbf{w}$ and $b$ are weights and bias of the regression function, respectively.

Instead of the empirical risk minimization (ERM) which is to minimize the empirical risk (i.e., training error), the structural risk minimization (SRM) induction principle is used to construct SVM. According to the SRM induction principle, the learning objective of a SVM is to minimize both the empirical risk and the model complexity. The use of the SRM induction principle results in a better generalization ability of SVM and avoids over-training of the model. Based on the SRM induction principle, $\mathbf{w}$ and $b$ are estimated by minimizing the following structural risk function (Vapnik 1995):

$$R = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N_d} L_\varepsilon(\hat{y}_i) \tag{7}$$

where $L_\varepsilon$ is the Vapnik's $\varepsilon$-insensitive loss function. The typical loss function $L_\varepsilon$ with an error tolerance of $\varepsilon$ is defined as (Vapnik 1995):

$$L_\varepsilon(\hat{y}) = \begin{cases} 0 & \text{if } |\hat{y} - y| < \varepsilon \\ |\hat{y} - y| - \varepsilon & \text{if } |\hat{y} - y| \geq \varepsilon \end{cases} \tag{8}$$

The SVR problem can be expressed as an optimization problem (Vapnik 1995):

$$\text{minimize } \mathrm{R}\left(\mathbf{w}, b, \xi, \xi'\right) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N_d}\left(\xi_i + \xi_i'\right) \tag{9a}$$

subject to:

$$\begin{aligned} &y_i - \hat{y}_i = y_i - \left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) \leq \varepsilon + \xi_i \\ &\hat{y}_i - y_i = \left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) - y_i \leq \varepsilon + \xi_i' \\ &\xi_i \geq 0,\ i = 1, 2, \cdots, N_d \\ &\xi_i' \geq 0, i = 1, 2, \cdots, N_d \end{aligned} \tag{9b}$$

where $\xi$ and $\xi'$ are slack variables representing the upper and the lower training errors, respectively. The optimization problem can be solved in its dual form using Lagrange multipliers method. Rewriting Equation (4) in its dual form and differentiating with respect to the primal variables $(\mathbf{w}, b, \xi, \xi')$ yields:

$$\begin{aligned} \text{maximize } &\sum_{i=1}^{N_d} y_i(\alpha_i - \alpha_i') - \varepsilon\sum_{i=1}^{N_d}(\alpha_i - \alpha_i') \\ &-\frac{1}{2}\sum_{i=1}^{N_d}\sum_{j=1}^{N_d}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \end{aligned} \tag{10a}$$

subject to:

$$\begin{aligned} &\sum_{i=0}^{N_d}\left(\alpha_i - \alpha_i'\right) = 0 \\ &0 \leq \alpha_i \leq C,\ i = 1, 2, \cdots, N_d \\ &0 \leq \alpha_i' \leq C,\ i = 1, 2, \cdots, N_d \end{aligned} \tag{10b}$$

where $\alpha$ and $\alpha'$ are the dual Lagrange multipliers. The optimal Lagrange multipliers $\alpha^*$ are solved by the standard quadratic programming algorithm. Some of the solved Lagrange multipliers are zero and should be eliminated from the regression function. Hence, the regression function can be expressed in terms of only the nonzero Lagrange multipliers and the corresponding input vectors of the training data (called support vectors). The final regression function becomes:

$$f(\mathbf{x}) = \sum_{k=1}^{N_{sv}} \alpha_k K(\mathbf{x}_k, \mathbf{x}) + b \tag{11}$$

where $N_{sv}$ is the number of support vectors, $\mathbf{x}_k$ is the $k$ th support vector and $K(\mathbf{x}_i,\mathbf{x})$ is the kernel function. The radial basis function is adopted herein.

## The spatial expansion step

The third step (the spatial expansion step) is to expand the forecasted results from point to region. In recent years, the NN-based models have been used in spatial expansion. Lopez *et al.* (2008) proposed a NN-based model to estimate the average wind speed at the selected site based on data from nearby stations, and found that the NN-based model provides a reliable estimate with an error below 2%. Chung *et al.* (2012) presented the reliability of the NN-based model in estimating the spatial distribution of evaporation by taking topography characteristics into consideration. The aforementioned studies prompted us to develop the spatial expansion module by using NN-based models.

To develop an accurate spatial expansion module, the SVM is also used as the computational method. A graphical illustration of the development of the spatial expansion module is presented in Figure 2. As shown in Figure 2, the inundation data are pre-analyzed by the $K$-means clustering algorithm for identifying the control points and the category of each grid. Then, for each cluster, the spatial expansion module is constructed using the data consisting of the forecasted depths at the control points (obtained from the point forecasting step) and the geographic information (such as elevation, coordinates, and rainfall) of the forecast grids. Finally, the inundation depths of corresponding grids are forecasted by the spatial expansion module for each cluster. The spatial expansion modules can be written in a general form as:

$$D_{n,t+\Delta t} = f_{\text{spatial}}\left(D_{\text{CP},t+\Delta t}, E_n, X_n, Y_n, R_n\right) \tag{12}$$

where $E_n$ is the elevation of grid $n$, $X_n$ and $Y_n$ are the coordinates of grid $n$, $R_n$ is the rainfall of grid $n$, and $D_{n,t+\Delta t}$ is the forecasted inundation depth of grid $n$ at time $t+\Delta t$. Once

