Probabilistic forecasting of coastal wave height during typhoon warning period using machine learning methods

Shien-Tsung Chen

ABSTRACT

This study applied machine learning methods to perform the probabilistic forecasting of coastal wave height during the typhoon warning period. The probabilistic forecasts comprise a deterministic forecast and the probability distribution of a forecast error. A support vector machine was used to develop a real-time forecasting model for generating deterministic wave height forecasts. The forecast errors of deterministic forecasting were then used as a database to generate probabilistic forecasts by using the modified fuzzy inference model. The innovation of the modified fuzzy inference model includes calculating the similarity of the data by performing fuzzy implication and resampling the potential data from the fuzzy database for probability distribution. The probabilistic forecasting method was applied to the east coast of Taiwan, where typhoons frequently cause large waves. Hourly wave height data from an offshore buoy and various typhoon characteristics were used as inputs of the probabilistic forecasting model. Validation results from real typhoon events verified that the proposed probabilistic forecasting model can generate the predicted confidence interval, which can properly enclose the observed wave height data, excluding some cases with extreme wave heights. Moreover, an objective measure was used to validate the proposed probabilistic forecasting method.

Key words | fuzzy inference, probabilistic forecasting, support vector machine, wave height

Shien-Tsung Chen
Department of Water Resources Engineering and Conservation,
Feng Chia University,
100, Wenhua Road, Taichung City, 40724
Taiwan
E-mail: *stchen@fcu.edu.tw*

INTRODUCTION

Typhoons or hurricanes, causing large waves with high energy, frequently threaten coastal and offshore areas. An efficient coastal hazard warning system that can rapidly calculate typhoon waves can benefit such activities as beach protection, coastal structure management, and safe navigation. Therefore, real-time wave height forecasting during typhoons is a crucial concern in coastal and offshore engineering and management. The real-time forecasting information can be useful for authorities and decision-makers to prepare for resources planning and allocation for possible disasters.

Wave hindcasting and forecasting have been primary subjects in coastal engineering for decades. First, empirical regression methods based on wind speed, wind direction, wind duration, and fetch length were proposed to estimate ocean waves (Ippen 1966; Bretschneider & Tamaye 1976; Donelan *et al.* 1985; Hurdle & Stive 1989). The traditional empirical methods are limited in that ocean waves depend on not only the present wind field but also previous wind fields, bathymetric effect, pre-existing waves subjected to other wind systems, and the entire wave generation process (Chang *et al.* 2011). Recently, because of the improvement in computer efficiency and the knowledge of physical processes regarding wave formation, numerical models based on a wind-wave energy spectrum have been proposed to forecast wave characteristics (e.g., Chen & Wang 1983; WAMDI Group 1988; SWAMP Group 2013). The numerical models are more effective and systematic in calculation than are the empirical methods, and such models can account for the physical interaction between wave parameters and their

doi: 10.2166/hydro.2019.115

driving factors. However, these numerical models require additional information, such as bathymetric data, the preparation of which requires considerable amount of work (Goda 2003), thus making the numerical models unfavorable for practical applications. Moreover, numerical models can significantly delay forecasts because of the time required for data collection and processing (Zamani *et al.* 2008), and were observed to be limited in predicting extreme situations (Chang *et al.* 2011).

A favorable alternative to forecasting the ocean waves is the machine learning method that has low computational cost and is easy to implement for researchers and engineers. Machine learning methods are widely used in hydrologic forecasting. For example, artificial neural networks (ANNs) have been successfully applied to forecast rainfalls and floods for decades (French *et al.* 1992; Dawson & Wilby 1998; Chang *et al.* 2007; Lin & Wu 2009; Chen *et al.* 2011; Lin *et al.* 2013; Jhong *et al.* 2018). Support vector machines (SVMs) have recently gained attention and demonstrated superior performance in rainfall and flood forecasting (Liong & Sivapragasam 2002; Bray & Han 2004; Han *et al.* 2007; Lin *et al.* 2009; Chen 2013; Yu *et al.* 2017) and the projection of hydrologic variables under climate change scenarios (Tripathi *et al.* 2006; Anandhi *et al.* 2008; Ghosh & Mujumdar 2008; Chen *et al.* 2010; Yang *et al.* 2011). Moreover, various fuzzy models that are based on the fuzzy set theory have been proposed to perform hydrologic forecasting (Yu *et al.* 2000; Chang & Chen 2001; Nayak *et al.* 2004; Yu & Chen 2005; Yu *et al.* 2005; Chen *et al.* 2013; Yarar 2014). These machine learning methods have been validated in various hydrologic forecasting and applications. Therefore, numerous studies have applied machine learning methods to wave forecasting.

The studies that adopted machine learning methods to forecast ocean waves have generally used ANN, SVM, and fuzzy methods. For example, Deo & Sridhar Naidu (1998) used the current observed wave height data in ANN models to forecast real-time wave heights. Deo *et al.* (2001) developed an ANN model to obtain the wave height and period by using the generated wind speeds as inputs. Tsai *et al.* (2002) applied ANNs to predict the significant wave height by using wave height data from two sites as inputs. Mandal & Prabaharan (2006) used various combinations of the wave height data in recurrent neural networks to forecast the significant wave height. Moreover, Kazeminezhad *et al.* (2005) proposed a fuzzy logic model to simulate wave parameters by using the wind speed and fetch length in Lake Ontario. Özger & Şen (2007) investigated the relation between the wind speed and wave characteristics and applied a fuzzy logic approach to forecast the significant wave height for the California coast. Mahjoobi *et al.* (2008) used the wind field data in a fuzzy logic model and an ANN model to hindcast the wave height and period for Lake Ontario. Mahjoobi & Etemad-Shahidi (2008) developed a wave height prediction model based on classification and regression trees by using the wind speed and direction as input variables. Mahjoobi & Mosabbeb (2009) employed the SVM to simulate the significant wave height for Lake Michigan by using current and previous wind speed data as inputs. Chang *et al.* (2011) employed an ANN model to modify the unsatisfactory wave calculations of a numerical model in special cases by using the simulated wind speed and wave height from the numerical model as inputs. Malekmohamadi *et al.* (2011) evaluated the efficacy of various machine learning models on wave height forecasting in Lake Superior and concluded that most models provide the suitable prediction of wave heights. Elbisy (2013) input the wind speed and direction into an SVM model calibrated using a genetic algorithm to predict the significant wave height, wave direction, and peak spectral period. Wei (2018) used ANN and SVM to predict typhoon-induced wave height near coastal areas in northeast Taiwan.

The studies that applied machine learning methods to wave forecasting have generally used the wave height and wind field data as model inputs. A few studies have adopted typhoon characteristics as the inputs to simulate or forecast wave parameters. Chang & Chien (2006a; 2006b) applied a fuzzy-neural hybrid model and a neural network model with a multi-trend simulating transfer function to simulate the wave height by adopting the typhoon moving speed and location information. Tsai *et al.* (2017) applied the ANN to forecast wave heights along the route of a ship during hurricanes by using hurricane characteristics (location, pressure, and intensity) generated by a quadtree-adaptive numerical model (Tsai *et al.* 2013).

The aforementioned studies that have applied machine learning methods to wave forecasting are related to

deterministic forecasting. However, a deterministic forecast, which may provide the user with an illusion of certainty, can lead to a suboptimal action (Krzysztofowicz 2001). By contrast, probabilistic forecasting is scientifically more practical than deterministic forecasting and can allow the authorities to establish risk-based criteria for warnings and responses. This study presents the probabilistic forecasting of coastal wave heights during the typhoon warning period through machine learning.

This study adopted the probabilistic forecasting method that combines deterministic forecasting and the probability distribution of the forecast error for probabilistic forecasting (Chen & Yu 2007a). This method provides probabilistic forecasts with their corresponding exceeding probabilities, and the probabilistic forecasting results can be demonstrated by the confidence interval with a certain significant level. The SVM was used to develop a wave height forecasting model to produce the deterministic forecast of the significant wave height. The forecast errors of deterministic forecasting were used as the database to construct the forecasting model for predicting the probability distribution of the forecast error, which was derived using the modified fuzzy inference method proposed in this study. The modified fuzzy inference method is easier to understand and implement than the original method proposed by Chen & Yu (2007a). The proposed probabilistic forecasting method was used to forecast the significant wave heights offshore the east coast of Taiwan when a real-time typhoon warning was issued. Validation results pertaining to real typhoon events indicated that the proposed method can appropriately forecast the probability distribution of significant wave heights. The predictive confidence interval of the wave height and a wave height forecast with an exceedance probability can be derived accordingly.

