

Article

Nearshore Wave Predictions Using Data Mining Techniques during Typhoons: A Case Study near Taiwan's Northeastern Coast

Chih-Chiang Wei 

Department of Marine Environmental Informatics, National Taiwan Ocean University, No. 2, Beining Rd., Jhonghng District, Keelung City 20224, Taiwan; ccwei@ntou.edu.tw; Tel.: +886-2-2462-2192

Received: 23 October 2017; Accepted: 19 December 2017; Published: 21 December 2017

Abstract: Seasonal typhoons provide energy into the wave field in summer and autumn in Taiwan. Typhoons lead to abundant wave energy near the coastal area and cause storm surges that can destroy offshore facilities. The potential for wave energy can be obtained from analyzing the wave height. To develop an effective model for predicting typhoon-induced wave height near coastal areas, this study employed various popular data mining models—namely k-nearest neighbors (kNN), linear regressions (LR), model trees (M5), multilayer perceptron (MLP) neural network, and support vector regression (SVR) algorithms—as forecasting techniques. The principal component analysis (PCA) was then performed to reduce the potential variables from the original data at the first stage of data preprocessing. The experimental site was the Longdong buoy off the northeastern coast of Taiwan. Data on typhoons that occurred during 2002–2011 and 2012–2013 were collected for training and testing, respectively. This study designed four PCA cases, namely EV1, TV90, TV95, and ORI: EV1 used eigenvalues higher than 1.0 as principal components; TV90 and TV95 used the total variance percentages of 90% and 95%, respectively; and ORI used the original data. The forecast horizons varying from 1 h to 6 h were evaluated. The results show that (1) in the PCA model cases, when the number of attributes decreases, computing time decreases and prediction error increases; (2) regarding classified wave heights, M5 provides excellent outcomes at the small wavelet wavelet level; MLP has favorable outcomes at the large wavelet and small/moderate wave levels; meanwhile, SVR gives optimal outcomes at the long wave and high/very high wave levels; and (3) for performance of lead times, MLP and SVR achieve more favorable relative weighted performance without consideration of computational complexity; however, MLP and SVR might obtain lower performance when computational complexity is considered.

Keywords: wave height; data mining; prediction; typhoon

1. Introduction

Taiwan is situated on one of the main paths of western North Pacific typhoons and is affected by an average of four typhoons each year [1]. As a typhoon approaches Taiwan, the sea level rises abnormally due to storm surges caused by strong winds and atmospheric pressure disturbances [2]. Typhoons lead to abundant wave energy near the coastal area and cause storm surges that can destroy offshore facilities.

The central mountain range (CMR, shown in Figure 1), which is 340 km long and 80 km wide with an average height of 2500 m, is Taiwan's principal mountain range. As a typhoon approaches Taiwan, the topography significantly influences its track and intensity [3,4]. Consequently, typhoon circulation interaction with the topography produces considerable mesoscale variations in pressure, wind, precipitation, and offshore waves near Taiwan [5–8]. Typhoon-complicated wind distribution often causes storm surges that can destroy offshore facilities and coastal infrastructure as well as

ships [2]. The potential for wave energy can be obtained from analyzing the wave climate [9]. Therefore, accurately simulating the wave fields produced by typhoons is vital for preventing damage and understanding the risks posed by this natural hazard [10].

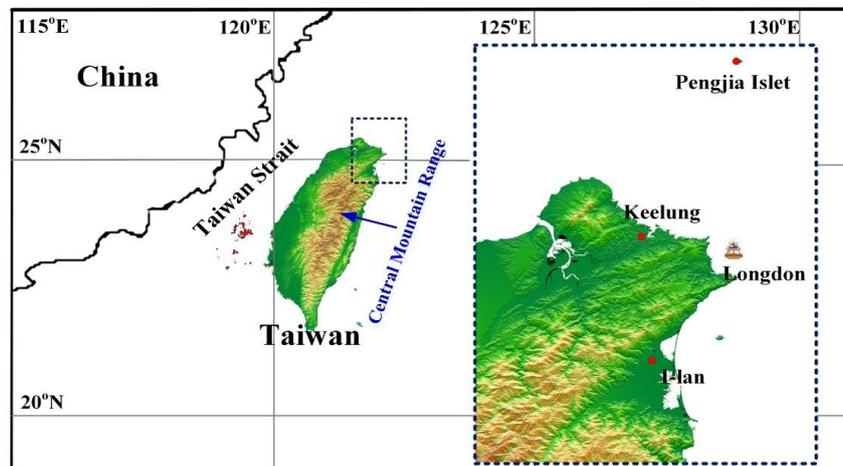


Figure 1. Location of the meteorological stations and a buoy.

A useful scheme for nearshore wave forecasts during typhoons is highly desirable in Taiwan. Data mining, which has popularized data-driven models, has been in widespread use in coastal and oceanic applications. These methods are based on the analysis of all data characterizing the system under study to determine an unknown mapping or dependencies between the system's input and output from the available data [11]. Numerous applications are used in water-related modeling, including some in wind-wave modeling [12–15]. For example, Londhe [16] used artificial neural networks (ANNs) and genetic programming for the estimation of missing wave heights in the eastern Gulf of Mexico. Etemad-Shahidi and Mahjoobi [17] used model trees (M5) for the same purpose, and the results were compared with those of ANNs. Tsai and Tsai [18] established a backpropagation (BP) neural network model to predict four surface waves by using acoustic wave and pressure data measured by ultrasonic wave gauges. Deka and Prahlada [19] developed a hybrid model of wavelets and an ANN to forecast significant wave heights for higher lead times up to 48 h on the west coast of India. Asma et al. [20] used multiple linear regressions (LR) and ANNs to describe the significant wave height off Goa, located on the west coast of India. The aforementioned studies have been widely used in hindcasting and forecasting of wave parameters. They indicated that neural networks can provide a good alternative to statistical regression, time series analysis and numerical methods. The advantages are due to the improved accuracy, simplicity, smaller computational efforts and in some cases less data requirements. However, the complicated meteorological conditions resulting from interactions between typhoons and terrain effects in Taiwan continue to pose tremendous challenges to current wave models in the predicting of accurate wave heights.

The presented data mining models are designed for hindcasting significant wave heights for nearshore during tropical cyclones (typhoons) in the paper. As mentioned, when a typhoon approaches Taiwan, the nearshore waves are significantly interacted by a typhoon's circulation and topography factor. This study developed a methodology to address the complicated problem of wave forecasts during typhoons, and designed a series of steps to deal with the modeling process (see Section 3). The novelty of this research has two main points. First, this study investigated several observations from meteorological stations, which are near the study site (a buoy), and collected the typhoon characteristics, and buoy atmospheric and maritime properties. Second, because a large number of attributes would be employed as model inputs, this study developed an improved approach that combines the principal component analysis (PCA) technique and data mining models. Some researchers, such as [21–24], concluded that PCA can be employed to find a set of orthogonal

components that minimize the error in the reconstructed data. This paper presents popular data mining techniques, including k-nearest neighbors (kNN), LR, M5, multilayer perceptron (MLP) neural network, and support vector regressions (SVR) that are often used for ensemble forecasting problems. The paper selected these models because the kNN, a typical non-parametric classifier, involves using neighbor search algorithms to achieve computational tractability [25]. LR is a traditional regression analysis modeling technique, and can be used as a benchmark model. M5, developed by [26], is a technique for dealing with continuous class learning problems that provides a structural representation of the data and a piecewise linear fit of the class. An ANN-based MLP network is extensively used to model an unknown system with observable inputs and outputs, and widely employed because of its flexibility and ease of use. Support vector machines (SVM), developed by [27], adhere to the principle of structural risk minimization seeking to minimize an upper bound of generalization error. SVM has been extended to solve nonlinear regression estimation problems, known as SVR. These models were employed for forecasting hourly wave heights during typhoon invasions.

The remainder of this paper is organized as follows. Section 2 describes the study area and selected typhoons and data. Section 3 describes the proposed prediction model. Section 4 outlines the model development regarding PCA approach and data mining models. Section 5 evaluates the performance of predicted results, and various lead time forecasts. Finally, Section 6 offers some conclusions.

2. Location and Typhoons

Figure 1 shows the experimental area and location (121.92° E, 25.10° N) of the Longdong buoy near Taiwan's northeastern coast. As mentioned, when a typhoon approaches Taiwan, the terrain of the CMR significantly affects a typhoon's circulation and its track. Therefore, this study investigated three meteorological stations, namely Keelung, Penjia Islet, and I-lan, which are near the Longdong buoy, and collected their meteorological observations.