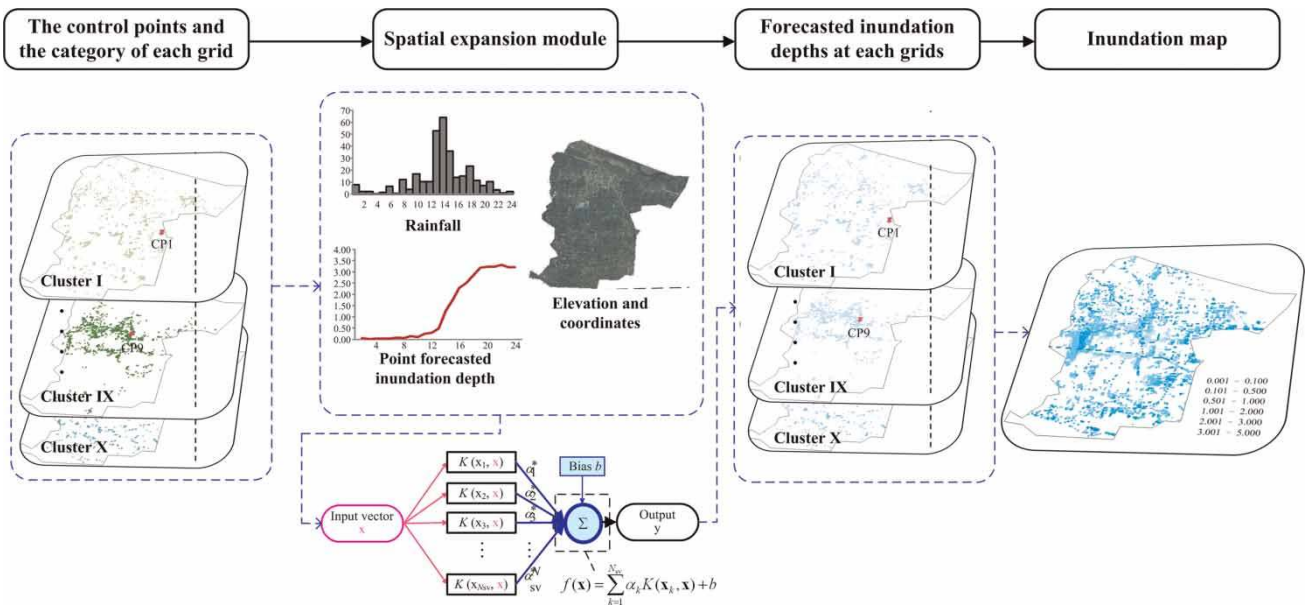


**Figure 2** | Graphical illustration of the spatial expansion module.

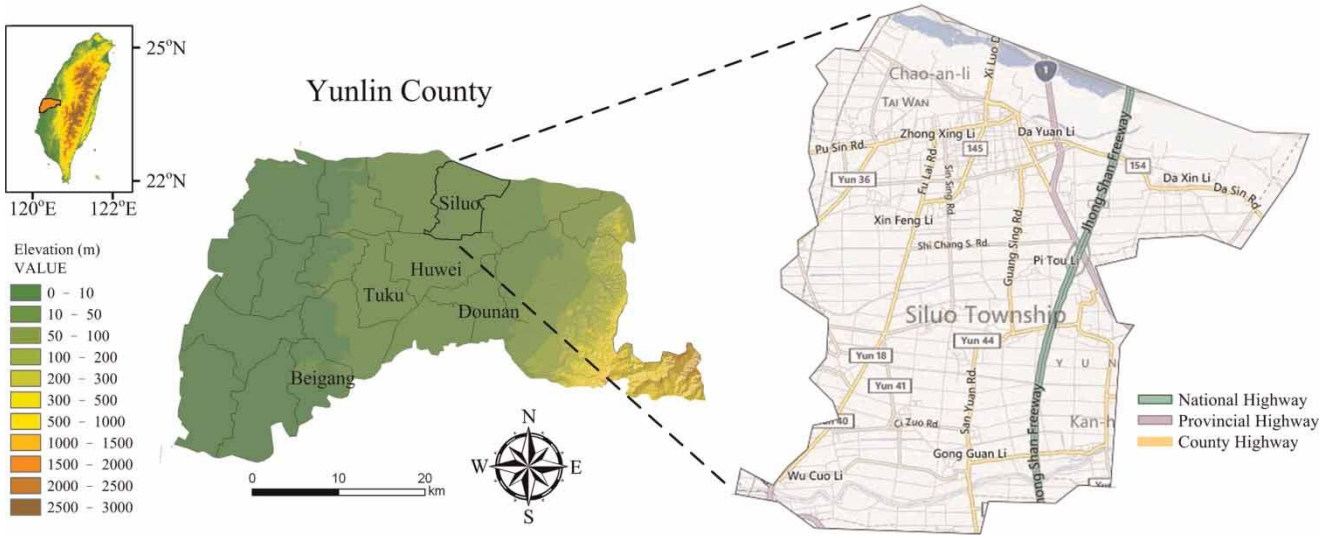


**Figure 3** | The study area.

the forecasted results of all the inundated grids in the study area are obtained, the inundation map is then constructed.

## APPLICATION

### The study area and data

The study area is Siluo Township, which is an urban township with a population of about fifty thousand people, and the Zhuoshui River (the longest in Taiwan) passes through the northern part of the study area. The study area (Figure 3) has an area of 49.8 km$^2$ and an average slope of 4°. Although the average annual rainfall of the study area is about 1,110–1,250 mm, the temporal distribution of rainfall is uneven. The rainy season (from April through to August) produces about 75% of the average annual rainfall to the study area. Heavy typhoon rainfall and the topography characteristics (low-elevation and gentle slope) have frequently caused inundation disasters in Siluo Township. Therefore, a well-performing and efficient real-time inundation forecasting model is needed for the study area.

The rainfall data were obtained from the Taiwan Water Resources Agency. The database of inundation used in the model development was published by the Taiwan Water Resources Agency, and the inundation depths in the database were simulated by a 2-D overland-flow model, which was calibrated and validated using survey data of inundation extent and depths (Chang *et al*. 2010; Pan *et al*. 2011). Rainfall and inundation data are hourly. Typhoon events with rainfall and inundation data available simultaneously were collected. Hence, a total of seven typhoon events was used to establish the forecasting model herein. Figure 4 presents the typhoon tracks used in this paper. Table 1 summarizes the date of occurrence and the maximum 24-h rainfall of these seven typhoon events. The topographic data were obtained from the Taiwan Ministry of the Interior, and the spatial resolution of DEM (Digital Elevation Model) data is 40 m.

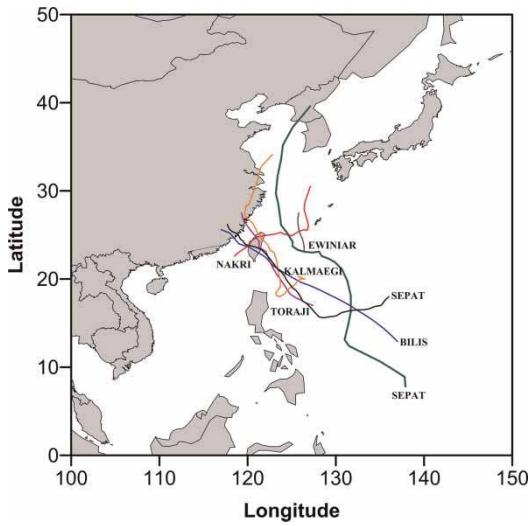


**Figure 4** | The tracks of typhoons.

**Table 1** | Descriptions of typhoon events used in the modeling

| Number | Name | Date (yyyy/mm/dd) | Maximum 24-h rainfall (mm) |
|---|---|---|---|
| 1 | Bilis | 2000/8/21 | 75.51 |
| 2 | Prapiroon | 2000/8/27 | 31.64 |
| 3 | Toraji | 2001/7/28 | 370.04 |
| 4 | Nakri | 2002/7/9 | 78.09 |
| 5 | Ewiniar | 2006/7/7 | 30.66 |
| 6 | Sepat | 2007/8/16 | 209.55 |
| 7 | Kalmaegi | 2008/7/16 | 315.51 |

### Cross validation and performance indices

During the construction of the proposed model, the collected events are usually partitioned into two sets: training and testing. Different selections of training and testing data yield different results and sometimes lead to different conclusions. For evaluating the performance of the accuracy and the robustness of the proposed model, a statistical technique called cross validation is conducted. Each single typhoon event (except Typhoon Toraji) is used to test the proposed models in turn. Typhoon Toraji yielded the maximum rainfall and the maximum inundation depth, and hence it should be used as training data. Then, conclusions are drawn based on the overall performance for the six testing events. Two performance indices that are commonly used to evaluate the model performance are employed:

1. RMSE:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\left(\hat{D}_t - D_t\right)^2} \tag{13}$$

where $D_t$ and $\hat{D}_t$ denote the inundation data and the forecasted depth at time $t$, respectively, and $m$ is the number of forecasts.

2. Correlation coefficient (CC):

$$\mathrm{CC} = \frac{\sum_{t=1}^{m}\left(D_t - \overline{D}\right)\left(\hat{E}_t - \overline{\hat{D}}\right)}{\sqrt{\sum_{t=1}^{m}\left(D_t - \overline{D}\right)^2 \sum_{t=1}^{m}\left(\hat{D}_t - \overline{\hat{D}}\right)^2}} \tag{14}$$

where $\overline{\hat{D}}$ is the average of the forecasted depth.