METHODOLOGY

Probabilistic forecasting

This study adopted the probabilistic forecasting method proposed by Chen & Yu (2007a), where the probabilistic forecast is the combination of the deterministic forecast and the probability distribution of the forecast error. The methodology of probabilistic forecasting is briefly described as follows.

The primary concept of the probabilistic method is that many probabilistic forecasts can be obtained by adding numerous probable forecast errors to one deterministic forecast. Let the forecast error e be expressed as:

$$e = H - \hat{H} \tag{1}$$

where H and $\hat{H}$ are observation and the deterministic forecast, respectively. The variance of the deterministic forecast $\hat{H}$ is identical to the variance of the forecast error e because H is an observed value without uncertainty:

$$\mathrm{var}[\hat{H}] = \mathrm{var}[e] \tag{2}$$

Therefore, adopting the uncertainty of the forecast error can account for the total uncertainty of the forecast.

The probability distribution of the forecast error Π_e depends on the forecasting model $f(\cdot)$ and input vector $\mathbf{x}$:

$$\Pi_e = P(e|f(\mathbf{x})) \tag{3}$$

where $P(\cdot)$ denotes the probability distribution function. When the probability distribution of forecast error Π_e is estimated based on the model and its inputs $f(\mathbf{x})$, the predictive probability distribution of forecast Π_H can be obtained by adding error distribution to the deterministic single forecast, which is expressed as:

$$\Pi_H = \hat{H} + \Pi_e \tag{4}$$

In addition, the confidence interval estimated with a certain confidence level can be constructed from the derived probability distribution.

Support vector machines

The methodology of applying an SVM to a regression problem is briefly introduced in this section. Some studies have provided a detailed description of the SVM theory (Vapnik 1995, 1998; Yu *et al.* 2006; Chen 2015).

Considering a training data set with n examples, let $\mathbf{x}_i$ and y_i be the input vector and the corresponding output

($i = 1, 2, \ldots, n$), respectively. A transfer function $\Phi(\mathbf{x})$ is used to develop a nonlinear SVM regression function as follows:

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \Phi(\mathbf{x}) + b \tag{5}$$

where $\mathbf{w}$ and b are the parameters of the regression function. The SVM model is developed by introducing the Vapnik's ε-insensitive loss function that defines the tolerated error within the coverage of the ε-tube and the penalized losses L when the data are outside the tube:

$$L(y_i) = \begin{cases} 0 & \text{where } |y_i - [\mathbf{w}^T \cdot \Phi(\mathbf{x}_i) + b]| < \varepsilon \\ |y_i - [\mathbf{w}^T \cdot \Phi(\mathbf{x}_i) + b]| - \varepsilon, & \text{where } |y_i - [\mathbf{w}^T \cdot \Phi(\mathbf{x}_i) + b]| \geq \varepsilon \end{cases} \tag{6}$$

The SVM model can be expressed as the following optimization problem based on the theory of structural risk minimization:

$$\begin{aligned} \min_{\mathbf{w},b,\xi,\xi^*} \quad & \frac{1}{2}\mathbf{w}^T \cdot \mathbf{w} + C\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) \\ \text{subject to} \quad & y_i - [\mathbf{w}^T \cdot \Phi(\mathbf{x}_i) + b] \leq \varepsilon + \xi_i^+ \\ & [\mathbf{w}^T \cdot \Phi(\mathbf{x}_i) + b] - y_i \leq \varepsilon + \xi_i^- \\ & \xi_i^+, \xi_i^- \geq 0, i = 1, 2, ..., n \end{aligned} \tag{7}$$

where ξ_i^+ and ξ_i^- are slack parameters that denote the upper and lower training errors depending on the error tolerance ε, respectively. C is the penalty parameter with a positive value. By introducing a set of two Lagrange multipliers, namely, α_i^+ and α_i^-, the optimization problem can be solved by adopting a quadratic programming algorithm. After the determination of Lagrange multipliers, parameters $\mathbf{w}$ and b in the SVM function can be estimated under complementarity conditions. Therefore, the SVM regression function can be formulated as:

$$f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)\Phi(\mathbf{x}_i)^T \cdot \Phi(\mathbf{x}) + b \tag{8}$$

The inner products in this equation can be calculated in a simple form by using a kernel function (Vapnik 1999):

$$K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i)^T \cdot \Phi(\mathbf{x}) \tag{9}$$

The prevalently used kernels in the SVM include linear, polynomial, sigmoid, and radial basis functions. Dibike *et al.* (2001) and Han & Cluckie (2004) explained the advantages of the radial basis function over other kernels. Moreover, several studies have exhibited the superior performance of the radial basis function in hydrologic and wave forecasting (Liong & Sivapragasam 2002; Choy & Chan 2003; Elbisy 2013). Therefore, the radial basis function kernel with a parameter γ was used in this study:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma|\mathbf{x}_i - \mathbf{x}|^2) \tag{10}$$

According to Equation (8), only the data corresponding to nonzero Lagrange multipliers ($\alpha_i^+ - \alpha_i^-$) are used to generate the SVM regression function. The data with nonzero Lagrange multipliers are called support vectors that support the construction of the regression function. Finally, the SVM model can be rewritten as:

$$f(\mathbf{x}) = \sum_{k=1}^{m}(\alpha_i^+ - \alpha_i^-) \cdot \exp(-\gamma|\mathbf{x}_k - \mathbf{x}|^2) + b \tag{11}$$

where $\mathbf{x}_k$ and m denote the support vector and the number of support vectors, respectively.

Parameters in the SVM model are the penalty parameter C, error tolerance ε, and kernel parameter γ. In this study, the parameters are optimized using a two-step grid search method with a cross-validation technique (Chen & Yu 2007b).

Modified fuzzy inference method

Fuzzy inference is to develop a fuzzy logic system that emulates the reasoning process of a human expert within a particular domain of knowledge (Klir & Yuan 1995). This study proposed a modified fuzzy inference method that combines the fuzzy rule database with a resampling technique to output probability distribution. The proposed fuzzy inference method involves the following four steps: (1) fuzzifying inputs, (2) formulating fuzzy rules, (3) fuzzy implication as a measure of similarity, and (4) resampling to generate probability distribution. The methodology is described as follows.

Fuzzifying inputs

Input variables, which are crisp numerical values, are first converted into fuzzy sets by applying fuzzy membership

functions. The degree of similarity between an input variable and the defined fuzzy set can then be calculated using the membership function. Various fuzzy membership functions have been presented. When selecting the fuzzy membership function, subjective expert decisions are generally required.

Formulating fuzzy rules

A fuzzy rule is in the form of an 'IF–THEN' statement, where the 'IF' component is the premise of the rule, and the 'THEN' component is the consequence. A fuzzy rule should be logical and should adopt the most appropriate variables. The premise and consequence components can involve more than one variable. An example of fuzzy rule can be expressed as:

$$\begin{aligned}&\text{IF } (x_1 \text{ is } A_1) \text{ AND } (x_2 \text{ is } A_2) \ \ldots \ \text{AND } (x_m \text{ is } A_m)\\&\text{THEN } (y_1 \text{ is } B_1) \text{ AND } (y_2 \text{ is } B_2) \ \ldots \ \text{AND } (y_n \text{ is } B_n)\end{aligned} \tag{12}$$

where $x_1, x_2, \ldots, x_m$ are input variables; $y_1, y_2, \ldots, y_n$ are output variables; and $A_1, A_2, \ldots, A_m$ and $B_1, B_2, \ldots, B_n$ are fuzzy sets defined by membership functions.

If the premise is true to a certain degree, then the consequence is also true to the same degree. The formulated fuzzy rule can be either a linguistic statement or the relation between numerical input and output data. When the numerical data are used to formulate the fuzzy rules, these rules represent the fuzzified relation between the input and output data. These formulated rules are stored in a fuzzy rule database. The following resampling procedure is then performed to sample the preferred data out of the database.