The typhoon data used in this study were recorded in the western Pacific Ocean, northeast of Taiwan. Table A1 of Appendix A lists 67 typhoons affecting offshore northeast Taiwan in the years 2002–2013 and their corresponding wind intensity scales. The Saffir–Simpson wind scale (Table 1) classifies tropical cyclones into seven categories (i.e., categories 1–5, tropical storm, and tropical depression) depending on the intensities of their maximum sustained winds. Table 1 shows the classification of the 67 typhoons into these seven categories. In total, 24 typhoons were classified as tropical storms, and 10 as category 3, the strongest among the studied typhoons. Whenever a typhoon approached the study area, various meteorological and oceanic characteristics are measured by the Central Weather Bureau (CWB) of Taiwan.

Table 1. Hurricane wind scale and number of 67 collected typhoon events.

Wind Intensity Scale		Ranges	Event Amounts
Saffir–Simpson wind category	Category 5	≥70 m/s	0
	Category 4	58–70 m/s	0
	Category 3	50–58 m/s	10
	Category 2	43–50 m/s	16
	Category 1	33–43 m/s	17
Additional classifications	Tropical storm	18–33 m/s	24
	Tropical depression	<18 m/s	0

The data comprised 3654 records measured at hourly intervals. The mean, minimal, and maximal values of the data attributes are shown in Table 2. The attributes used in this study can be divided into four.

Table 2. Range and average values of data attributes.

Attribute	Min–Max, Mean	Attribute	Min–Max, Mean	Attribute	Min–Max, Mean
h_1	15.0–28.5, 22.5	$m_{2,1}$	950.2–1007.4, 990.6	$m_{3,7}$	0.1–25.2, 3.8
h_2	113.9–131.4, 122.4	$m_{2,2}$	961.6–1019.7, 1002.5	$m_{3,8}$	0.0–360.0, 161.8
h_3	910.0–1000.0, 966.8	$m_{2,3}$	16.4–32.7, 26.3	$m_{3,9}$	0.0–70.5, 1.8
h_4	0.0–300.0, 201.1	$m_{2,4}$	13.7–29.3, 24.2	$m_{3,10}$	0.0–1.0, 0.4
h_5	0.0–50.0, 16.5	$m_{2,5}$	58.0–100.0, 88.7	$m_{3,11}$	0.0–3.8, 0.4
h_6	15.0–55.0, 33.6	$m_{2,6}$	15.7–40.8, 30.4	p_1	1.1–26.7, 7.0
$m_{1,1}$	956.9–1014.8, 998.5	$m_{2,7}$	0.0–50.3, 11.7	p_2	1.0–360.0, 174.5
$m_{1,2}$	959.9–1018.1, 1001.6	$m_{2,8}$	0.0–360.0, 130.1	p_3	1.6–36.1, 10.4
$m_{1,3}$	15.6–36.7, 27.4	$m_{2,9}$	0.0–186.0, 1.4	p_4	16.1–30.6, 26.4
$m_{1,4}$	14.5–27.4, 23.4	$m_{2,10}$	0.0–1.0, 0.3	p_5	955.6–1017.4, 1001.2
$m_{1,5}$	49.0–100.0, 79.6	$m_{2,11}$	0.0–3.9, 0.5	b_1	0.2–12.8, 2.2
$m_{1,6}$	16.5–36.5, 28.9	$m_{3,1}$	954.5–1018.0, 1000.7	b_2	28.0–193.0, 102.2
$m_{1,7}$	0.0–24.0, 4.9	$m_{3,2}$	955.4–1018.9, 1001.6	b_3	33.0–129.0, 67.5
$m_{1,8}$	0.0–360.0, 123.1	$m_{3,3}$	17.0–38.1, 27.3	b_4	11.0–337.0, 90.4
$m_{1,9}$	0.0–75.5, 1.4	$m_{3,4}$	14.7–28.9, 23.6	b_5	0.0–33.9, 26.8
$m_{1,10}$	0.0–1.0, 0.4	$m_{3,5}$	38.0–100.0, 81.0	-	-
$m_{1,11}$	0.0–3.9, 0.4	$m_{3,6}$	16.7–39.8, 29.2	-	-

2.1. Typhoon Characteristics

The attributes of typhoon characteristics, denoted as $\{H\}$, are latitude and longitude of the typhoon center (unit: deg) (denoted as h_1 and h_2), pressure at the typhoon center (mb) (h_3), radius of the typhoon (km) (h_4), moving speed of the typhoon (km/h) (h_5), and intensity of the typhoon (m/s) (h_6). For the 10 category 3 typhoons, the mean value of typhoon radius is 237.1 km (which is 1.18 times the average of all records) and the mean value of typhoon intensity is 45.7 m/s (1.36 times).

2.2. Ground Meteorological Properties

The ground meteorological properties, denoted as $\{M\}$, are ground air pressure (hPa) (m_1), air pressure at sea level (hPa) (m_2), ground temperature ($^{\circ}\text{C}$) (m_3), ground dew point temperature ($^{\circ}\text{C}$) (m_4), ground relative humidity (%) (m_5), ground vapor pressure (hPa) (m_6), surface wind velocity (maximum 10-min mean, 10 m above the surface) (m/s) (m_7), surface wind direction (deg) (m_8), ground precipitation (mm/h) (m_9), rainfall duration within 1 h (h) (m_{10}), and ground global solar radiation ($\text{MJ}\cdot\text{m}^{-2}$) (m_{11}). In the table, the suffix of the subset $\{m_{i,j}\}$ refers to the ground meteorological attribute j at meteorological station i (where $i = 1, 2$, and 3 indicate Keelung, Penjia Islet, and I-lan stations, respectively). For the 10 category 3 typhoons, the mean values of surface wind velocity at Keelung, Penjia Islet, and I-lan stations are 6.6, 12.9, and 4.9 m/s (which are 1.34, 1.11, and 1.28 times the average of all records), respectively.

2.3. Buoy Atmospheric Properties

The atmospheric properties of the Longdong buoy at 3 m height, denoted as $\{P\}$, are wind speed at the buoy (m/s) (p_1), wind direction at the buoy (deg) (p_2), gusty wind speed at the buoy (m/s) (p_3), temperature above the buoy ($^{\circ}\text{C}$) (p_4), and atmospheric pressure above the buoy (hPa) (p_5). For the 10 category 3 typhoons, the mean value of wind speed at the buoy is 8.5 m/s (which is 1.21 times the average of all records).

2.4. Buoy Maritime Properties

The maritime properties of the Longdong buoy, denoted as $\{B\}$, are wave height (h) at the buoy (m) (b_1), maximum wave period (s) (b_2), average wave period (s) (b_3), wave direction (deg) (b_4), and sea surface temperature ($^{\circ}\text{C}$) (b_5). Attribute b_1 refers to significant wave height, $h_{1/3}$. Significant wave height is defined as the average of the highest one-third of the crest to trough waves in that segment of time series by sifting

through each individual trough to crest waves in the data [28]. For the 10 category 3 typhoons, the mean value of wave height at the buoy is 2.8 m (which is 1.31 times the average of all records).

3. Methodology

This section presents a methodology for developing a usable scheme for forecasting wave heights during typhoon periods. As described in the previous section, the model inputs comprise the six attributes $\{H\}$ for typhoon characteristics, 33 attributes $\{M\}$ at three ground meteorological stations (11 attributes per station), and 10 attributes at a buoy (five attributes $\{P\}$ for buoy atmospheric properties and five attributes $\{B\}$ for buoy maritime properties). In total, 50 input attributes are used for modeling. First, for lead time = 1 h, the model target is the 1-h ahead observed wave heights $\{b_{1,t+1}\}$ at the buoy. To reduce the dimension of model inputs, the study employed the PCA approach (theorem see Section 3.1) to generate input–output relationships. PCA can gather highly correlated independent variables into a principal component, and all principal components are independent of each other so that the analysis' only function is to transform a set of correlated variables into a set of uncorrelated principal components [29].

The main processing stages are plotted in Figure 2. First, the collected attributes are preprocessed as the model inputs. This study then employs the PCA approach to derive new variables from original attributes, and select the eigenvalues greater than a threshold as principal components. Thus, the new PCA-derived attributes can be obtained and used to formulate the wave-height data-driven prediction models using PCA-derived attributes. The five data mining algorithms, kNN, LR, M5, MLP, and SVR, are involved. Here, the height data-driven prediction models using the original attributes are employed as a benchmark model. To evaluate the utility of these data-driven approaches, results were compared with those from various model cases, namely EV1, TV90, TV95, and ORI: EV1 used eigenvalues higher than 1.0 as principal components; TV90 and TV95 used the total variance percentages of 90% and 95%, respectively; and ORI used the original data. To examine the robustness for the various lead times, the forecast horizons vary from 1 to 6 h in the nearshore of northeast Taiwan. Finally, the most accurate model is suggested as the prediction model. The theories of PCA technique and five data mining algorithms were described in the sequence sections.