The RMSE is used to measure the difference between the inundation data and the forecasted depth. The smaller the RMSE value, the better the forecasts. In contrast, the CC is used to measure the similarity between the observed and the forecasted depths. The smaller the CC value, the worse the forecasts.

## RESULTS AND DISCUSSION

In the first subsection, characteristics of clusters identified by the classification module are discussed. In the second subsection, the performance of 1- to 3-h lead time point forecasts is investigated. In the third subsection, the performance of 1- to 3-h lead time spatial forecasts is presented. Finally, the forecasting performance for two typhoon events, Nakri in 2002 and Kalmaegi in 2008, is discussed in depth.

### Characteristics of clusters identified by the classification step

The characteristics of the inundation depth within each homogeneous cluster identified by the classification module are discussed in this section. The inundation depths are the input to the *K*-means clustering algorithm which is used to obtain the relative information of each grid. In this paper, the determination of the number of clusters is based on RSQ and SPRSQ. When the *K*-means clustering algorithm is performed, the clustering result of grids is obtained and the control point of each cluster is identified.

The distribution of the grid clustering result in the study area is presented in Figure 5(a), and the location of each control point is presented in Figure 5(b). The characteristics of inundation and the number of grids in each cluster are summarized in Table 2. As shown in Figure 5(a), the main inundated area is located in the western and central region of the study area. The inundation depths at each control point are also identified (Figure 6). As shown in Figure 6, the inundation depths are implicitly grouped according to their peak depths and the increasing rate of inundation. These clusters have distinct types of peak inundation depths, namely, Clusters 1–4 in the

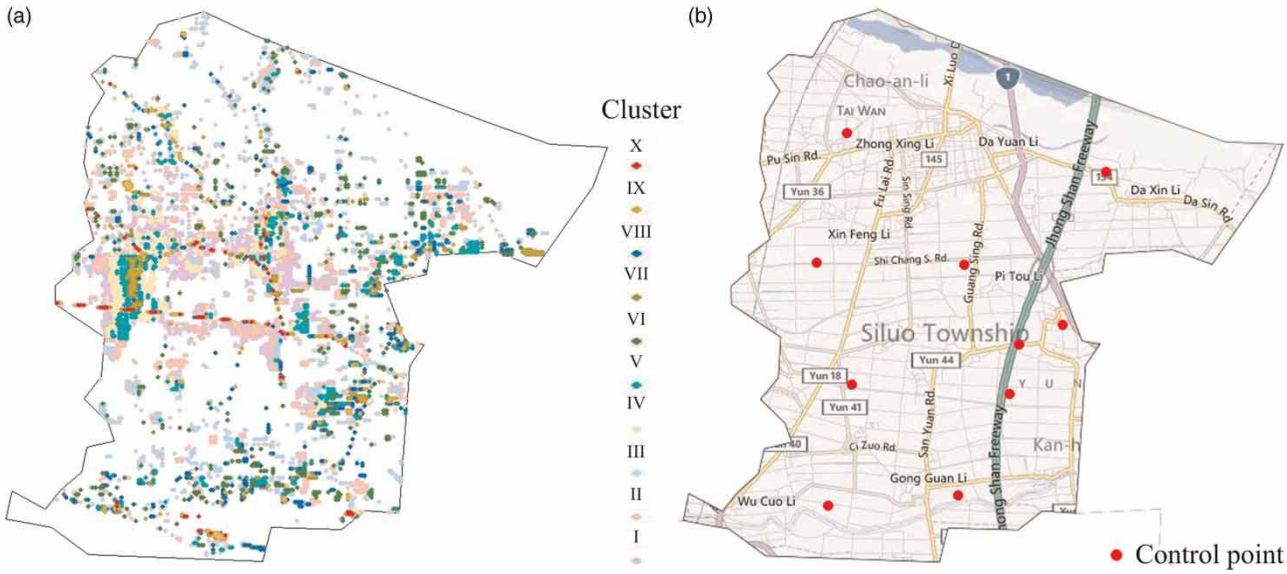


**Figure 5** | (a) The results of data clustering based on *K*-means and (b) locations of control points.

**Table 2** | The characteristics of inundation for each cluster

| Cluster | Number of grids | Percentage (%) | Maximum depth (m) | Mean depth (m) |
|---|---|---|---|---|
| I | 1,279 | 20.6 | 0.66 | 0.18 |
| II | 1,290 | 20.8 | 0.59 | 0.25 |
| III | 862 | 13.9 | 0.76 | 0.47 |
| IV | 820 | 13.2 | 0.77 | 0.33 |
| V | 638 | 10.3 | 1.11 | 0.46 |
| VI | 516 | 8.3 | 1.19 | 0.65 |
| VII | 253 | 4.1 | 1.49 | 0.66 |
| VIII | 300 | 4.8 | 1.60 | 0.90 |
| IX | 173 | 2.8 | 2.27 | 1.24 |
| X | 78 | 1.3 | 3.39 | 1.90 |
| Total | 6,209 | 100.0 | | |

'$D_t < 1$ m' level of the inundation depth, Clusters 5–7 in the '1 m $< D_t <$ 2 m' level, and Clusters 8–10 in the '$D_t > 2$ m' level. Moreover, the increasing rate of 10 inundation depths is different. For example, although Clusters 1 and 2 are in the '$< 1$ m' level, the inundation increasing rate of Cluster 2 is greater than that of Cluster 1 at time 10–14 h. A conclusion can also be drawn that the peak depth and the increasing rate of inundation have significant influence on the classification of the inundation depths.

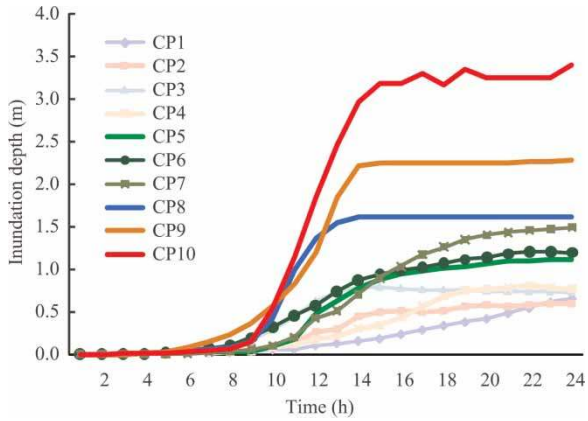


**Figure 6** | The inundation depth versus time at each control point.

## Performance of the point forecasting module

In this subsection, the forecasting performance of the point forecasting module is discussed. In this paper, the source code of the SVM model is from Chang & Lin (2001). The parameters of SVM are set as follows. Parameter $C$ (the trade-off between the model complexity and the empirical error) is set to 1 herein. It means that the model complexity is as important as the empirical error. In addition, it is acceptable to set the error tolerance $\varepsilon$ of 1% and kernel parameter $\gamma$ of 0.017 for inundation forecasting. The lag lengths of rainfall $L_R$ and inundation $L_D$ are selected as 1 and 2, respectively.

Figures 7 and 8 present the point forecasting results for Typhoons Kalmaegi and Nakri, respectively. Comparisons of the inundation data with the forecasts resulting from the point forecasting module are presented in Figures 7(a) and 8(a); the scatter plots for the inundation data and the forecast results in Figures 7(b) and 8(b). As shown in Figures 7 and 8, the forecasted inundation yielded by the proposed point forecasting module is in good agreement with the inundation data.