Fuzzy implication as a measure of similarity

Fuzzy implication is used as a measure of similarity between the fuzzy rules (the exemplar data stored in the database) and input data. For each fuzzy rule, an implication process generates a fuzzy subset of the output variable in the consequence. Figure 1 depicts an example of the fuzzy implication process that has two input variables in the premise and one output variable in the consequence. Membership grades μ_1 and μ_2 of both input data can be determined using the membership functions of input variables in the fuzzy rule. A fuzzy intersection operation is used to select the minimum membership grade as μ_2 for the output variable. The resulting membership grade μ_2 for the rule is termed firing strength. Firing strength is used as a measure of similarity between the input variable and fuzzy rule in the database. In this example, the inferred output has a similarity degree of μ_2 with the output variable in Figure 1. The repetitive process of fuzzy implication for each rule provides the similarity of an input data set to each fuzzy rule.

Resampling to construct probability distribution

After performing the procedure described in the previous section, each rule has a similarity with respect to a certain input data set. Given an input data set, the output is more likely to be the output variable in a rule with high similarity and less possible to be the output variable in a rule with low similarity. Thus, a resampling procedure can be applied on the basis of the similarity degree: the output variable in a rule with higher similarity has higher probability to be

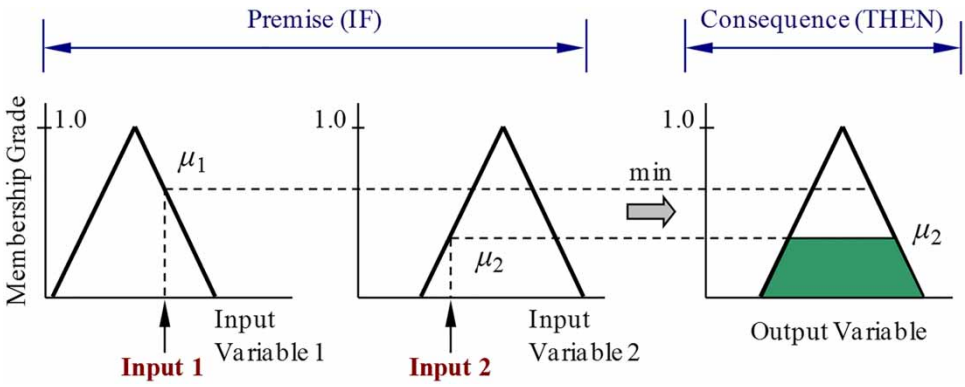

Figure 1 | Fuzzy implication process.

sampled out than that with lower similarity. Probability p_i that the output variable in a fuzzy rule is selected is expressed as:

$$p_i = \mu_i / \sum_{i=1}^{l} \mu_i \tag{13}$$

where μ_i is the similarity degree corresponding to the i-th rule, and l is the number of rules. By using a repetitive resampling procedure, the large number of sampled outcomes can form a probability distribution with desired likelihood.

APPLICATION

Typhoon and wave height data

In Taiwan, approximately four typhoons occur annually, several of which move from the Pacific Ocean directly toward the east coast of Taiwan. Hualien, located on the east coast of Taiwan, was used as the study area because the long-term wave data are available for this offshore region (Figure 2). The hourly significant wave height data are recorded using a wave rider buoy operated by the Central Weather Bureau (CWB) of Taiwan. The buoy is placed at 121.63°E and 24.03°N, approximately 1 km from the coast at a water depth of approximately 30 m. The hourly typhoon data of central pressure (hPa), central wind velocity (m/s), cyclonic radius (km), and coordinates of latitude and longitude are available from the CWB typhoon archives.

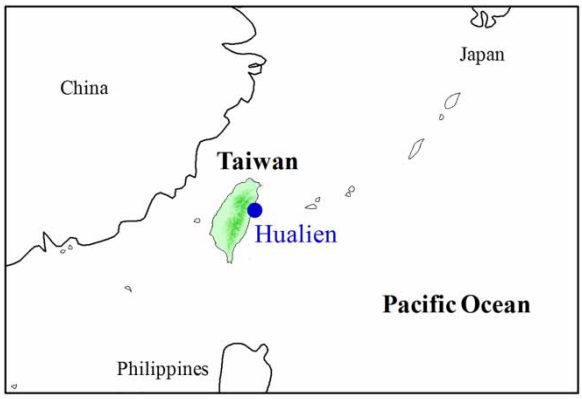

Figure 2 | Location of Hualien Wave Buoy.

Complete hourly wave height and typhoon data during the typhoon warning period issued by the CWB are available for 27 typhoon events from 2006 to 2017. Table 1 lists the typhoon and the wave data, including the name and date of typhoon, the lowest central pressure, highest wind speed, cyclonic radius, moving track (classified by the CWB, cf. Figure 3), maximum significant wave height, and number of data points during the typhoon warning period. Table 1 indicates that the maximum wave height during a typhoon corresponds with the moving tracks. Typhoon events with maximum wave heights higher than 600 cm belong to two, three, and four tracks, and the typhoons in these tracks directly strike the east coast of Taiwan. For the collected 27 typhoon events, the former 20 events were used to calibrate the proposed probabilistic forecasting model, and the remaining seven events were used for validation.

Deterministic wave height forecasting

SVM was used to perform real-time deterministic wave height forecasting. Variables, such as the central pressure (P), central wind velocity (V), and cyclonic radius (R), that can characterize the typhoon scale are used as input variables of the SVM forecasting model. In addition, the significant wave height (H) corresponds to the track of typhoon. The relative position of a typhoon to the study area was mapped using a polar coordinate system, with the location of Hualien Buoy set as the origin. The position of the typhoon is defined using the great-circle distance (D) and angle (θ). The angle directly facing east is set as 0°, and the counterclockwise direction is set as positive (Figure 4); these settings were applied because typhoons generally approach Taiwan from the east, and a typhoon north or south of the Hualien Buoy has a different effect on wave development. Moreover, this definition ensures the continuity of the angle data when a typhoon is to the east of the study area and prevents discontinuity when the angle is ±180° in the west. The current observed significant wave height was also used as an input variable to forecast wave heights. Six input variables were used to develop the forecasting model to predict wave heights during the typhoon warning period. This study applied real-time forecasting for various lead times. The architecture of forecasting

Table 1 | Typhoon events and wave data used in this study

Event no.	Name of typhoon	Date (yyyy/mm/dd)	Lowest pressure (hPa)	Highest wind (m/s)	Cyclonic radius (km)	Moving track	Max. wave height (cm)	No. of data
1	Bilis	2006/07/12	978	25	300	2	441	73
2	Kaemi	2006/07/23	960	28	200	3	486	54
3	Bopha	2006/08/07	985	23	120	4	324	39
4	Pabuk	2007/08/06	980	28	150	4	355	36
5	Wutip	2007/08/08	992	18	100	3	414	24
6	Sepat	2007/08/16	920	53	250	3	937	79
7	Wipha	2007/09/17	935	48	200	1	321	52
8	Krosa	2007/10/04	925	51	300	2	1,013	79
9	Kalmaegi	2008/07/16	970	33	120	2	298	58
10	Fungwong	2008/07/26	948	43	220	3	958	77
11	Sinlaku	2008/09/11	925	51	250	2	635	109
12	Hagupit	2008/09/21	940	45	280	*	343	43
13	Jangmi	2008/09/26	925	53	280	2	953	73
14	Parma	2009/10/03	945	43	250	10	331	85
15	Fanapi	2010/09/17	940	45	200	4	1,189	64
16	Megi	2010/10/21	935	48	250	9	443	70
17	Tembin	2012/08/21	945	45	180	10	350	158
18	Jelawat	2012/09/27	910	55	250	*	535	43
19	Fitow	2013/10/04	960	38	250	1	549	58
20	Matmo	2014/07/21	960	38	200	3	748	55
21	Soudelor	2015/08/06	930	48	300	3	1,195	70
22	Dujuan	2015/09/27	925	51	220	2	663	58
23	Nepartak	2016/07/06	905	58	200	4	396	72
24	Meranti	2016/09/12	900	60	220	7	508	61
25	Malakas	2016/09/15	940	45	180	*	393	58
26	Megi	2016/09/25	940	45	250	3	1,056	67
27	Nesat	2017/07/28	955	40	180	2	378	51

Note: Asterisk *indicates that typhoon tracks were not classified by CWB.

models can be expressed as follows:

$$\hat{H}(t+1) = f_{\text{SVM\_1}}[H(t), P(t), V(t), R(t), D(t), \theta(t)] \quad (14)$$

$$\hat{H}(t+2) = f_{\text{SVM\_2}}[H(t), P(t), V(t), R(t), D(t), \theta(t)] \quad (15)$$

$$\hat{H}(t+3) = f_{\text{SVM\_3}}[H(t), P(t), V(t), R(t), D(t), \theta(t)] \quad (16)$$

where $\hat{H}(t+1)$, $\hat{H}(t+2)$, and $\hat{H}(t+3)$ are forecasted wave heights with lead times of 1, 2, and 3 h, respectively; $f_{\text{SVM\_1}}$, $f_{\text{SVM\_2}}$, and $f_{\text{SVM\_3}}$ denote the SVM forecasting models for lead times of 1, 2, and 3 h, respectively; and t denotes time.