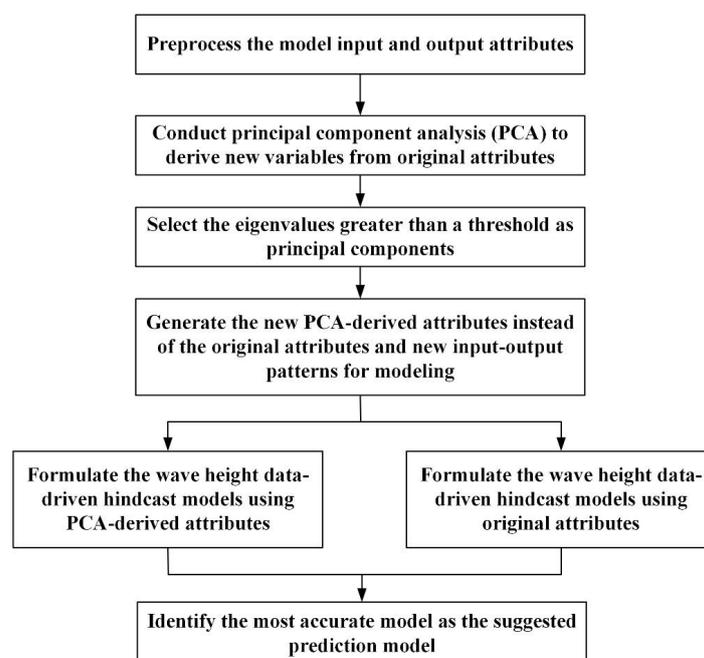


Figure 2. Flowchart of the proposed principal component analysis (PCA)-based data-driven wave height prediction models.

3.1. PCA Approach

In data analysis, reducing the dimensionality of data is vital because it helps understand the data and decreases computational cost [30]. The PCA approach is used in the proposed prediction model to simplify and orthogonalize the original data set in order to optimize the models because a small set of uncorrelated variables is easier to understand and use in further analyses than is a larger set of correlated variables [22].

PCA was first introduced by Pearson [31] and developed by Hotelling [32]. PCA is a linear decomposition of data that results in a set of eigenvectors that form the basis for a new coordinate system [33]. Principal components performed on a data matrix of (N samples \times P variables) yield linearly transformed random variables that exhibit special properties in terms of variances. In effect, transforming the original vector variables to the vector of principal components to a rotation of coordinate axes to a new coordinate system that possesses inherent statistical properties [34]. The first principal component (Y_1) can be defined as a linear combination of the elements of the data matrix, that is [23]:

$$Y_1 = e_{11}X_1 + e_{12}X_2 + \dots + e_{1P}X_P \quad (1)$$

where coefficients chosen to maximize the variance represented by the first principal component are simply the eigenvectors of the symmetric covariance matrix. The eigenvalues of the covariance matrix represent the variation of each principal component, where $Var(Y_i) = \lambda_i$. Ideally, a PCA yields several components that describe the majority of the total variation of the data set.

3.2. Data Mining Models

3.2.1. kNN

The difficulty in using kNN is determining the nearest k neighbors for a query point from a given data set. This difficulty occurs in scientific and engineering applications, including pattern recognition, data clustering, and function approximation [35]. kNN is a type of instance-based learning in which a function is only approximated locally, and all calculations are deferred until classification [36]. This method can be used for regression by simply assigning the property value for an object as the average of the values of its kNN [37]. The kNN can be defined by [38]. Given a set of N points, $[p_i]_{i=1}^N$, defined in real d -dimensional space, $S \subset \mathbb{R}^d$, and a query point, $q \in \mathbb{R}^d$, the k points of S with a minimal Euclidean distance to q are determined, where $k \geq 1$. This problem is solved simply by calculating the distances between the N points in S to the query point q . The k points with the smallest distances comprise the desired subset of S ; that is, the kNN to q .

3.2.2. LR

LR is the most widely used of all statistical techniques. LR estimates the most efficient-fitting linear equation for predicting the output field according to the input fields. The regression equation represents a straight line or plane that minimizes the squared differences between predicted and actual output values [39,40]. The multiple LR can be defined as

$$Y = X\beta + E \quad (2)$$

where Y is the $(p \times 1)$ vector of observations, X is the $(p \times q)$ vector containing q input factors, β is the $(q \times 1)$ vector containing the regression coefficients, and E is the $(p \times 1)$ vector containing the noise terms.

3.2.3. M5

M5 combines a conventional decision tree with the possibility of LR functions at the leaves [41]. An M5 tree-based model is constructed using the divide-and-conquer method. Splitting in M5 ceases when the class values of all the instances that reach a node vary only slightly, or when only a few

instances remain. The splitting criterion is based on treating the standard deviation of the class values that reach a node as a measure of the error at that node, and calculating the expected reduction in error as a result of testing each attribute at the node [42]. The attribute that maximizes the expected error reduction is chosen. Standard deviation reduction (SDR) is calculated by the formula

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (3)$$

where T is the set of examples that reaches the node, T_i is the subset of cases that have the i th outcome of the potential test, and $sd()$ is the function of the standard deviation.

3.2.4. MLP

ANNs are the information processing systems that are inspired by the models of biological neural networks. MLP is a feedforward network in which there may be one or more hidden layers besides one input and one output layer. Studies have suggested that one hidden layer may be sufficient for most problems [43–45]. All layers except the output layer contain a bias or threshold node whose output is set to a fixed value of 1. All the nodes of a lower layer are connected to all the nodes of an upper layer through links called weights. For MLP, the procedure used to perform the learning process is called the learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective [46]. The BP algorithm, a generalized steepest descent algorithm, is the most popular learning technique used to train the MLP. The weights of the MLP are updated by using the BP algorithm during the training phase. The knowledge acquired by the network after learning is stored in its weights in a distributed manner.

3.2.5. SVR

SVR is a learning machine based on statistical learning theory. Let

$$f(\mathbf{x}) = \omega^T \mathbf{x} + b\omega, \mathbf{x} \in R^d \quad (4)$$

where \mathbf{x} is the input vector; $f(\mathbf{x})$ is the output vector; ω and b are the slope and offset of the regression line, respectively; and R^d is d -dimensional input space. In SVR, the regression function is calculated by minimizing:

$$\frac{1}{2} \omega^T \omega + \frac{1}{M} \sum_{i=1}^M c(f(\mathbf{x}_i), y_i) \quad (5)$$

where $(1/2) \|\omega\|^2$ is the term characterizing the model complexity, M is the number of support vectors, and $c(f(\mathbf{x}_i), y_i)$ is the loss function determining how the distance between $f(\mathbf{x}_i)$ and the target values y_i should be penalized.

The constrained optimization problem can be reformulated into dual problem formalism by using Lagrange multipliers [27], which leads to the solution:

$$f(\mathbf{x}) = \sum_{i=1}^M (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (6)$$

where α_i and α_i^* are the Lagrange multipliers, and $K(\mathbf{x}_i, \mathbf{x})$ represents the kernel function. A suitable kernel function makes it possible to map a nonlinear input space to a high-dimensional feature space, in which LR can be performed [47]. The researcher adopted the prevalent radial basis function (RBF) kernel. The Euclidean distance-based RBF-kernel function is defined as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (7)$$

where σ is the standard deviation that determines the width of the RBF kernel.

4. Model Development

The collected typhoons were classified into training and testing sets when modeling. The training data (years 2002–2011, 56 typhoons) were used to build the models, and the testing data (years 2012–2013, 11 typhoons) were used to evaluate the performance levels of the built models.

4.1. PCA

PCA was performed on the original data to explore possibilities for data dimension reduction. Figure 3 plots the eigenvalue of components and total variance explained for wave-height predictions. On the basis of [48], the components with an eigenvalue greater than one are suggested to replace the original attribute values for further analysis. Thus, the researcher selected eigenvalues greater than 1.0 as principal components. As Figure 3 shows, 10 components are greater than 1.0, and the corresponding total variance percentage of these components is 77.29%. Through PCA, the PCA-derived components (variables) then replaced the original ones. This model case is called EV1 in the subsequent section. The relationships between new variables and original inputs can be expressed in a matrix form, such as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_{10} \end{bmatrix} = \begin{bmatrix} 0.170 & -0.012 & -0.054 & 0.089 & \cdots & -0.005 \\ 0.019 & 0.072 & 0.011 & 0.003 & \cdots & 0.251 \\ -0.048 & -0.201 & -0.351 & 0.302 & \cdots & 0.056 \\ -0.269 & -0.052 & -0.114 & 0.036 & \cdots & -0.050 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -0.026 & -0.042 & 0.105 & -0.134 & \cdots & 0.166 \end{bmatrix} \begin{bmatrix} h'_1 \\ h'_2 \\ h'_3 \\ h'_4 \\ \vdots \\ b'_5 \end{bmatrix} \quad (8)$$

where h'_1 is the normalized input h_1 , and y_j are the PCA-derived inputs ($j = 1, 10$).