The performance indices of the point forecasting module for Typhoons Kalmaegi and Nakri are presented in Tables 3 and 4, respectively. As shown in Tables 3 and 4, for each control point, the RMSE values increase with increasing forecast lead time, and the RMSE values decrease with the decreasing number of control points. It is quite reasonable that the RMSE values of the control point CP10 are higher than those of CP1 for 1- to 3-h lead time forecasts, because CP10 and CP1 are the control points with the maximum and the minimum inundation depths, respectively. The RMSE values increase from 0.083 to 0.272 m for 1- to 3-h lead time forecasts for Typhoon Kalmaegi; from 0.057 to 0.132 m, for Typhoon Nakri. We note that in this region, the maximum inundation depths are 3.29 and 1.17 m for Typhoons Kalmaegi and

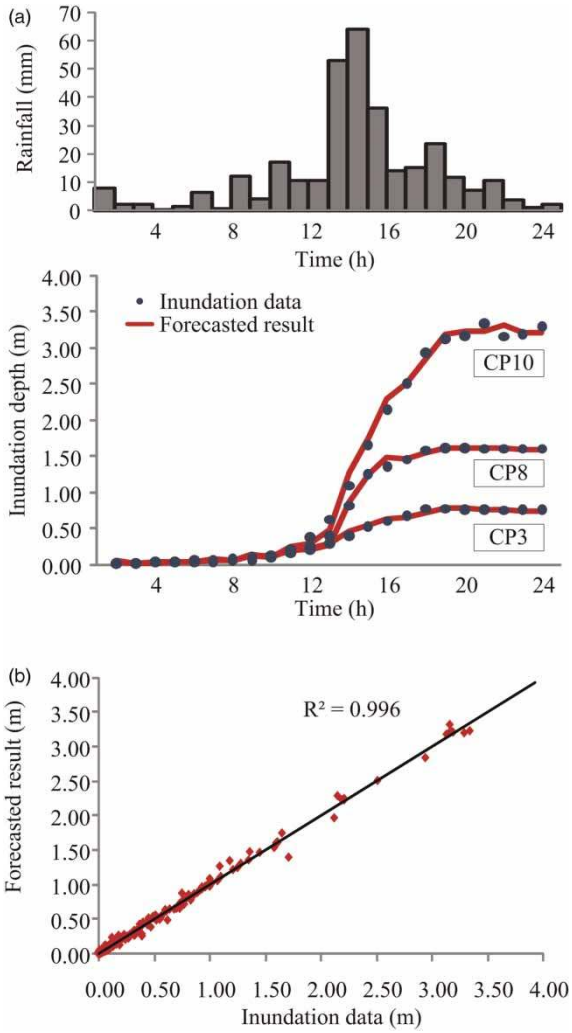


**Figure 7** | (a) The rainfall-inundation depth of Typhoon Kalmaegi (for CP3, CP8, CP10) and (b) the scatter plot of the inundation data and forecasted inundation depth resulting from the point forecasting module (for 10 control points).

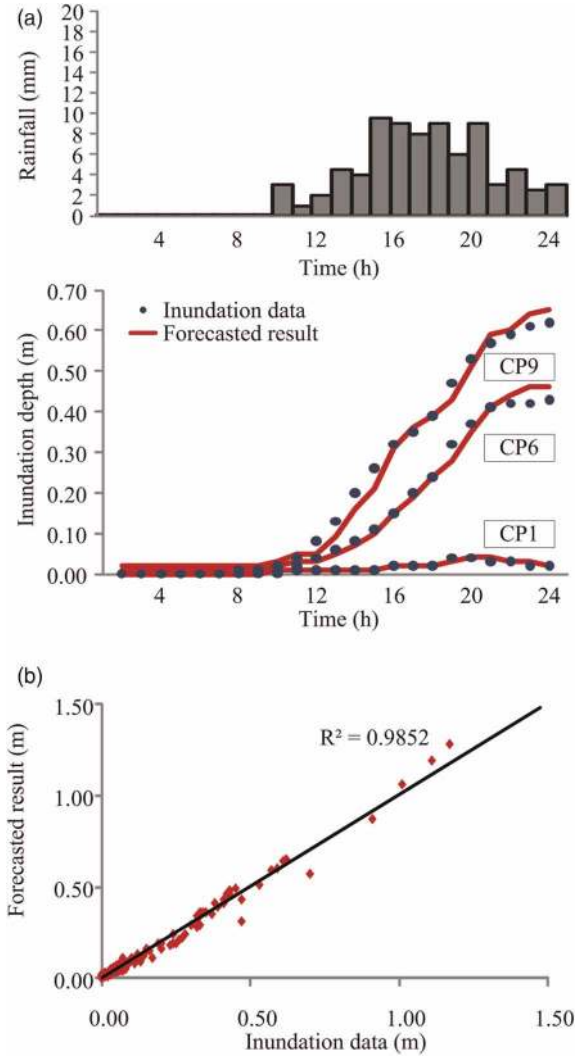


**Figure 8** | (a) The rainfall-inundation depth of Typhoon Nakri (for CP1, CP6, CP9) and (b) the scatter plot of the inundation data and forecasted inundation depth resulting from the point forecasting module (for 10 control points).

**Table 3** | Root mean square error (RMSE) and correlation coefficient (CC) of Typhoon Kalmaegi (2008) for the point forecasting module

| | Control point | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Lead time (h)** | **CP1** | **CP2** | **CP3** | **CP4** | **CP5** | **CP6** | **CP7** | **CP8** | **CP9** | **CP10** |
| | RMSE (m) | | | | | | | | | |
| 1 | 0.012 | 0.032 | 0.024 | 0.015 | 0.036 | 0.027 | 0.029 | 0.042 | 0.092 | 0.083 |
| 2 | 0.022 | 0.047 | 0.051 | 0.030 | 0.031 | 0.058 | 0.040 | 0.123 | 0.163 | 0.170 |
| 3 | 0.032 | 0.075 | 0.067 | 0.056 | 0.077 | 0.073 | 0.086 | 0.220 | 0.222 | 0.272 |
| | CC | | | | | | | | | |
| 1 | 0.994 | 0.989 | 0.997 | 0.998 | 0.996 | 0.998 | 0.998 | 0.998 | 0.995 | 0.998 |
| 2 | 0.983 | 0.976 | 0.986 | 0.993 | 0.997 | 0.989 | 0.997 | 0.985 | 0.983 | 0.993 |
| 3 | 0.969 | 0.933 | 0.977 | 0.979 | 0.982 | 0.983 | 0.984 | 0.953 | 0.970 | 0.981 |

**Table 4** | Root mean square error (RMSE) and correlation coefficient (CC) of Typhoon Nakri (2002) for the point forecasting module

| | Control point | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Lead time (h)** | **CP1** | **CP2** | **CP3** | **CP4** | **CP5** | **CP6** | **CP7** | **CP8** | **CP9** | **CP10** |
| | RMSE (m) | | | | | | | | | |
| 1 | 0.004 | 0.006 | 0.016 | 0.008 | 0.010 | 0.017 | 0.021 | 0.023 | 0.025 | 0.057 |
| 2 | 0.009 | 0.011 | 0.024 | 0.009 | 0.017 | 0.026 | 0.033 | 0.035 | 0.051 | 0.104 |
| 3 | 0.010 | 0.015 | 0.032 | 0.011 | 0.020 | 0.039 | 0.040 | 0.041 | 0.066 | 0.132 |
| | CC | | | | | | | | | |
| 1 | 0.951 | 0.973 | 0.993 | 0.977 | 0.913 | 0.995 | 0.991 | 0.991 | 0.995 | 0.990 |
| 2 | 0.863 | 0.926 | 0.985 | 0.972 | 0.749 | 0.988 | 0.972 | 0.979 | 0.979 | 0.967 |
| 3 | 0.808 | 0.893 | 0.969 | 0.959 | 0.723 | 0.973 | 0.912 | 0.970 | 0.961 | 0.949 |

Nakri, respectively. For both typhoon events, the RMSE values of 1- to 3-h lead time forecasts are all lower than 0.3 m, which indicates a high accuracy of the point inundation forecasts.