The variables differ in their values and units, which are not in the same order of magnitude. Therefore, these variables must be normalized to prevent the model from being dominated by variables with large values. Wave height, central pressure, central wind velocity, cyclonic radius, and distance are normalized to the interval of [0, 1] by using the following equation:

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (17)$$

where X is the original value; X_n is the normalized value; and $X_{\min}$ and $X_{\max}$ are the parameters for scaling. Table 2

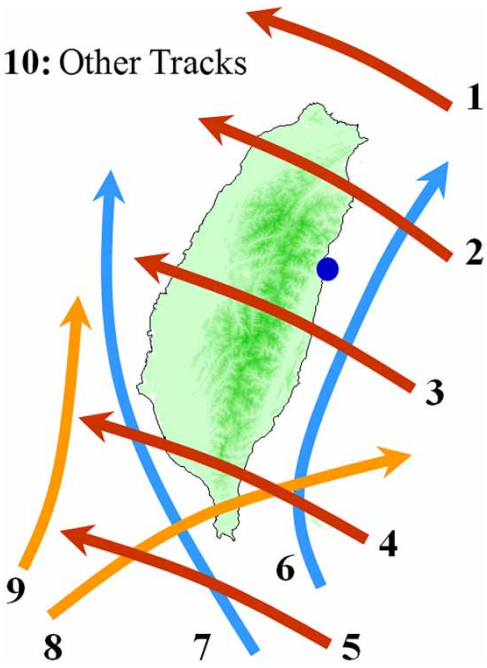

Figure 3 | Moving tracks of typhoons classified by CWB.

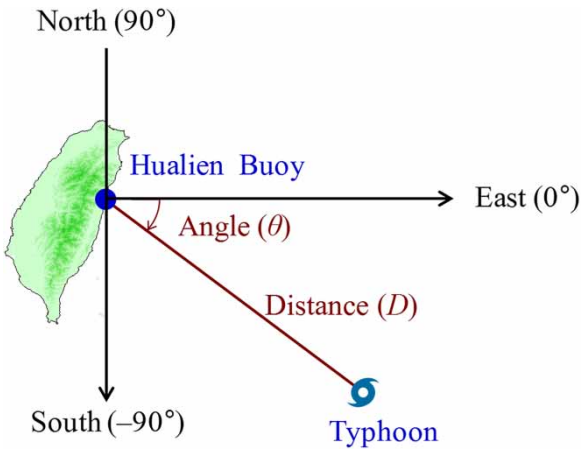

Figure 4 | Definition of distance and angle of typhoon with respect to Hualien Buoy.

Table 2 | Scaling parameters for variables

Parameter	Wave height (cm)	Central pressure (hPa)	Wind velocity (m/s)	Cyclonic radius (km)	Distance (km)
$X_{\min}$	0	900	0	0	0
$X_{\max}$	1,200	1,000	60	300	1,000

lists the parameters for scaling according to the properties of the calibration data. The angle data are linearly normalized into [−1, 1] corresponding to the original interval of [−180°, 180°]. Thus, the original outputs of the wave height from SVM forecasting models were normalized data. These normalized outputs were subsequently transformed into wave height forecasts with their real scale (cm).

The hourly data from calibration events were used to calibrate the three SVM forecasting models with multiple hours ahead. Root mean squared error (RMSE) was used as an objective function to determine the optimal parameters of the forecasting models. RMSE is an overall estimate of the error between model forecasts and observed data. An RMSE index highlights large errors, such as errors around the peak wave height. Table 3 lists the calibrated parameters of the penalty parameter C, error tolerance ε, and kernel parameter γ of the SVM models and presents the RMSEs in the calibration phase. The RMSE of the significant wave height for the 1-h forecasting model during the calibration phase was 49.7 cm. The RMSEs for the 2-h and 3-h forecasting models were 64.8 and 75.6 cm, respectively. The RMSE values of wave height during typhoon period were tolerable. Figure 5 shows the 1-h simulation results pertaining to calibration events, which indicates that the 1-h forecasting model can satisfactorily replicate the wave height process excluding some cases around the peak wave height. Figure 6 presents scatter plots for multiple-hour-ahead simulation results. In general, the model outputs are similar to the observed data; however, the simulation for larger wave heights exhibits larger deviations.

The calibrated SVM models were executed in accordance with validation events to examine real-time forecasting performance. Figure 7 shows the deterministic forecasting performance in terms of RMSE for the calibration and validation events. The RMSEs for validation are greater than those for calibration, with a difference of

Table 3 | Calibrated parameters and RMSEs for SVM models

Models	Penalty parameter C	Error tolerance ϵ	Kernel parameter γ	RMSE (cm)
$f_{\text{SVM\_1}}$	1.8	0.008	0.10	49.7
$f_{\text{SVM\_2}}$	1.8	0.020	0.11	64.8
$f_{\text{SVM\_3}}$	1.8	0.019	0.10	75.6

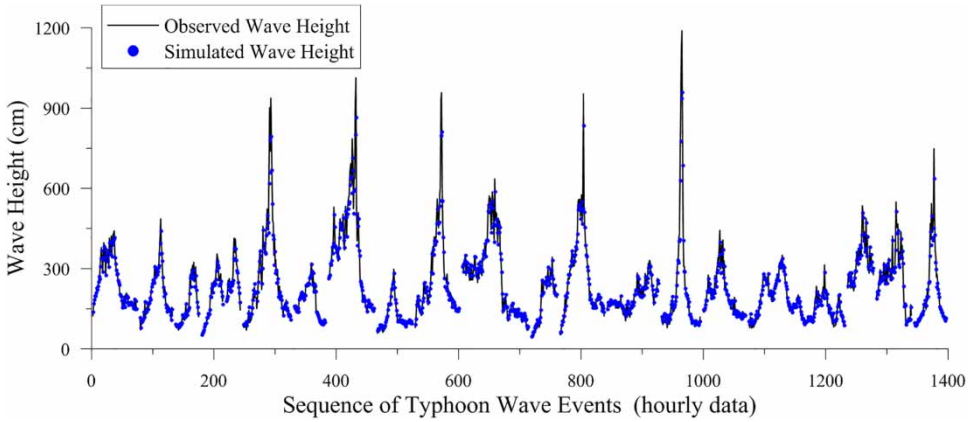

Figure 5 | Simulation results pertaining to calibration events.

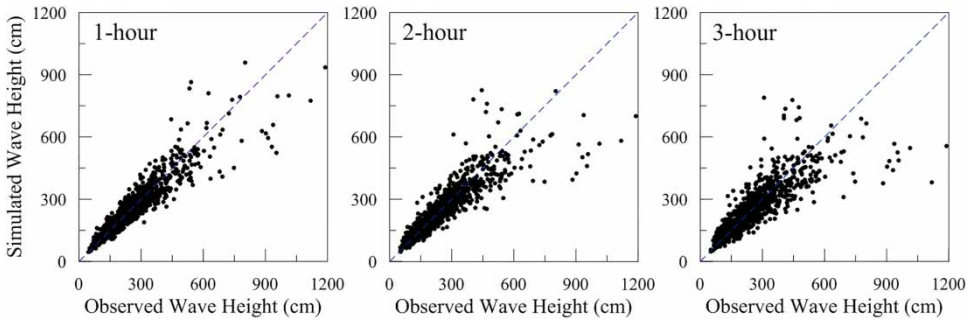

Figure 6 | Scatter plot for multiple-hour-ahead simulation results.