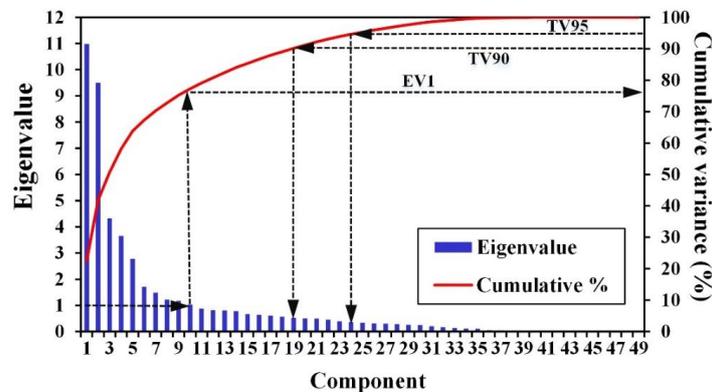


Figure 3. Eigenvalue and total variance explained for 1-h ahead wave-height predictions.

To compare with EV1, the researcher also selected the total variance percentage of 90% and 95%, whose corresponding components are 19 and 25, respectively (Figure 3). The two model cases were named TV90 and TV95. For TV90, the relationships between new variables and original inputs was expressed as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_{19} \end{bmatrix} = \begin{bmatrix} 0.170 & -0.012 & -0.054 & 0.089 & \cdots & -0.005 \\ 0.019 & 0.072 & 0.011 & 0.003 & \cdots & 0.251 \\ -0.048 & -0.201 & -0.351 & 0.302 & \cdots & 0.056 \\ -0.269 & -0.052 & -0.114 & 0.036 & \cdots & -0.050 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -0.089 & 0.019 & 0.066 & 0.200 & \cdots & -0.056 \end{bmatrix} \begin{bmatrix} h'_1 \\ h'_2 \\ h'_3 \\ h'_4 \\ \vdots \\ b'_5 \end{bmatrix} \quad (9)$$

For TV95, the relationships between new variables and original inputs was expressed as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_{25} \end{bmatrix} = \begin{bmatrix} 0.170 & -0.012 & -0.054 & 0.089 & \cdots & -0.005 \\ 0.019 & 0.072 & 0.011 & 0.003 & \cdots & 0.251 \\ -0.048 & -0.201 & -0.351 & 0.302 & \cdots & 0.056 \\ -0.269 & -0.052 & -0.114 & 0.036 & \cdots & -0.050 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -0.002 & 0.162 & 0.051 & 0.044 & \cdots & 0.104 \end{bmatrix} \begin{bmatrix} h'_1 \\ h'_2 \\ h'_3 \\ h'_4 \\ \vdots \\ b'_5 \end{bmatrix} \quad (10)$$

Additionally, the model case that used the original data was designed as another case, namely, ORI.

4.2. Model Constructions

The parameter calibrations of various model cases were subjected to sensitivity analysis. The root mean square error (RMSE) was used to measure the errors of the results, that is,

$$\text{RMSE} = \sqrt{\frac{1}{n_r} \sum_{i=1}^{n_r} (O_i^{\text{pre}} - O_i^{\text{obs}})^2} \quad (11)$$

where O_i^{pre} is the prediction for record i , O_i^{obs} is the observation for record i , and n_r is the number of records. The smaller the RMSE criteria, the more favorable is the performance of the predicted outcome.

First, the processes of modeling the EV1 are described. For kNN, the favorable choice of k depends on the data. In this study, the k neighbor points were chosen by performing sensitivity analysis. Figure 4a plots the RMSE calculations using the testing data set. In the figure, the optimal RMSE value was at $k = 14$. For LR, the researcher used a stepwise regression method and specified selection criteria based on the statistical probability (p value) associated with each field. The criteria and stepwise estimation were used to add and remove fields. The p value used was 0.05.

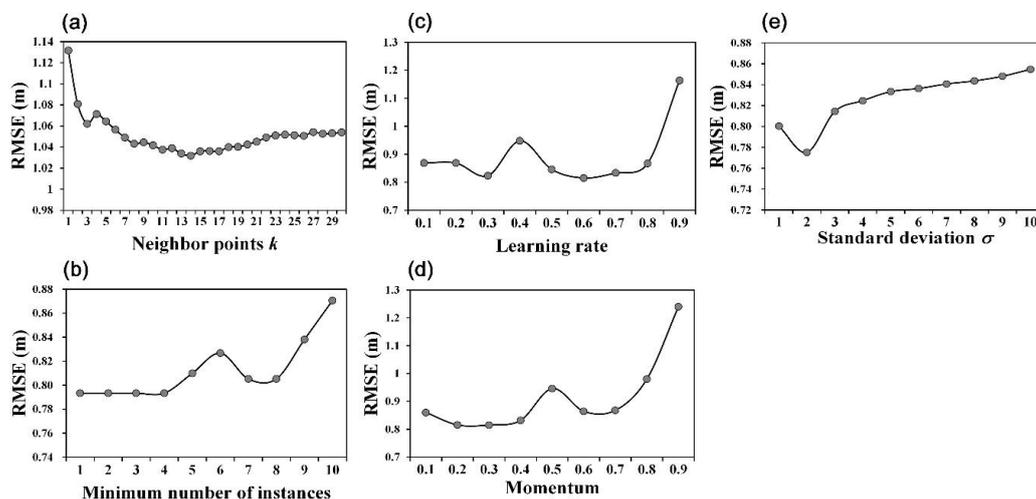


Figure 4. Sensitivity of model parameters on the root mean square error (RMSE): (a) neighbor points for k -nearest neighbors (kNN), (b) minimum number of instances for M5, (c) learning rate for multilayer perceptron (MLP), (d) momentum for MLP, and (e) standard deviation for support vector regression (SVR).

For M5, the minimum number of instances to allow at a leaf node were chosen by performing sensitivity analysis. According to Figure 4b, the suitable minimum number of instances can be chosen between 1 and 4. To train an MLP network, a three-layer feedforward network was created. This study introduced sigmoid and linear activity functions in the hidden and output layers, respectively. The number of neurons in the hidden layer was determined according to the method proposed by [49], namely adding

up the number of neurons in the input and output layers, minus the sum by 1, and dividing this number by 2. In addition, the parameters of learning rate and momentum were validated according to a sensitivity analysis. According to Figure 4c,d, the suitable learning rate and momentum can be set at 0.6 and 0.3, respectively. For SVR, the parameters involve the penalty parameter and small positive number, in which both parameters are embedded in dual problem formalism. Here, this study sets penalty parameter and small positive number at 1.0 and 0.001, respectively. As mentioned, the RBF kernel was employed. According to sensitivity analysis (Figure 4e), the favor σ value can be set at 2. The other model cases (ORI, TV95, and TV90) were calibrated by using the same processes as those in EV1.

5. Evaluation

As stated in the previous section, the various trained models were validated by the 11 testing typhoons (years 2012–2013). Figure 5 plots the prediction results of testing typhoons for four PCA cases using five models.