The overall performance indices, which are drawn on the basis of all testing events, are summarized in Table 5. For the 3-h lead time forecasts, the RMSE and CC values of the proposed point forecasting module for CP10 are 0.191 and 0.806, respectively. Thus, for the 3-h lead time forecasts, the performance of the proposed point forecasting module is still acceptable. It is seen that the proposed point forecasting module produced accurate point inundation forecasts up to 3 h, and the use of the proposed point forecasting module effectively decreases the negative impact of increasing forecast lead time, which is consistent with the conclusion of Lin *et al.* (2009a, b). It is shown that the proposed point forecasting module effectively improves the forecasting performance.

## Performance of the spatial expansion module

In this section, the performance of the proposed module in spatial forecasting is discussed. To quantify the module performance, the RMSE and CC are also employed. The overall performance indices, which are drawn on the basis of all testing events, are summarized in Table 6. As shown in Table 6, for each cluster, the RMSE values increase with increasing forecast lead time, and the RMSE values decrease with decreasing number of clusters. It is quite reasonable that the RMSE values of Cluster X are higher than those of Cluster I for 1- to 3-h lead time forecasts because Cluster X is a category with higher inundation depth than Cluster I. For the 1- to 3-h lead time forecasts in Cluster X, the RMSE values of the proposed module are 0.318, 0.332, 0.357 m, respectively. A reasonable explanation for this phenomenon is discussed herein. As the forecast lead time increases, the

**Table 5** | Root mean square error (RMSE) and correlation coefficient (CC) of six events for the point forecasting module

| | Control point | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Lead time (h)** | **CP1** | **CP2** | **CP3** | **CP4** | **CP5** | **CP6** | **CP7** | **CP8** | **CP9** | **CP10** |
| | RMSE (m) | | | | | | | | | |
| 1 | 0.008 | 0.017 | 0.020 | 0.015 | 0.025 | 0.018 | 0.025 | 0.038 | 0.051 | 0.071 |
| 2 | 0.013 | 0.028 | 0.036 | 0.021 | 0.036 | 0.036 | 0.042 | 0.080 | 0.098 | 0.125 |
| 3 | 0.019 | 0.043 | 0.053 | 0.032 | 0.054 | 0.050 | 0.063 | 0.122 | 0.141 | 0.191 |
| | CC | | | | | | | | | |
| 1 | 0.834 | 0.950 | 0.987 | 0.952 | 0.816 | 0.996 | 0.955 | 0.975 | 0.994 | 0.954 |
| 2 | 0.763 | 0.882 | 0.962 | 0.934 | 0.745 | 0.982 | 0.909 | 0.905 | 0.974 | 0.857 |
| 3 | 0.718 | 0.833 | 0.930 | 0.891 | 0.715 | 0.963 | 0.801 | 0.863 | 0.944 | 0.806 |

**Table 6** | Root mean square error (RMSE) and correlation coefficient (CC) of six events for the spatial expansion module

| | Cluster | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Lead time (h)** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** | **IX** | **X** |
| | RMSE (m) | | | | | | | | | |
| 1 | 0.043 | 0.063 | 0.084 | 0.055 | 0.075 | 0.108 | 0.095 | 0.174 | 0.273 | 0.318 |
| 2 | 0.045 | 0.066 | 0.088 | 0.056 | 0.077 | 0.114 | 0.102 | 0.183 | 0.278 | 0.332 |
| 3 | 0.048 | 0.074 | 0.100 | 0.065 | 0.082 | 0.124 | 0.115 | 0.196 | 0.290 | 0.357 |
| | CC | | | | | | | | | |
| 1 | 0.500 | 0.687 | 0.789 | 0.561 | 0.650 | 0.833 | 0.720 | 0.810 | 0.763 | 0.816 |
| 2 | 0.490 | 0.665 | 0.767 | 0.533 | 0.644 | 0.819 | 0.705 | 0.784 | 0.752 | 0.788 |
| 3 | 0.474 | 0.634 | 0.740 | 0.529 | 0.630 | 0.796 | 0.678 | 0.756 | 0.733 | 0.760 |

correlation between inputs and desired output decreases, and the data used for the long lead time spatial expansion module include more noise. Hence, the RMSE values increase for 1- to 3-h lead time forecasts. However, the proposed spatial expansion module provided reasonable spatial inundation forecasts for 1- to 3-h lead time forecasts, and the results indicate that the spatial expansion module was capable of dealing with the nonlinear relationships between inputs and desired output (the inundation depth of forecast grids $D_n$). As for the performance index of CC, similar conclusions are also drawn. For the 1- to 3-h lead time forecasts in Cluster X, the CC values are 0.816, 0.788, 0.760, respectively. The performance of the proposed spatial expansion module is still acceptable for longer lead times up to 3 h. Again, the results confirm that the proposed spatial expansion module produced acceptable spatial inundation forecasts.

Efficiency is another important issue for forecasting models. Based on statistical learning theory, the determination of the architecture and weights of SVM-based models can be rapidly solved by a standard programming algorithm. This advantage is very helpful in constructing efficient forecasting models with large amounts of data (over 6,000 inundated grids in this study area), especially for the spatial expansion module. To construct the spatial expansion module in this paper, we have tried different kinds of NN, BPN, RBFN, and self-organizing linear output (SOLO). It is shown that the spatial expansion module with SVM used as the computational method performs the best. Therefore, the proposed model could be more rapidly trained with real-time data and is more suitable to be integrated with the decision support system.