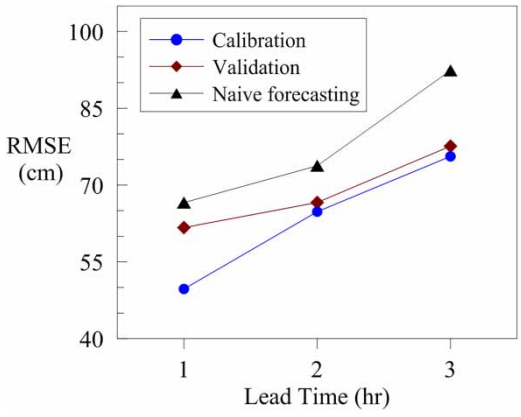

Figure 7 | Deterministic and naïve forecasting performance in terms of RMSE.

12 cm for 1-h forecasting and approximately 2 cm for 2-h and 3-h forecasting. In general, forecasting performance in validation is only slightly inferior to that in model simulation. Figures 8–10 exhibit deterministic wave height forecasting with lead times of 1, 2, and 3 h, respectively. The wave heights forecasted 1 to 3 h ahead generally coincide with the observed wave heights. However, the deterministic forecasting models do not successfully forecast peak wave heights for large events, such as Events 21 and 26. Moreover, a minor phase lag can be observed for forecasts with longer lead times, such as Events 24–27 (Figure 10). The naïve forecasting (or persistence forecasting) is a simple model that provides the same forecasts as the observations with a pure phase lag. Figure 7 also illustrates the forecasting performance with respect to the naïve forecasting, revealing that the SVM outperforms the naïve model. The issue of addressing the phase error has been studied and additional insights can be found in, e.g., Babovic *et al.* (2000), Sannasiraj *et al.* (2004), and Sun *et al.* (2010). Figure 11 presents the scatter plots for real-time wave height forecasting results. Figures 6 and 11 show that for larger wave heights, the simulation or

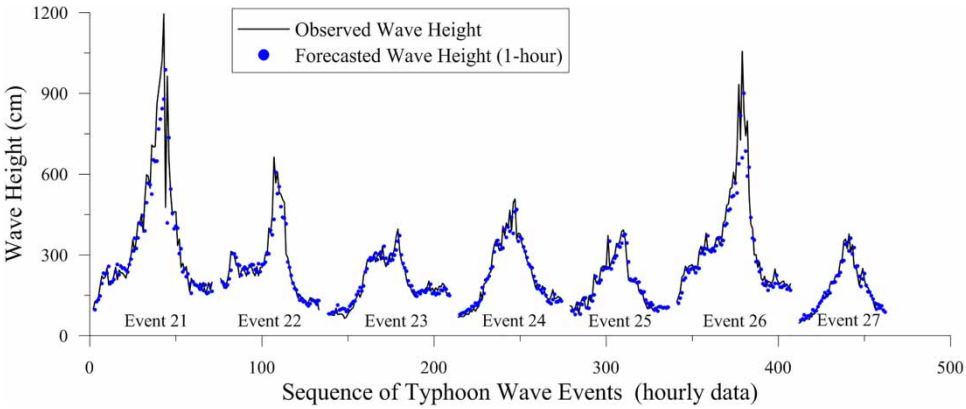

Figure 8 | Deterministic wave height forecasting with 1-h lead time.

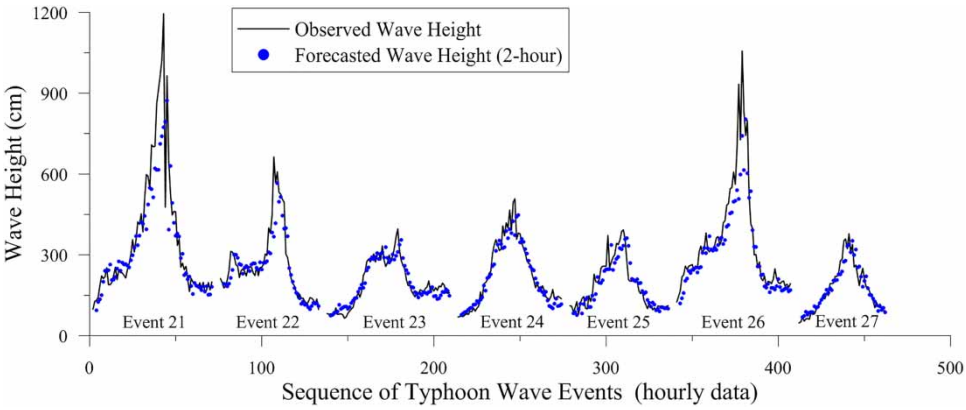

Figure 9 | Deterministic wave height forecasting with 2-h lead time.

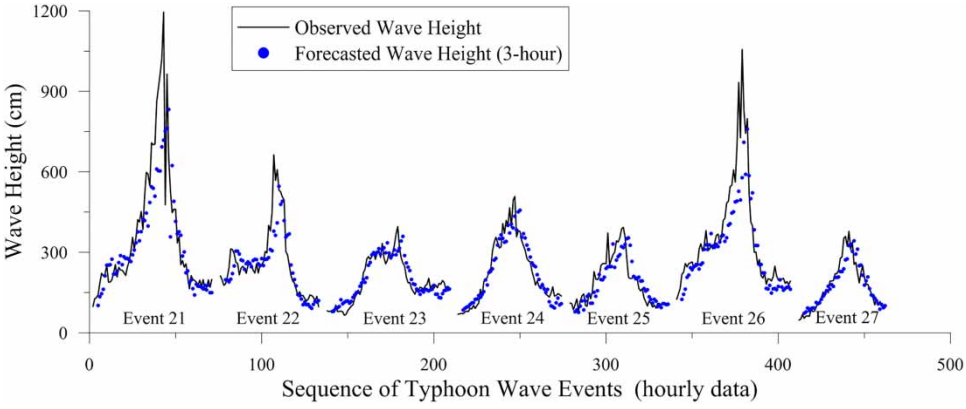

Figure 10 | Deterministic wave height forecasting with 3-h lead time.

forecasting results exhibit larger errors. Usually, there are less extreme data than the normal data in nature and in the collected database. Therefore, the extreme data have higher uncertainty and the deterministic forecasting for extreme waves has larger errors. Although SVM models have satisfactory forecasting performance in predicting the

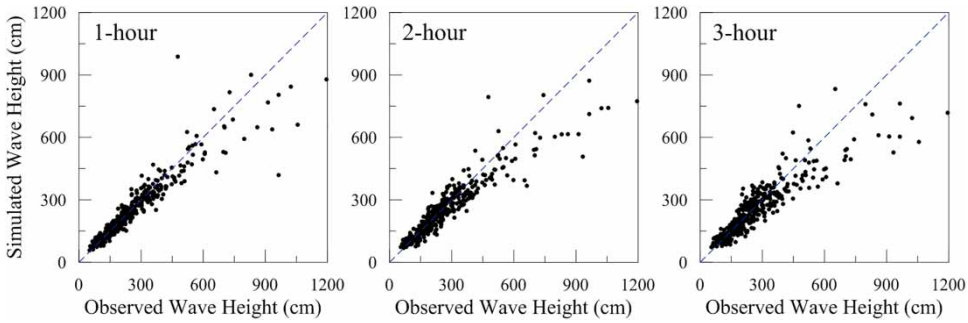

Figure 11 | Scatter plot for deterministic wave height forecasting results.

typhoon wave height, the deterministic forecasting scheme has minor phase lag and large errors around the peak wave height. Alternatively, the probabilistic forecasting can consider the forecasting uncertainty in the way that many possible forecasts with their exceedance probabilities are provided. Thus, the probabilistic forecasting method was used to provide comprehensive forecasting information.