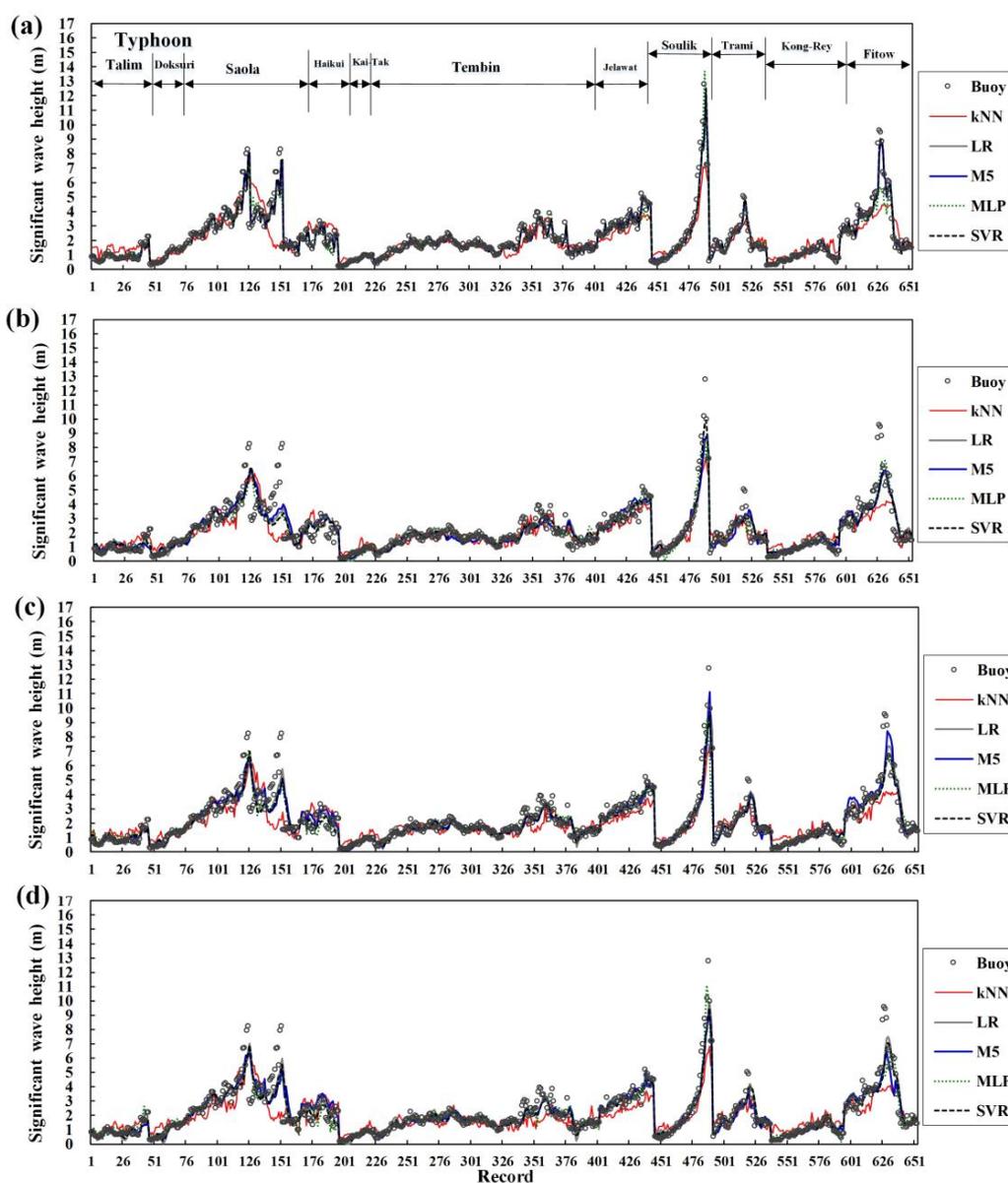


Figure 5. Prediction results of testing typhoons (2012–2013) for (a) ORI, (b) EV1, (c) TV90, and (d) TV95. LR: linear regressions; M5: model trees; MLP: multilayer perceptron; SVR: support vector regression.

5.1. Effect of Dimension Reduction

First, the researcher evaluated the effect of data dimension reduction using a PCA method on the prediction ability and computation efficiency. Figure 6 illustrates the relationships among RMSE prediction errors, computational time, and four designed cases by using a testing set. The total numbers of attributes (comprising input attributes and a target) are 11, 20, 26, and 51 for EV1, TV90, TV95, and ORI, respectively. As seen in these subfigures, when the number of model attributes decreases, the computing efficiency increases because of a decrease in computational time. For the kNN model, the computing time increased from 0.8 to 2.0 s using the four cases. For other models, the computing times range of 0.02–0.78 s, 2.95–4.70 s, 7.0–237.94 s, and 22.31–570.43 s were for the LR, M5, MLP, and SVR models, respectively. This study found that the computing times for kNN, LR, and M5 are short regardless of input attribute amounts. By contrast, the computing time on MLP and SVR rise considerably when the number of input attributes increases. For prediction errors, the results show that the RMSE values are in the range 0.988–1.032 m, 0.555–0.816 m, 0.553–0.793 m, 0.553–0.766 m, and 0.557–0.775 m for the kNN, LR, M5, MLP, and SVR models, respectively. LR takes advantage of computational time, because LR whose regressions depend linearly on their unknown parameters are easier to fit than are models (such as MLP and SVR) that are nonlinearly related to their parameters. Moreover, the researcher found that when the number of model attributes decreases, the number of prediction errors increase. The reason for this might be that although PCA processes data dimension reduction, it may partially lose pattern information.

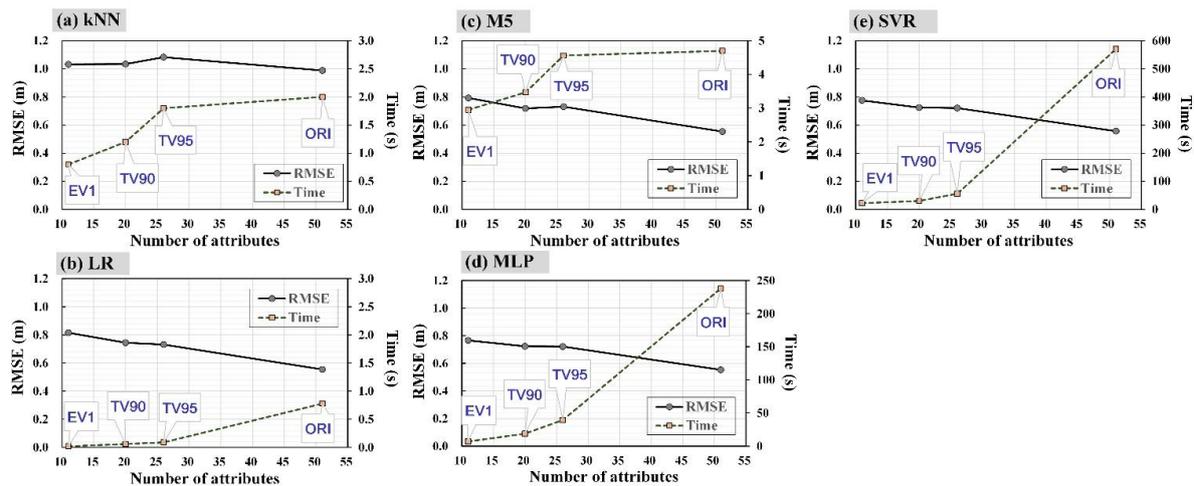


Figure 6. Effect of data reduction on computational time and prediction errors for (a) kNN, (b) LR, (c) M5, (d) MLP, and (e) SVR.

To easily present the overall performance with respect to the prediction errors and computational time, Table 3 lists the average of RMSE measures and computational times of the four PCA cases. As Table 3 shows, (1) in terms of prediction errors, the researcher determined that the optimal one occurs at MLP (RMSE = 0.691 m), and the secondary favorable one occurs at SVR (RMSE = 0.694 m), in which the RMSE is close to that in MLP; meanwhile, the least favorable one appears at kNN (RMSE = 1.034 m). (2) In terms of computational time, the favorable one is LR (0.2 s), and the most time-consuming is SVR (169.6 s).

Table 3. Average performance of RMSE and computational time of four PCA cases on five models.

Performance	kNN	LR	M5	MLP	SVR
Time (s)	1.5	0.2	3.9	75.8	169.6
RMSE (m)	1.034	0.712	0.699	0.691	0.694

5.2. Performance of Classified Wave Heights

5.2.1. Error Measures

To evaluate the performance of different wave strengths which are defined by CWB, the wave heights were classified into five levels, namely, small wavelets (<0.6 m), large wavelets (0.6–1.5 m), small and moderate waves (1.5–2.5 m), long waves (2.5–5.0 m), and high and very high waves (>5.0 m). The criterion used to assess the wave strengths were the mean absolute error (MAE), RMSE and correlation coefficient (r). MAE is defined as:

$$\text{MAE} = \frac{1}{n_r} \sum_{i=1}^{n_r} |O_i^{\text{pre}} - O_i^{\text{obs}}| \quad (12)$$

Then, the r is defined as:

$$r = \frac{\sum_{i=1}^{n_r} (O_i^{\text{obs}} - \bar{O}^{\text{obs}}) (O_i^{\text{pre}} - \bar{O}^{\text{pre}})}{\sqrt{\sum_{i=1}^{n_r} (O_i^{\text{obs}} - \bar{O}^{\text{obs}})^2 \sum_{i=1}^{n_r} (O_i^{\text{pre}} - \bar{O}^{\text{pre}})^2}} \quad (13)$$

where \bar{O}^{obs} is the average of the observations, and \bar{O}^{pre} is the average of the predictions. Generally, lower MAE values indicate more accurate performance measures and larger r values indicate favorable performance measures.