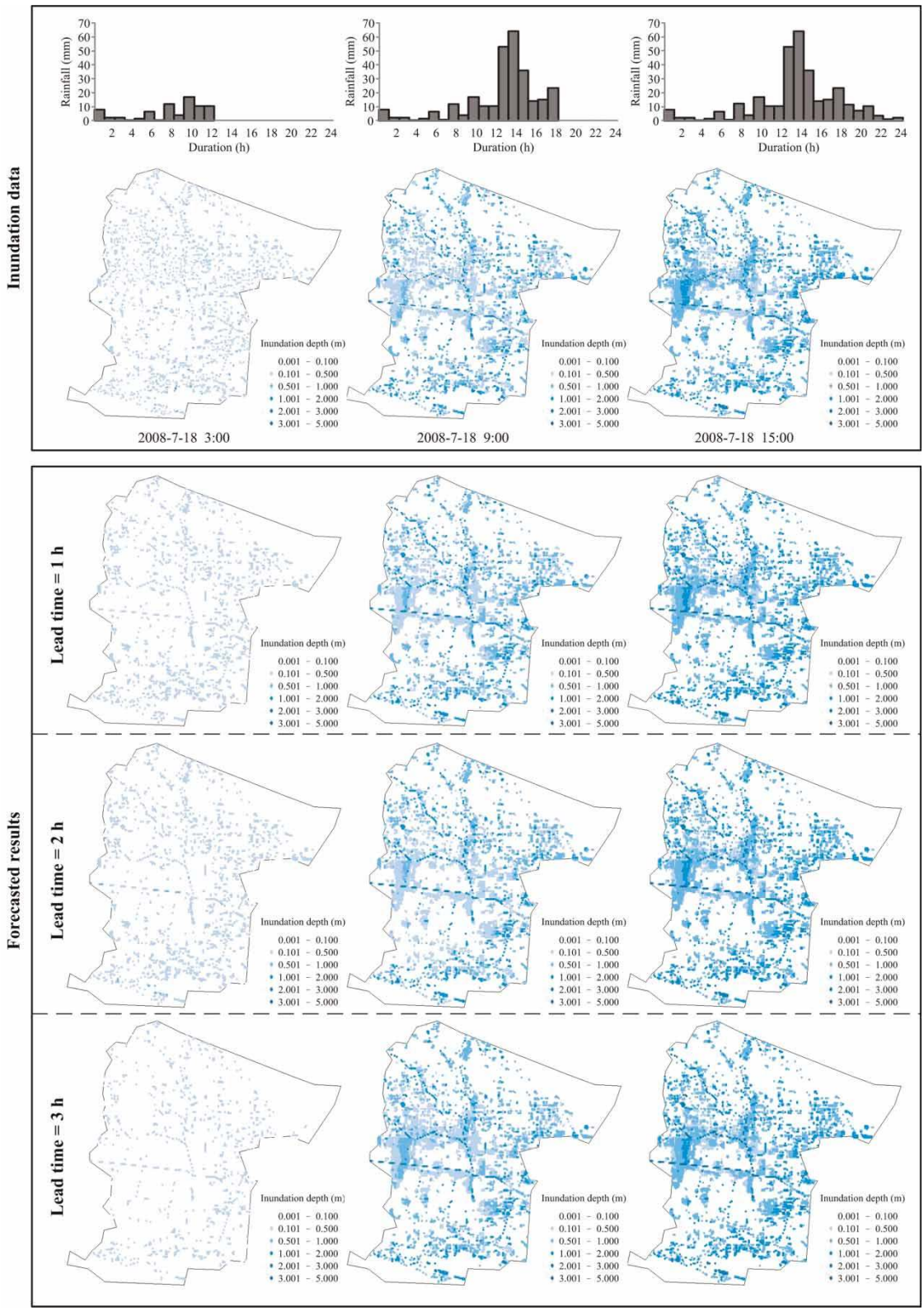


**Figure 9** | Comparison of the inundation data with the forecasts of the proposed model for Typhoon Kalmaegi.

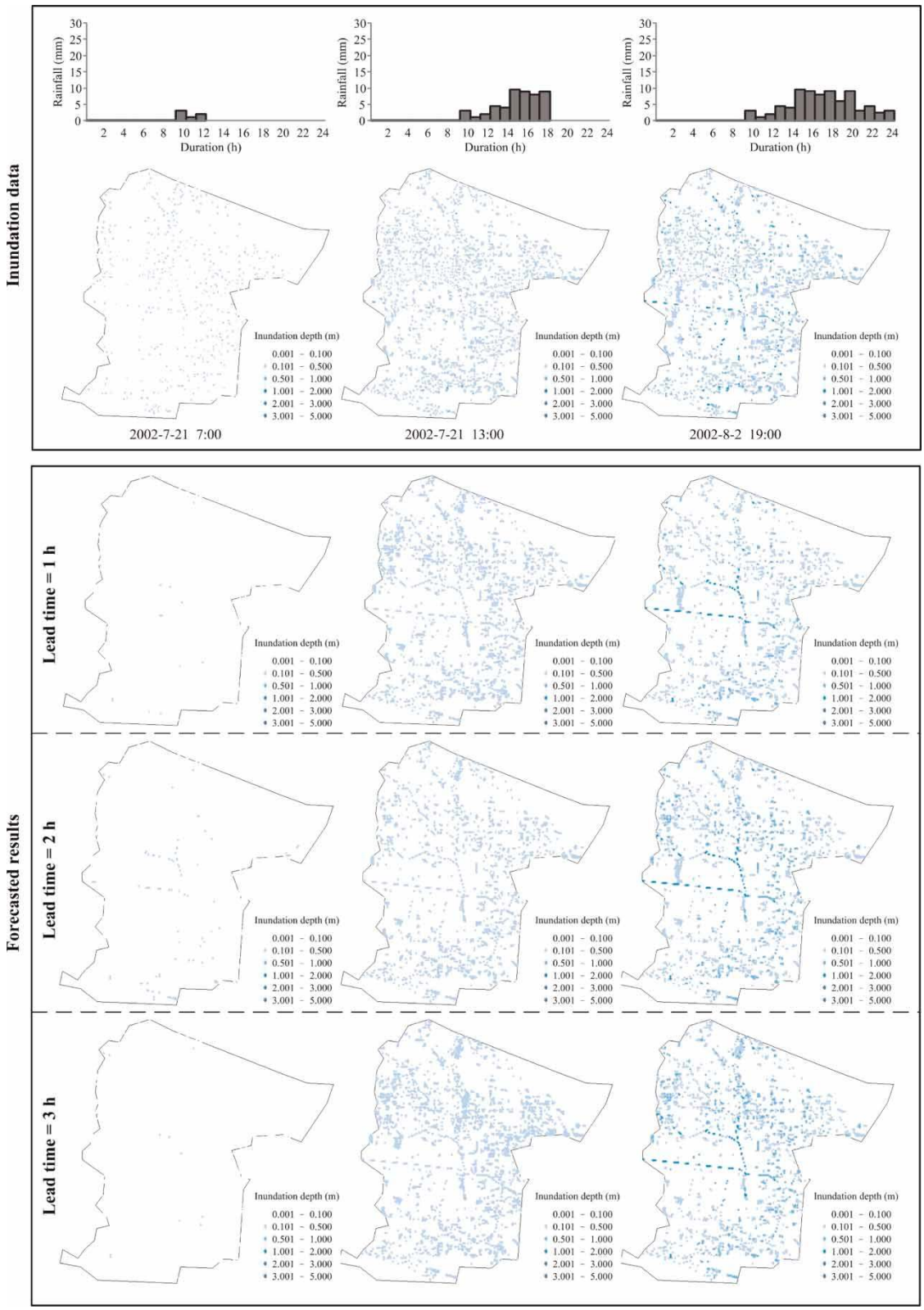


**Figure 10** | Comparison of the inundation data with the forecasts of the proposed model for Typhoon Nakri.

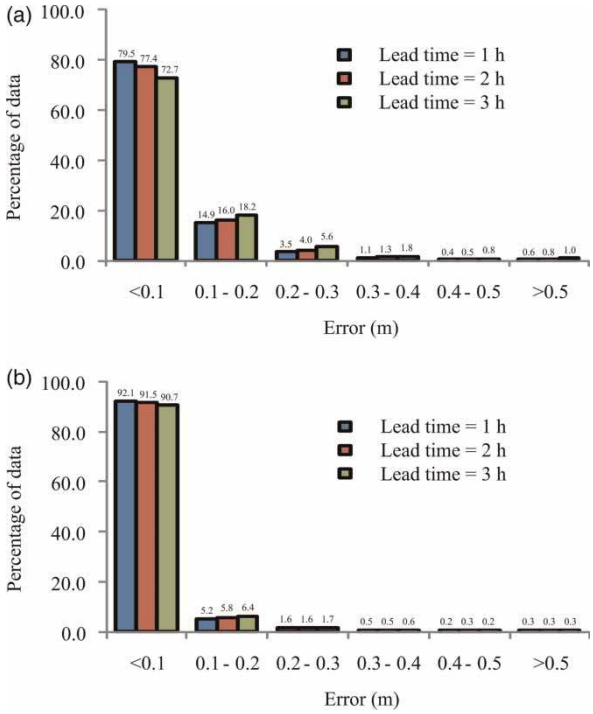


**Figure 11** | Histograms of forecasted error distributions at 1- to 3-h lead times for (a) Typhoon Kalmaegi and (b) Typhoon Nakri.