Probabilistic wave height forecasting

Probabilistic forecasts are the combination of a deterministic forecast and probability distribution of the forecast error. The errors of deterministic forecasting during the calibration period provide the database for establishing probability distribution. The modified fuzzy inference method was proposed to deduce error probability distributions, which is described in the following.

First, the data of input variables during the calibration period must be transformed into fuzzy sets by applying the fuzzy membership function. In this study, the Gaussian membership function was used. The membership function defines the value of an input variable m in the database as a fuzzy set:

$$\mu(x) = \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right) \tag{18}$$

where σ is a parameter representing the spread of the membership function, and $(x-m)$ corresponds to the distance between an input value x and the center of the membership function m. Therefore, the membership grade $\mu(x)$ measures the similarity of x with m. When x is close to m, the membership grade (similarity) is large, whereas when x is away from m, the similarity is exponentially reduced. Gaussian function was employed because of its improved differentiability than the prevalently used triangular membership function. Parameter σ is adopted as the standard deviation of the calibration data of the input variable (Chen & Yu 2007a). The parameters of membership functions for variables of wave height (H), pressure (P), velocity (V), radius (R), distance (D), and angle (θ) in the normalized scale were determined to be 0.1179, 0.2172, 0.1615, 0.1824, 0.1823, and 0.4978, respectively.

Equation (3) indicates that the probability distribution of the forecast error depends on the model and inputs. Therefore, the fuzzy rule is formulated using input variables of the deterministic forecasting model in the premise and the forecast error in the consequence. Thus, the fuzzy rule for 1-h forecasting is formulated as:

$$\begin{aligned}&\text{IF } (H(t) \text{ is } F_H) \text{ AND } (P(t) \text{ is } F_P) \text{ AND } (V(t) \text{ is } F_V)\\ &\text{AND } (R(t) \text{ is } F_R) \text{ AND } (D(t) \text{ is } F_D) \text{ AND } (\theta(t) \text{ is } F_\theta)\\ &\text{THEN } (\hat{E}(t+1) \text{ is } F_E)\end{aligned} \tag{19}$$

where F_H, F_P, F_V, F_R, F_D, and F_θ indicate the defined fuzzy sets with respect to variables of wave height (H), pressure (P), velocity (V), radius (R), distance (D), and angle (θ) corresponding to the calibration data at time t, respectively. F_E is the fuzzy sets defined by 1-h simulation errors in the database, and $\hat{E}(t+1)$ denotes the inferred error for 1-h forecasting. Similarly, the rules for 2-h and 3-h forecasting are analogous to Equation (19); however, the variables in the consequence are $\hat{E}(t+2)$ and $\hat{E}(t+3)$.

When the data of six input variables at present time t were available, fuzzy implication was executed to obtain the similarity of this input data set to each rule in the database. The resampling procedure was then applied to sample out numerous forecast errors form the fuzzy rule database to obtain predictive probability distribution. In this study, 10,000 sets of the forecast error were generated at each time step, and predictive probability distribution was constructed using the Weibull plotting position formula. The derived probability distribution of the forecast error was added to the single deterministic forecast with respect to each lead-time forecasting to form probabilistic forecasting.

Figures 12–14 present probabilistic forecasting results with a 95% confidence interval (shaded area) for the forecasting lead time of 1–3 h, respectively. The regions with 95% confidence, which widen with the increasing forecast lead time, can reasonably enclose most of the observed wave heights. The predicted confidence intervals are

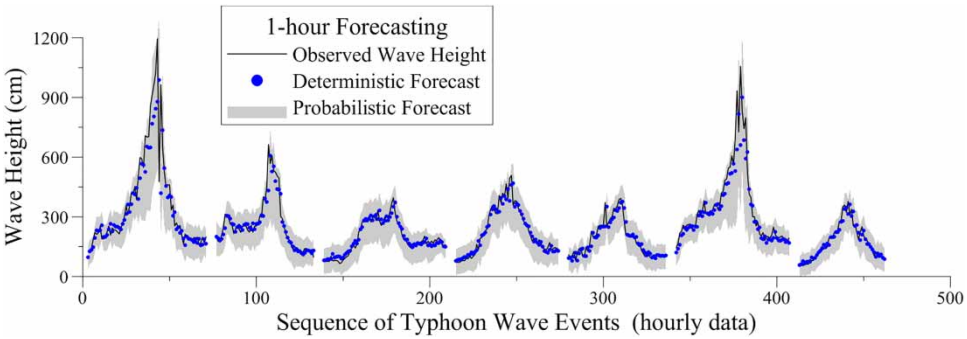

Figure 12 | Probabilistic wave height forecasting with 95% confidence interval (1-h ahead).

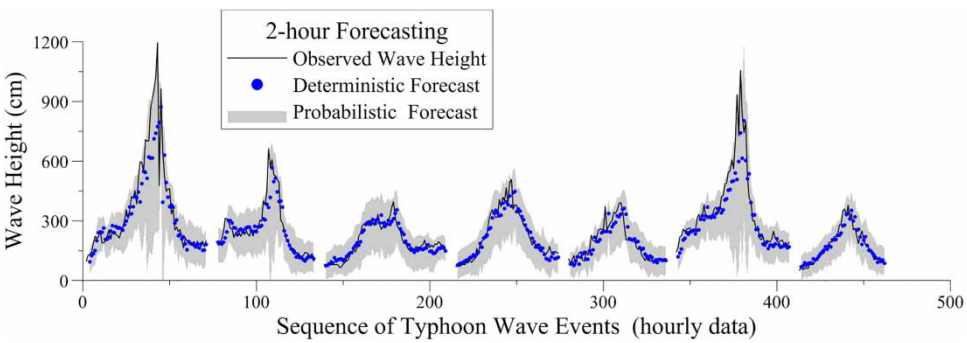

Figure 13 | Probabilistic wave height forecasting with 95% confidence interval (2-h ahead).

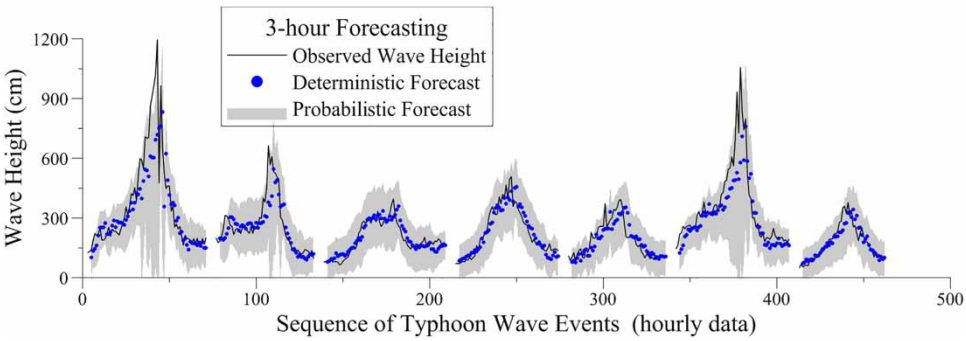

Figure 14 | Probabilistic wave height forecasting with 95% confidence interval (3-h ahead).

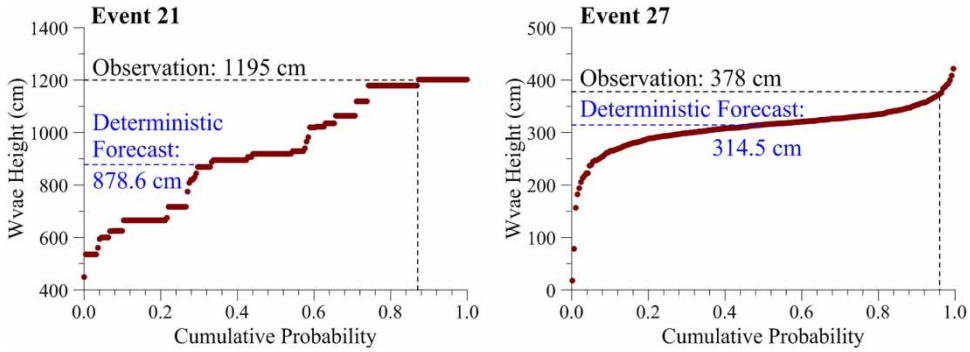

Figure 15 | Examples of predictive probability distribution for peak wave heights.

smooth, and their amplitude is small for 1-h forecasting, demonstrating its practical usefulness. However, for 3-h forecasting, the predicted confidence intervals around the peak wave heights display fluctuations, and the interval is large in some events. Furthermore, the 3-h predictive 95% confidence interval does not include some extreme wave heights for Event 21 because the fuzzy rule database does not contain enough data for such extremely high wave heights for the inference model to produce perfect probabilistic forecasts. Nevertheless, the proposed modified fuzzy inference method can almost successfully perform probabilistic forecasting.