Figure 7 illustrates the performance measures of the five levels of wave heights. The figure plots the performance measures regarding MAE (Figure 7a–d), RMSE (Figure 7e–h), and r (Figure 7i–l) for the four cases (ORI, EV1, TV90, and TV95). The results from MAE (Figure 7a–d) and RMSE (Figure 7e–h) show that (1) the highest predicted errors seem to appear at the high and very high wave level, whereas the lowest predicted errors occur at the small wavelet level. (2) The kNN resulted in more errors in the small wavelet, large wavelet, long wave, and high and very high wave levels (except for small and moderate wave levels) on the four model cases. Moreover, for the correlation between observed and predicted values (Figure 7i–l), the researcher determined that the highest r results might be at the large wavelet level, whereas the lowest r results are at the small and moderate wave level.

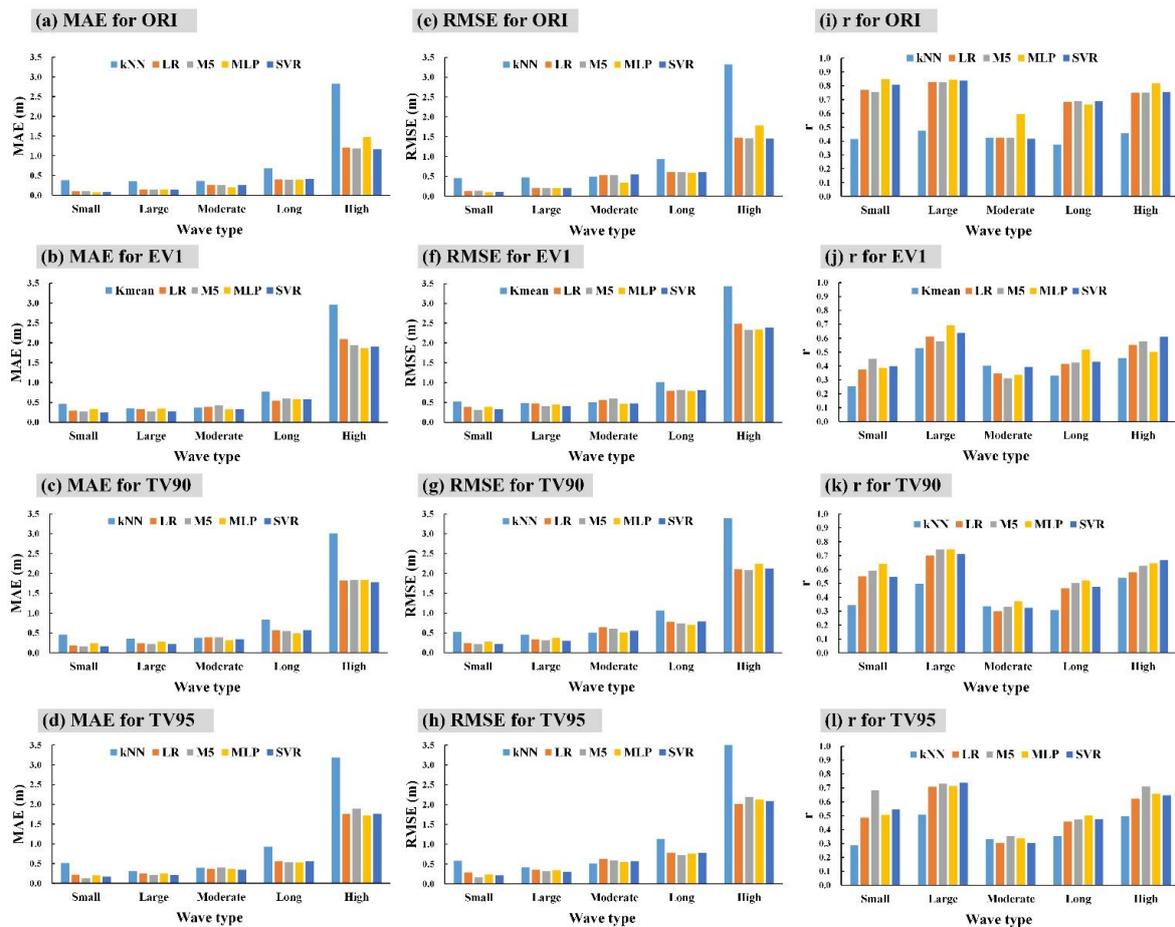


Figure 7. Performance measures of classified wave heights for the four PCA cases: (a–d) MAE, (e–h) RMSE, and (i–l) r. MAE: mean absolute error.

The analysis shows that the error measures from MAE and RMSE belong to “absolute errors.” To assess the “relative errors” regarding the classified wave-height predictions, the relative MAE (rMAE) and coefficient of variation of the RMSE (CV_{RMSE}) were calculated. rMAE is computed using

$$rMAE = \frac{MAE}{O_{obs}} \tag{14}$$

Then, the CV_{RMSE} is computed using

$$CV_{RMSE} = \frac{RMSE}{O_{obs}} \tag{15}$$

Generally, precise predictions are those whose rMAE and CV_{RMSE} are nearly 0. Figure 8 plots the performance rMAE and CV_{RMSE} measures of the classified wave heights for the four model cases. Results from rMAE (Figure 8a–d) and RMSE (Figure 8h) show that the highest relative predicted errors seem to appear at the small wavelet level, whereas the lowest predicted errors occur at the long wave level.

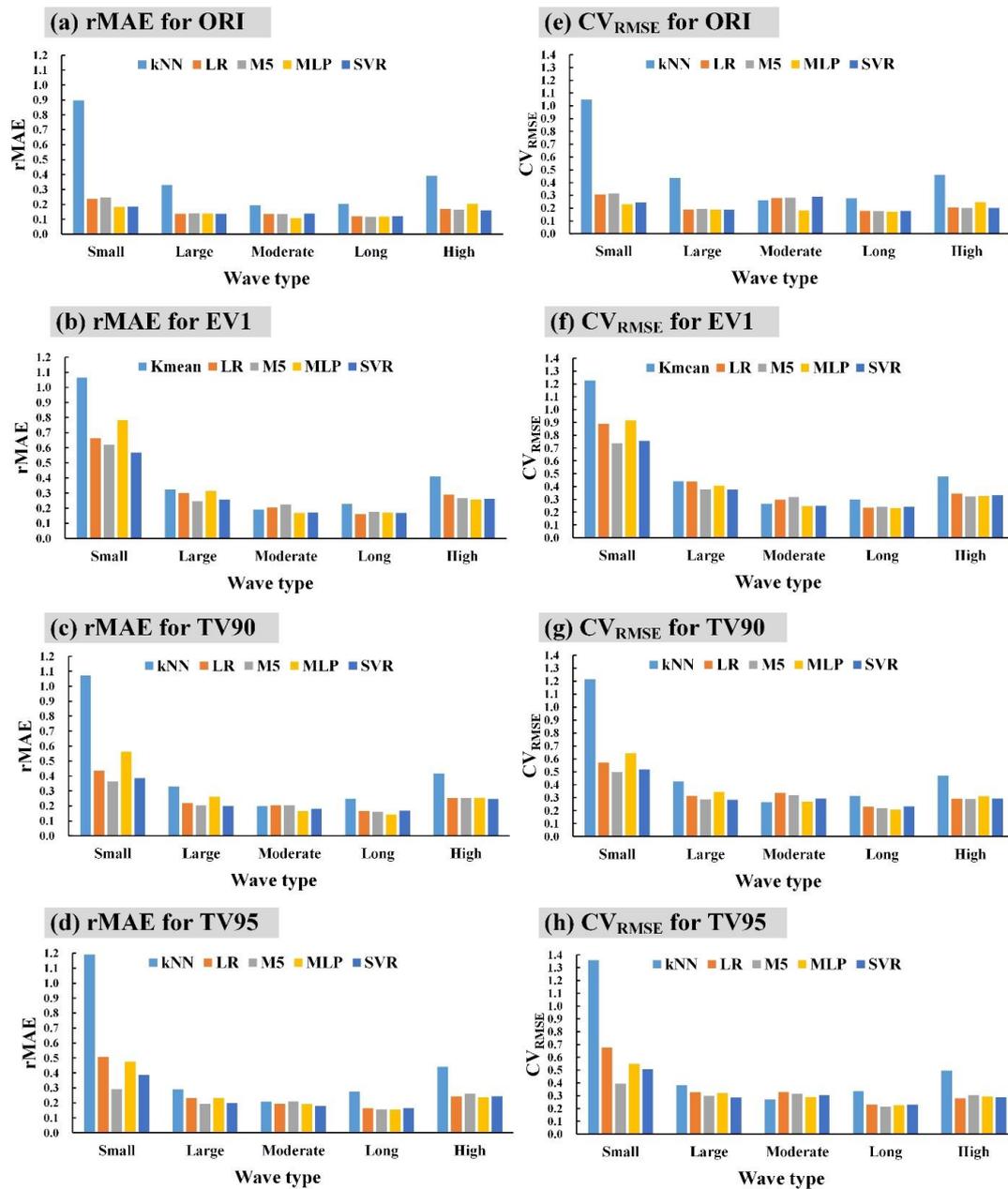


Figure 8. Performance measures of classified wave heights for the four PCA cases: (a–d) rMAE, and (e–h) CV_{RMSE} .