### Comparisons of the overall performance of the forecasting model

In this section, more performance comparisons are discussed in depth and these clearly demonstrate the superiority of the proposed model. To highlight the forecasting performance of the proposed model, two typhoon events (Typhoons Kalmaegi and Nakri) with the highest and the lowest inundation depths among all testing events are taken as examples. Typhoon Kalmaegi made landfall in Taiwan in the 2008 typhoon season; Typhoon Nakri in 2002. Typhoon Kalmaegi was upgraded to a tropical storm (35 knots) on 15 July 2008, reached peak intensity on 17 July 2008 with an observed maximum sustained wind of 75 knots, made landfall in northeast Taiwan (Ilan County) on 17 July 2008, and caused 19 deaths. Typhoon Nakri formed on 7 July 2002 in the South China Sea, and reached peak intensity on 10 July 2002 with an observed maximum sustained wind of 50 knots. When Typhoon Nakri passed over Taiwan, the accumulated rainfall reached 647 mm at Pengjia Islet.

The proposed model is performed to forecast the inundation depths for the study area. Figures 9 and 10 present the inundation data versus corresponding forecasts resulting from the proposed model at the 1- to 3-h lead times for Typhoon Kalmaegi and Typhoon Nakri, respectively. As shown in Figures 9 and 10, the difference between forecasts and inundation data increases with increasing forecast lead time. However, it is clearly observed that the forecasts from the proposed model are in good agreement with the inundation data.

Histograms of the forecasted error at the 1- to 3-h lead times for Typhoons Kalmaegi and Nakri are plotted in Figure 11. As shown in Figure 11, the proposed model has the highest percentage of data in the '$< 0.1$ m' level and the lowest percentage of data in the '$0.4\ \text{m} < D_t < 0.5$ m' level for these two typhoons. This phenomenon also indicates that most of the forecasted inundation depths resulting from the proposed model are in good agreement with the inundation data. These results demonstrate the accuracy and reliability of the proposed forecasting model to produce 1- to 3-h lead time inundation maps, and the forecasts results can provide information to assist the emergency response measures.

## SUMMARY AND CONCLUSIONS

The objective of this paper is to develop a well-performing and efficient real-time regional inundation depth forecasting model for inundation warning systems during typhoon-warning periods. For this purpose, we proposed a forecasting model which is composed of three steps: classification, point forecasting, and spatial expansion. In the first step, the cluster analysis of the inundation depths is first performed by the classification module. Then, in the second step, the rainfall and inundation data are used as input and SVM is used as the computational method to develop the point forecasting module to yield the inundation depth forecasts at control points. Finally, based on the point forecasts and the geographic information, the point forecasts are expanded to the spatial forecasts by using the spatial expansion module.

An application to Siluo Township in central Taiwan is performed to clearly demonstrate the superiority of the proposed model. First, according to the cluster resulting from the *K*-means clustering algorithm, it is observed that the inundation depths in the study area are grouped into 10 clusters. The peak depth and the increasing rate of inundation have significant influence on the classification. Then, the

performance of the point forecasting module is presented, which shows that the proposed module can accurately yield 1- to 3-h lead time point forecasts of inundation depth at each control point. Finally, for the spatial expansion module, the results show that the proposed module can produce acceptable inundation maps for 1- to 3-h lead times. It is concluded that the proposed modeling technique is suitable and useful for improving the regional inundation forecasting. In addition, another advantage of the proposed model is its efficiency, which is very important for a real-time inundation warning system. The efficiency of the proposed model is also confirmed, and hence the proposed model is more suitable to be integrated with the decision support system. However, the proposed model is also expected to be helpful in supporting inundation warning systems. In the future, instead of *K*-means and SVM, it will be possible to try other kinds of clustering algorithm and data mining techniques to examine whether the model performance can be improved.

## ACKNOWLEDGEMENTS

This paper is based on research partially supported by the National Science Council, Taiwan, under grant NSC 101-2625-M-002-007. We would also like to thank the Associate Editor and three reviewers for their constructive suggestions that greatly improved the manuscript.

## REFERENCES

Adamowski, J., Chan, H. F., Prasher, S. O. & Sharda, V. N. 2012 Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J. Hydroinform.* **14** (3), 731–744.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a Artificial neural networks in hydrology, I: Preliminary concepts. *J. Hydrol. Eng.* **5** (2), 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000b Artificial neural networks in hydrology, II: Hydrological applications. *J. Hydrol. Eng.* **5** (2), 124–137.

Bates, P. D. & De Roo, A. P. J. 2000 A simple raster-based model for flood inundation simulation. *J. Hydrol.* **236** (1–2), 54–77.

Carsell, K. M., Pingel, N. D. & Ford, D. T. 2004 Quantifying the benefit of a flood warning system. *Nat. Hazards Rev.* **5** (3), 131–140.

Chang, C. C. & Lin, C. J. 2001 LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chang, L. C., Shen, H. Y., Wang, Y. F., Huang, J. Y. & Lin, Y. T. 2010 Clustering-based hybrid inundation model for forecasting flood inundation depths. *J. Hydrol.* **385** (1–4), 257–268.

Chau, K. W. & Wu, C. L. 2010 A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. *J. Hydroinform.* **12** (4), 458–473.

Chen, A. S., Hsu, M. H., Teng, W. H., Huang, C. J., Yeh, S. H. & Lien, W. Y. 2006 Establishing the database of inundation potential in Taiwan. *Nat. Hazards* **37** (1–2), 107–132.

Chen, L., Yeh, K. C., Wei, H. P. & Liu, G. R. 2011 An improved genetic programming to SSM/I estimation typhoon precipitation over ocean. *Hydrol. Process.* **25** (16), 2573–2583.

Chung, C. H., Chiang, Y. M. & Chang, F. J. 2012 A spatial neural fuzzy network for estimating pan evaporation at ungauged sites. *Hydrol. Earth Syst. Sci.* **16** (1), 255–266.

Cristianini, N. & Shaw-Taylor, J. 2000 *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge University Press, New York.

de Vos, N. J. & Rientjes, T. H. M. 2005 Constraints of artificial neural networks for rainfall–runoff modeling: trade-offs in hydrological state representation and model evaluation. *Hydrol. Earth Syst. Sci.* **9** (1–2), 111–126.

Giustolisi, O. & Simeone, V. 2006 Multi-objective strategy in artificial neural network construction. *Hydrol. Sci. J.* **51** (3), 502–523.

Grover, R. & Vriens, M. 2006 *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*. SAGE Publications, Thousand Oaks, CA.