Figure 15 exhibits 1-h predicted probability details of peak wave heights corresponding to the largest (Event 21) and smallest events (Event 27) in validation. For Event 21, the observed peak wave height was 1,195 cm; however, the deterministic forecast was estimated to be 878.6 cm. The predicted probability distribution could cover the peak wave height with an exceedance probability of 0.13 (a cumulative probability of 0.87), which notified the user with the probability of the occurrence of an extremely high wave. However, probability distribution is not smooth and continuous because of the insufficient number of similar large wave data in the database. For Event 27, probability distribution is smooth and continuous because the database comprised similar data for the peak wave height of Event 27. The deterministic forecast for the peak wave height was 314.5 cm, which was lower than the observed value of 378 cm. Although the deterministic forecast was underestimated, the addition of the predicted probability distribution of the forecast error could enclose the observed peak wave height with an accompanying exceedance probability of 0.04. The examples in Figure 15 are extreme cases. Although the predicted probability distribution does not appropriately include the observed peak wave height in its middle region, it informs the user regarding the occurrence of a certain wave height in the future with a probability value. For probabilistic forecasting of most of the data, the predicted 95% confidence intervals are preferable for 1-h to 3-h forecasting (Figures 12–14).

An essential approach to assessing the performance of probabilistic forecasting includes the accurate enclosure of the amount of data within the confidence region. If predictive probability distribution can effectively explain total uncertainty, then the percentage of data included in the confidence interval is equivalent to the confidence level (Chen

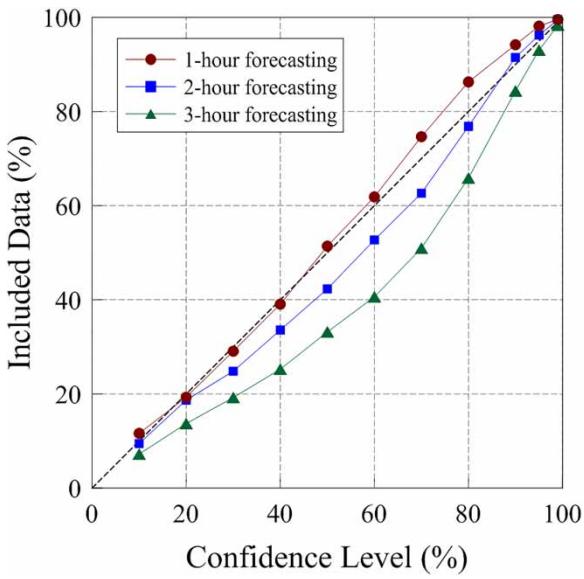

Figure 16 | Percentage of observed data included in confidence region.

& Yu 2007a). Figure 16 shows the percentages of observed data that are included in the confidence interval at various percentages of confidence levels. In Figure 16, the included data for 1-h probabilistic forecasting is ideal. The performance of 2-h and 3-h probabilistic forecasting is inferior to that of 1-h forecasting because less than expected data are accurately enclosed in the confidence region. Moreover, this observation is analogous to the probabilistic forecasting results (Figures 12–14), indicating that probabilistic forecasts with longer lead times are inferior to those with shorter lead times. Overall, the proposed probabilistic forecasting method can account for the total uncertainty during the forecasting process. The modified fuzzy inference method can resample the desired forecast errors from the database to form predictive probability distribution for all cases, excluding some extreme cases around the peak wave height.

CONCLUSION

Fast and robust wave forecasting models for the coastal hazard warning system are crucial for practical engineering applications. The use of machine learning methods to forecast wave parameters in a timely manner has attracted considerable attention in coastal engineering. This study applied machine learning methods to forecast real-time significant wave heights in a probabilistic framework, which is a novel investigation in probabilistic wave forecasting.

The probabilistic forecasting method combines a deterministic forecast and the probability distribution of a forecast error. An SVM was used to develop a deterministic forecasting model to forecast significant wave heights with lead times of 1–3 h. A modified fuzzy inference model was proposed in this study to forecast the probability distribution of the forecast error, with the innovation of calculating the similarity of data by implementing fuzzy implication and resampling potential data from a fuzzy database to form probability distribution. The east coast of Taiwan, where typhoons frequently occur, was used as the study area. The wave data observed using the Hualien Buoy and typhoon characteristics were used in the model. Validation results from seven typhoon events demonstrated that the proposed probabilistic forecasting model is practical and robust. The predicted confidence interval can satisfactorily enclose the real wave height data with smooth and small amplitudes. The forecasted probability distribution is smooth and continuous to provide users with instructive probability information for all cases, excluding some cases for extreme wave heights. Finally, a measure that calculates the percentage of the observed data included in the confidence region was used to objectively verify the proposed probabilistic forecasting method.

REFERENCES

Anandhi, A., Srinivas, V. V., Nanjundiah, R. S. & Nagesh Kumar, D. 2008 Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **28** (3), 401–420.

Babovic, V., Keijzer, M. & Bundzel, M. 2000 From global to local modelling: a case study in error correction of deterministic models. In: *Proceedings of the Fourth International Conference on Hydroinformatics*, Cedar Rapids, Iowa, USA.

Bray, M. & Han, D. 2004 Identification of support vector machines for runoff modeling. *Journal of Hydroinformatics* **6** (4), 265–280.

Bretschneider, C. L. & Tamaye, E. E. 1976 Hurricane wind and wave forecasting techniques. In: *Proceedings of the 15th Coastal Engineering Conference*, 1–17 July 1976, Hawaii, Vol. 1, pp. 202–237.

Chang, F. J. & Chen, Y. C. 2001 A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology* **245**, 153–164.

Chang, F. J., Chiang, Y. M. & Chang, L. C. 2007 Multi-step-ahead neural networks for flood forecasting. *Hydrological Sciences Journal* **52** (1), 114–130.

Chang, H. K. & Chien, W. A. 2006a A fuzzy–neural hybrid system of simulating typhoon waves. *Coastal Engineering* **53** (9), 737–748.

Chang, H. K. & Chien, W. A. 2006b Neural network with multi-trend simulating transfer function for forecasting typhoon wave. *Advances in Engineering Software* **37** (3), 184–194.

Chang, H. K., Liou, J. C., Liu, S. J. & Liaw, S. R. 2011 Simulated wave-driven ANN model for typhoon waves. *Advances in Engineering Software* **42** (1–2), 25–34.

Chen, S. T. 2013 Multiclass support vector classification to estimate typhoon rainfall distribution. *Disaster Advances* **6** (10), 110–121.

Chen, S. T. 2015 Mining informative hydrologic data by using support vector machines and elucidating mined data according to information entropy. *Entropy* **17** (3), 1023–1041.

Chen, Y. H. & Wang, H. 1983 Numerical model for nonstationary shallow water wave spectral transformations. *Journal of Geophysical Research: Oceans* **88** (C14), 9851–9863.

Chen, S. T. & Yu, P. S. 2007a Real-time probabilistic forecasting of flood stages. *Journal of Hydrology* **340** (1–2), 63–77.

Chen, S. T. & Yu, P. S. 2007b Pruning of support vector networks on flood forecasting. *Journal of Hydrology* **347** (1–2), 67–78.

Chen, S. T., Yu, P. S. & Tang, Y. H. 2010 Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology* **385** (1–4), 13–22.

Chen, S. T., Yu, P. S. & Liu, B. W. 2011 Comparison of neural network architectures and inputs for radar rainfall adjustment for typhoon events. *Journal of Hydrology* **405** (1–2), 150–160.