5.2.2. Performance Graphically Using Taylor Diagrams

In this section, the researcher employed Taylor diagrams to graphically indicate which of various prediction models is most realistic. The Taylor diagram, invented by [50], is an alternative means of understanding the relative merits of the forecast models. The diagram can provide a concise statistical summary of how well patterns match each other in terms of their correlation, their root-mean-square (RMS) difference, and the standard deviation.

The Taylor diagram shown in Figure 9 provides a summary of the relative skill with which five models simulate the various wave levels: (a) small wavelet, (b) large wavelet, (c) small/moderate wave, (d) long wave, and (e) high/very high wave. For Figure 9a, M5 agrees best with observations, with the lowest centered RMS error. Of the models, kNN has the lowest pattern correlation. For Figure 9b, MLP has a slightly higher correlation with observations and a slightly lower standard

deviation as the observed. Although kNN and LR have about the same centered RMS errors, LR predicts the wave heights much better than kNN, resulting in a higher correlation with observations. For Figure 9c, MLP has the lowest standard deviation, and the highest correlation with observations. For Figure 9d, SVR has a slightly higher correlation with observations and a slightly lower standard deviation as the observed. Finally, in Figure 9e, SVR, MLP, M5 and LR agree best with observations, each with about the same RMS errors. SVR, however, has a slightly higher correlation with observations. Hand et al. [51] reported that ANN-based models (e.g., MLP and SVR) can more easily overcome the high dimensionality problem, that is, the problem that arises when a high number of input variables are present relative to the number of available observations.

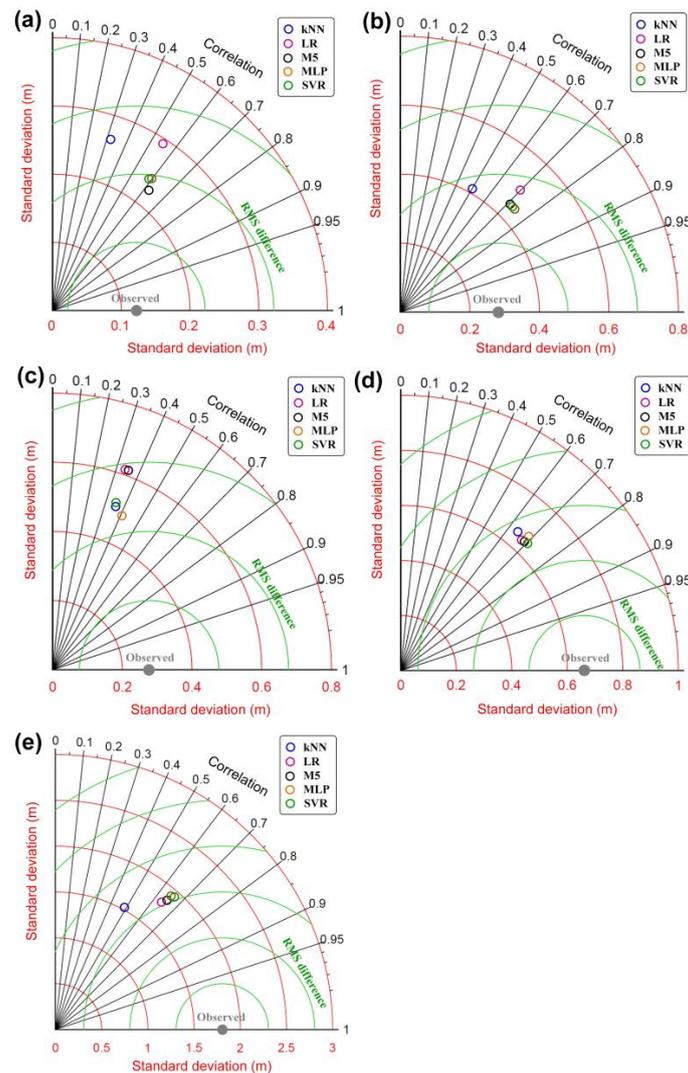


Figure 9. Taylor diagrams for: (a) small wavelet, (b) large wavelet, (c) small/moderate wave, (d) long wave, and (e) high/very high wave.

5.3. Evaluation of Various Lead Times

5.3.1. Error Measures

In this section, the researcher examines the forecast horizons varying from 1 to 6 h, and using the PCA case of TV90 as a demonstration. Figure 10 shows the observations and predictions at lead time from 2 h to 6 h (predictions of 1 h can see Figure 5c). In addition, Figure 11 shows the scattered plots contrasting the observations with the predictions at the lead time at 1, 3, and 6 h. The plots that the slopes calculated

are greater in the MLP model than in other models, and the kNN was the least correlation among the models. Figure 12 shows the model performance measures with corresponding $rMAE$, CV_{RMSE} , and r values to enable a comparison of forecast horizons varying from 1–6 h. According to figure, as expected, the researcher found that the 1-h-ahead predictions were more accurate than the 2-h-ahead predictions, and the 6-h-ahead predictions were the least accurate among the predictions. This indicated that increased forecasting horizons yielded increased errors in the prediction of wave heights.

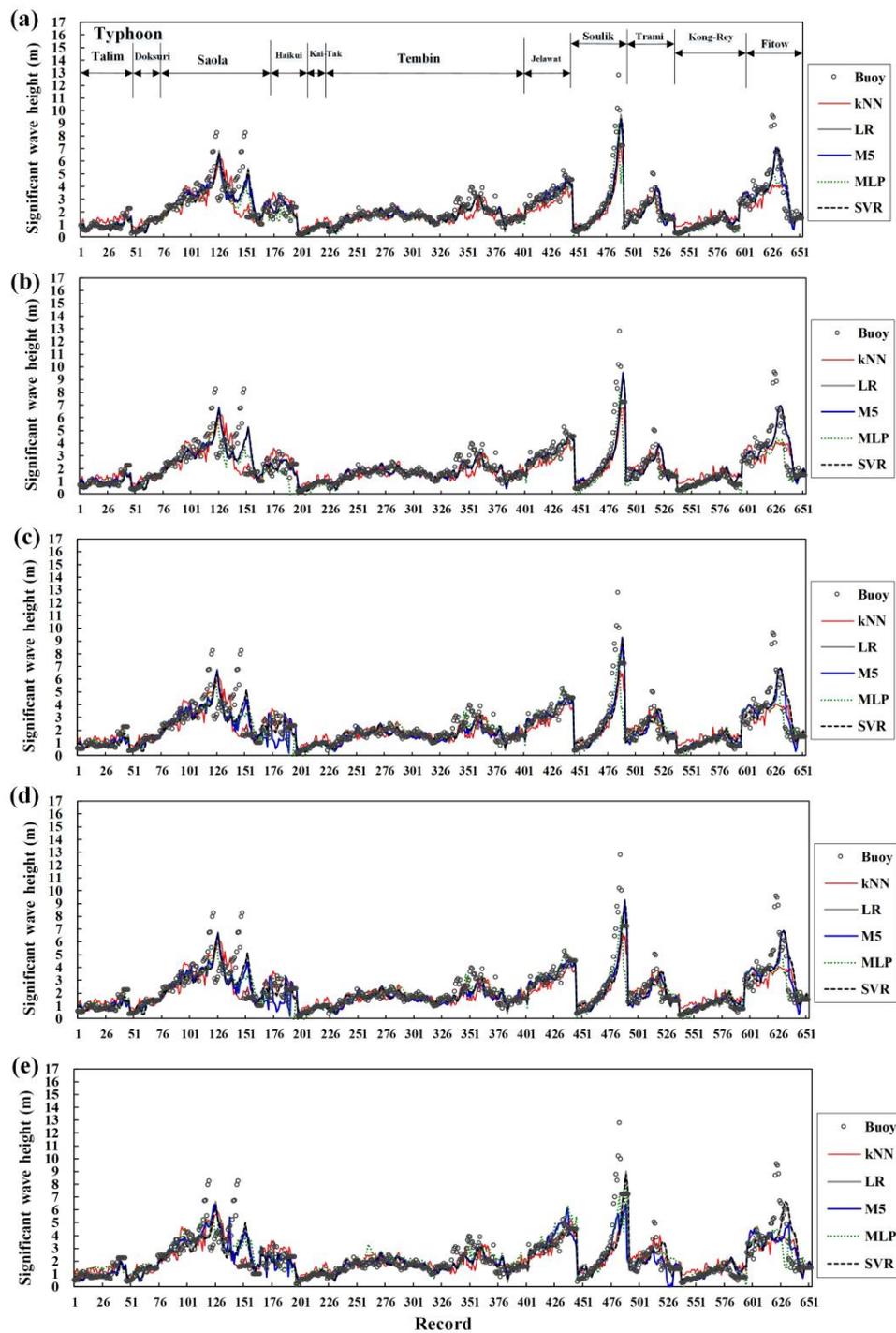


Figure 10. Predictions of wave heights using TV90 case for lead time at: (a) 2 h, (b) 3 h, (c) 4 h, (d) 5 h, and (e) 6 h.