Guo, W. D., Lai, J. S. & Lin, G. F. 2007 Hybrid flux-splitting finite-volume scheme for the shallow water flow simulations with source terms. *J. Mech.* **23** (4), 399–414.

Hu, T. S., Wu, F. Y. & Zhang, X. 2007 Rainfall–runoff modeling using principal component analysis and neural network. *Nord. Hydrol.* **38** (3), 235–248.

Isik, S. & Singh, V. P. 2008 Hydrologic regionalization of watersheds in Turkey. *J. Hydrol. Eng.* **13** (9), 824–834.

Johnson, R. A. & Wichern, D. W. 2002 *Applied Multivariate Statistical Analysis*, 5th edn. Prentice Hall, Upper Saddle River, NJ.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A. & Moradi, A. 2012 An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environ. Earth Sci.* **67** (1), 251–264.

Kuiry, S. N., Sen, D. & Bates, P. D. 2010 Coupled 1D-quasi-2D flood inundation model with unstructured grids. *J. Hydrol. Eng.* **136** (8), 493–506.

Lane, S. N. 1998 Hydraulic modelling in hydrology and geomorphology: a review of high resolution approaches. *Hydrol. Process.* **12** (8), 1131–1150.

Lin, G. F. & Chen, G. R. 2008 A systematic approach to the input determination for neural network rainfall–runoff models. *Hydrol. Process.* **22** (14), 2524–2530.

Lin, G. F., Chen, G. R. & Huang, P. Y. 2010 Effective typhoon characteristics and their effects on hourly

reservoir inflow forecasting. *Adv. Water Resour.* **33** (8), 887–898.

Lin, G. F., Chen, G. R., Huang, P. Y. & Chou, Y. C. 2009a Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *J. Hydrol.* **372** (1–4), 17–29.

Lin, G. F., Chen, G. R., Wu, M. C. & Chou, Y. C. 2009b Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resour. Res.* **45** (8), W08440.

Lin, G. F., Chen, L. H. & Kao, S. C. 2005 Development of regional design hyetographs. *Hydrol. Process.* **19** (4), 937–946.

Lin, J. Y., Cheng, C. T. & Chau, K.W. 2006 Using support vector machines for long-term discharge prediction. *Hydrol. Sci. J.* **51** (4), 599–612.

Lin, G. F., Chou, Y. C. & Wu, M. C. 2013a Typhoon flood forecasting using integrated two-stage support vector machine approach. *J. Hydrol.* **486**, 334–342.

Lin, G. F., Huang, P. K. & Lin, H. Y. 2013b Forecasting tropical cyclone intensity change in the western North Pacific. *J. Hydroinform.* **15** (3), 952–966.

Lin, G. F., Lin, H. Y. & Wu, M. C. 2012 Development of a support-vector-machine-based model for daily pan evaporation estimation. *Hydrol. Process.* doi: 10.1002/hyp.9428.

Liong, S. Y. & Sivapragasam, C. 2002 Flood stage forecasting with support vector machines. *J. Am. Water Resour. Assoc.* **38** (1), 173–186.

Lopez, P., Velo, R. & Maseda, F. 2008 Effect of direction on wind speed estimation in complex terrain using neural networks. *Renew. Energy* **33** (10), 2266–2272.

MacQueen, J. 1967 Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, CA, pp. 281–297.

Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environ. Modell. Softw.* **15**, 101–124.

Maity, R., Bhagwat, P. P. & Bhatnagar, A. 2010 Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrol. Process.* **24** (7), 917–923.

Meshgi, A. & Khalili, D. 2009 Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. I. A re-visit to regional homogeneity. *Stoch. Environ. Res. Risk Assess.* **23** (1), 119–135.

Neal, J., Villanueva, I., Wright, N., Willis, T., Fewtrell, T. & Bates, P. 2012 How much physical complexity is needed to model flood inundation. *Hydrol. Process.* **26** (15), 2264–2282.

Nguyen, P. K. T. & Chua, L. H. C. 2012 The data-driven approach as an operational real-time flood forecasting model. *Hydrol. Process.* **26** (19), 2878–2893.

Nourani, V., Baghanamd, A. H., Adamowskid, J. & Gebremichaele, M. 2013 Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. *J. Hydrol.* **476**, 228–243.

Nourani, V. & Kalantari, O. 2010 Integrated artificial neural network for spatiotemporal modeling of rainfall-runoff-sediment processes. *Environ. Eng. Sci.* **27** (5), 411–422.

Nourani, V., Kalantari, O. & Baghanam, A. 2012 Two semi distributed ANN-based models for estimation of suspended sediment load. *J. Hydrol. Eng.* **17** (12), 1368–1380.

Pan, T. Y., Lai, J. S., Chang, T. J., Chang, H. K., Chang, K. C. & Tan, Y. C. 2011 Hybrid neural networks in rainfall-inundation forecasting based on a synthetic potential inundation database. *Nat. Hazards Earth Syst. Sci.* **11** (3), 771–787.

Pramanik, N., Panda, R. K. & Singh, A. 2011 Daily river flow forecasting using wavelet ANN hybrid models. *J. Hydroinform.* **13** (1), 49–63.

Remesan, R., Ahmadi, A., Shamim, M. A. & Han, D. 2010 Effect of data time interval on real-time flood forecasting. *J. Hydroinform.* **12** (4), 396–407.

Saf, B. 2009 Regional flood frequency analysis using L-moments for the west Mediterranean region of Turkey. *Water Resour. Manage.* **23** (3), 531–551.

Satyanarayana, P. & Srinivas, V. V. 2008 Regional frequency analysis of precipitation using large-scale atmospheric variables. *J. Geophys. Res.* **113**, D24110.

Seyoum, S. D., Vojinovic, Z., Price, R. K. & Weesakul, S. 2012 Coupled 1D and noninertia 2D flood inundation model for simulation of urban flooding. *J. Hydrol. Eng.* **138** (1), 23–34.

Sivapragasam, C. & Liong, S. Y. 2005 Flow categorization model for improving forecasting. *Nord. Hydrol.* **36** (1), 37–48.

Vapnik, V. 1995 *The Nature of Statistical Learning Theory*. Springer, New York.

Vapnik, V. 1998 *Statistical Learning Theory*. John Wiley, New York.

Wang, W. C., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.* **374** (3–4), 294–306.

Wei, S. K., Song, J. X. & Khan, N. I. 2012 Simulating and predicting river discharge time series using a wavelet-neural network hybrid modelling approach. *Hydrol. Process.* **26** (2), 281–296.

Wu, C. L. & Chau, K. W. 2011 Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *J. Hydrol.* **399** (3–4), 394–409.

Wu, C. L., Chau, K. W. & Li, Y. S. 2008 River stage prediction based on a distributed support vector regression. *J. Hydrol.* **358** (1–2), 96–111.

Yang, C. Y., Wang, H. W. & Kuo, P. H. 2011 Application of landscape ecological decision and evaluation support system on flood mitigation strategies. *Irrig. Drain.* **60** (1), 71–76.

Yu, X. Y. & Liong, S. Y. 2007 Forecasting of hydrologic time series with ridge regression in feature space. *J. Hydrol.* **332** (3–4), 290–302.

Yu, X. Y., Liong, S. Y. & Babovic, V. 2004 EC-SVM approach for real-time hydrologic forecasting. *J. Hydroinform.* **6** (3), 209–223.

First received 25 December 2012; accepted in revised form 26 March 2013. Available online 29 April 2013