Chen, C. S., Jhong, Y. D., Wu, T. Y. & Chen, S. T. 2013 Typhoon event-based evolutionary fuzzy inference model for flood stage forecasting. *Journal of Hydrology* **490**, 134–143.

Choy, K. Y. & Chan, C. W. 2003 Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. *International Journal of Systems Science* **34** (14–15), 763–773.

Dawson, C. W. & Wilby, R. 1998 An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal* **43** (1), 47–66.

Deo, M. C. & Sridhar Naidu, C. 1998 Real time wave forecasting using neural networks. *Ocean Engineering* **26** (3), 191–203.

Deo, M. C., Jha, A., Chaphekar, A. S. & Ravikant, K. 2001 Neural networks for wave forecasting. *Ocean Engineering* **28** (7), 889–898.

Dibike, Y. B., Velickov, S., Solomatine, D. & Abbott, M. B. 2001 Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering* **15** (3), 208–216.

Donelan, M. A., Hamilton, J. & Hui, W. 1985 Directional spectra of wind-generated ocean waves. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **315** (1534), 509–562.

Elbisy, M. S. 2013 Sea wave parameters prediction by support vector machine using a genetic algorithm. *Journal of Coastal Research* **31** (4), 892–899.

French, M. N., Krajewski, W. F. & Cuykendall, R. R. 1992 Rainfall forecasting in space and time using a neural network. *Journal of Hydrology* **137** (1–4), 1–31.

Ghosh, S. & Mujumdar, P. P. 2008 Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources* **31** (1), 132–146.

Goda, Y. 2003 Revisiting Wilson's formulas for simplified wind-wave prediction. *Journal of Waterway, Port, Coastal, and Ocean Engineering* **129** (2), 93–95.

Han, D. & Cluckie, I. 2004 Support vector machines identification for runoff modeling. In: *Proceedings of the Sixth International Conference on Hydroinformatics, Singapore,* 21–24 June 2004 (S. Y. Liong, K. K. Phoon & V. Babovic, eds). World Scientific Publishing Co., Singapore, pp. 1597–1604.

Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. *Journal of Hydroinformatics* **9** (4), 267–276.

Hurdle, D. P. & Stive, R. J. H. 1989 Revision of SPM 1984 wave hindcast model to avoid inconsistencies in engineering applications. *Coastal Engineering* **12** (4), 339–351.

Ippen, A. T. 1996 *Estuary and Coastline Hydrodynamics*. McGraw-Hill, New York, USA.

Jhong, Y. D., Chen, C. S., Wu, W. Z. & Chen, S. T. 2018 Physical hybrid neural network model to forecast typhoon floods. *Water* **10** (5), 632.

Kazeminezhad, M. H., Etemad-Shahidi, A. & Mousavi, S. J. 2005 Application of fuzzy inference system in the prediction of wave parameters. *Ocean Engineering* **32** (14–15), 1709–1725.

Klir, G. J. & Yuan, B. 1995 *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, New Jersey, USA.

Krzysztofowicz, R. 2001 The case for probabilistic forecasting in hydrology. *Journal of Hydrology* **249**, 2–9.

Lin, G. F. & Wu, M. C. 2009 A hybrid neural network model for typhoon-rainfall forecasting. *Journal of Hydrology* **375**, 450–458.

Lin, G. F., Chen, G. R., Huang, P. Y. & Chou, Y. C. 2009 Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *Journal of Hydrology* **372**, 17–29.

Lin, G. F., Jhong, B. C. & Chang, C. C. 2013 Development of an effective data-driven model for hourly typhoon rainfall forecasting. *Journal of Hydrology* **495**, 52–63.

Liong, S. Y. & Sivapragasam, C. 2002 Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association* **38** (1), 173–196.

Mahjoobi, J. & Etemad-Shahidi, A. 2008 An alternative approach for the prediction of significant wave heights based on classification and regression trees. *Applied Ocean Research* **30** (3), 172–177.

Mahjoobi, J. & Mosabbeb, E. A. 2009 Prediction of significant wave height using regressive support vector machines. *Ocean Engineering* **36** (5), 339–347.

Mahjoobi, J., Etemad-Shahidi, A. & Kazeminezhad, M. H. 2008 Hindcasting of wave parameters using different soft computing methods. *Applied Ocean Research* **30** (1), 28–36.

Malekmohamadi, I., Bazargan-Lari, M. R., Kerachian, R., Nikoo, M. R. & Fallahnia, M. 2011 Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. *Ocean Engineering* **38** (2–3), 487–497.

Mandal, S. & Prabaharan, N. 2006 Ocean wave forecasting using recurrent neural networks. *Ocean Engineering* **33** (10), 1401–1410.

Nayak, P. C., Sudheer, K. P., Rangan, D. M. & Ramasastri, K. S. 2004 A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* **291** (1–2), 52–66.

Özger, M. & Şen, Z. 2007 Prediction of wave parameters by using fuzzy logic approach. *Ocean Engineering* **34** (3–4), 460–469.

Sannasiraj, S. A., Zhang, H., Babovic, V. & Chan, E. S. 2004 Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Advances in Water Resources* **27** (7), 761–772.

Sun, Y., Babovic, V. & Chan, E. S. 2010 Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *Journal of Hydrology* **395** (1–2), 109–116.

SWAMP Group 2013 *Ocean Wave Modeling*. Springer Science & Business Media, Berlin, Germany.

Tripathi, S., Srinivas, V. V. & Nanjundiah, R. S. 2006 Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330** (3–4), 621–640.

Tsai, C. P., Lin, C. & Shen, J. N. 2002 Neural network for wave forecasting among multi-stations. *Ocean Engineering* **29** (13), 1683–1695.

Tsai, C. C., Hou, T. H., Popinet, S. & Chao, Y. Y. 2013 Prediction of waves generated by tropical cyclones with a quadtree-adaptive model. *Coastal Engineering* **77**, 108–119.

Tsai, C. C., Wei, C. C., Hou, T. H. & Hsu, T. W. 2017 Artificial neural network for forecasting wave heights along a ship's route during hurricanes. *Journal of Waterway, Port, Coastal, and Ocean Engineering* **144** (2), 04017042.

Vapnik, V. N. 1995 *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, USA.

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

Vapnik, V. N. 1999 An overview of statistical learning theory. *IEEE Transactions on Neural Networks* **10** (5), 988–999.

WAMDI Group 1988 The WAM model – a third generation ocean wave prediction model. *Journal of Physical Oceanography* **18** (12), 1775–1810.

Wei, C. C. 2018 Nearshore wave predictions using data mining techniques during typhoons: a case study near Taiwan's northeastern coast. *Energies* **11** (1), 11.

Yang, T. C., Yu, P. S., Wei, C. M. & Chen, S. T. 2011 Projection of climate change for daily precipitation: a case study in Shih-Men Reservoir catchment in Taiwan. *Hydrological Processes* **25** (8), 1342–1354.

Yarar, A. 2014 A hybrid wavelet and neuro-fuzzy model for forecasting the monthly streamflow data. *Water Resources Management* **28** (2), 553–565.

Yu, P. S. & Chen, S. T. 2005 Updating real-time flood forecasting using a fuzzy rule-based model. *Hydrological Sciences Journal* **50** (2), 265–278.

Yu, P. S., Chen, C. J. & Chen, S. J. 2000 Application of gray and fuzzy methods for rainfall forecasting. *Journal of Hydrologic Engineering* **5** (4), 339–345.

Yu, P. S., Chen, S. T., Chen, C. J. & Yang, T. C. 2005 The potential of fuzzy multi-objective model for rainfall forecasting from typhoons. *Natural Hazards* **34** (2), 131–150.

Yu, P. S., Chen, S. T. & Chang, I. F. 2006 Support vector regression for real-time flood stage forecasting. *Journal of Hydrology* **328** (3–4), 704–716.

Yu, P. S., Yang, T. C., Chen, S. Y., Kuo, C. M. & Tseng, H. W. 2017 Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology* **552**, 92–104.

Zamani, A., Solomatine, D., Azimian, A. & Heemink, A. 2008 Learning from data for wind-wave forecasting. *Ocean Engineering* **35** (10), 953–962.

First received 14 October 2018; accepted in revised form 28 December 2018. Available online 4 February 2019