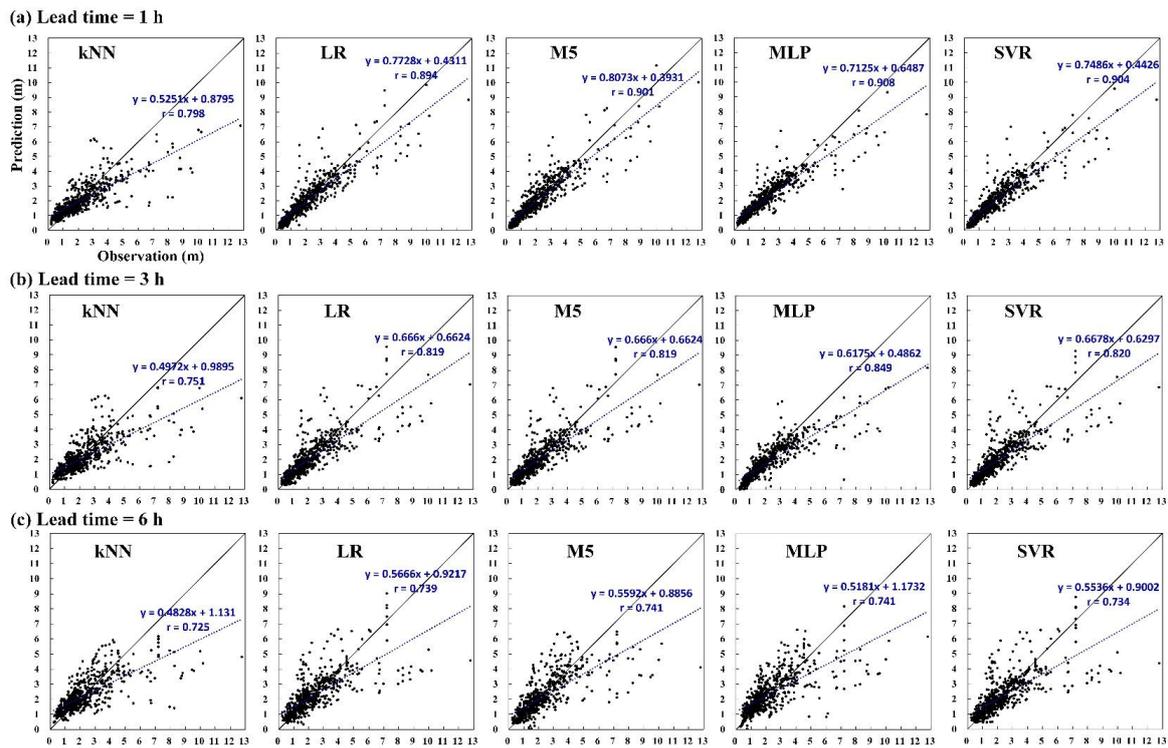


Figure 11. Scattering plots of wave heights using TV90 case for lead time at: (a) 1 h, (b) 3 h, and (c) 6 h.

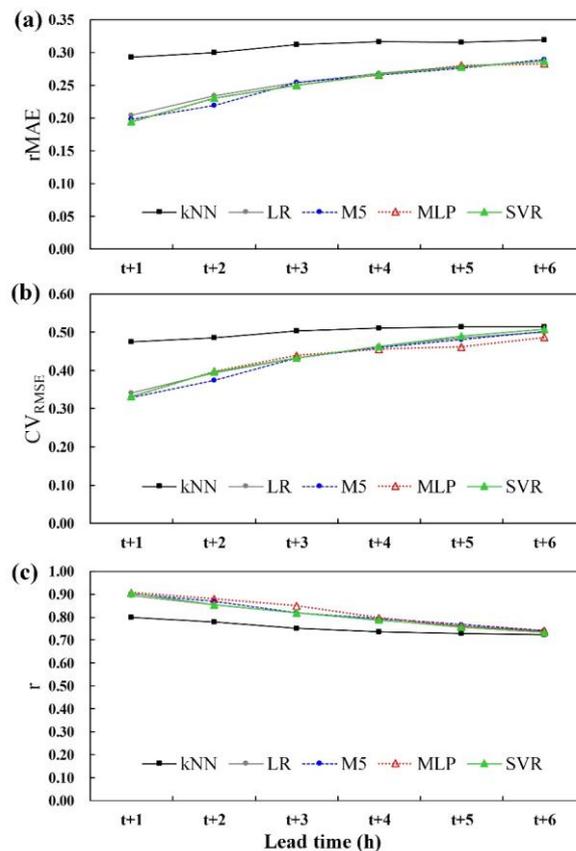


Figure 12. Performance measures at lead time from 1 h to 6 h using TV90 case: (a) rMAE, (b) CV_{RMSE}, and (c) r.

To determine the prediction ability, the researcher calculated the average performance measure for 1–6 h predictions (Table 4). The results showed that the MLP achieved more favorable performance (lower rMAE and CV_{RMSE} , and higher r values) than did other models.

Table 4. Average performance measures of relative error terms (rMAE and CV_{RMSE}) and correlation for 1–6 h predictions.

Performance	kNN	LR	M5	MLP	SVR
rMAE	0.309	0.254	0.252	0.251	0.251
CV_{RMSE}	0.500	0.436	0.432	0.429	0.437
r	0.753	0.809	0.816	0.823	0.809

5.3.2. Effect of Ranking Average and Computational Complexity

To further determine the most suitable model for different lead times, the model accuracy and complexity were estimated and then the relative weighted performance was computed. The relative weighted performance comprises two terms: the ranking performance, which was estimated for each of the models based on r , MAE and RMSE, and the computational complexity, which comprises both the model training time and the test set evaluation time [52]. For the ranking performance, the best performing model on each of these measures is assigned the rank of 1 and the worst is 0. For MAE and RMSE, the rank of the model i is calculated as [53,54]:

$$R_i = 1 - \frac{e_i - \min(e_i)}{\max(e_i) - \min(e_i)} \quad (16)$$

where e_i is the measured value. For r , the rank is calculated as

$$R_i = 1 - \frac{e_i - \max(e_i)}{\min(e_i) - \max(e_i)} \quad (17)$$

The total number of the best and worst ranking models is computed by

$$C_i = \frac{1}{\rho} \left(\frac{s_i - f_i}{n} \right) + \frac{1}{\rho} \quad (18)$$

where $\rho = 2$ is the weight shifting parameter, s_i is the total number of successes or best cases for the model i , f_i is the total number of failures or worst cases for the same model, and n is the total number of datasets.

Subsequently, the relative weighted performance can be measured by

$$Z = \alpha \cdot a_i + \beta \cdot t_i \quad (19)$$

where α and β are the weight parameters for ranking average accuracy against computational complexity. The average accuracy and computational complexity are denoted by a_i and t_i respectively. Higher Z values indicate superior metrics.

Figure 13 shows the relative weighted performance for the 1, 3, and 6 h lead forecast with respect to different β values. The Z values were calculated by assuming $\alpha = 1$ and varying β from 0 to 2. From Figure 13a, MLP obtained the highest Z values when $\beta = 0$ (without consideration of computational complexity); however, M5 obtained the highest Z values when the β value increases. Similarly, Figure 13b,c, MLP and SVR obtained the higher Z values than M5, LR, and kNN when $\beta = 0$; however, MLP and SVR might obtained lower Z values than others when considering computational complexity. This indicates that the training periods of the MLP and SVR were much longer than those of the kNN, LR, and M5, because fitting all the training data and generating a global approximation in the MLP and SVR training process is typically time consuming.

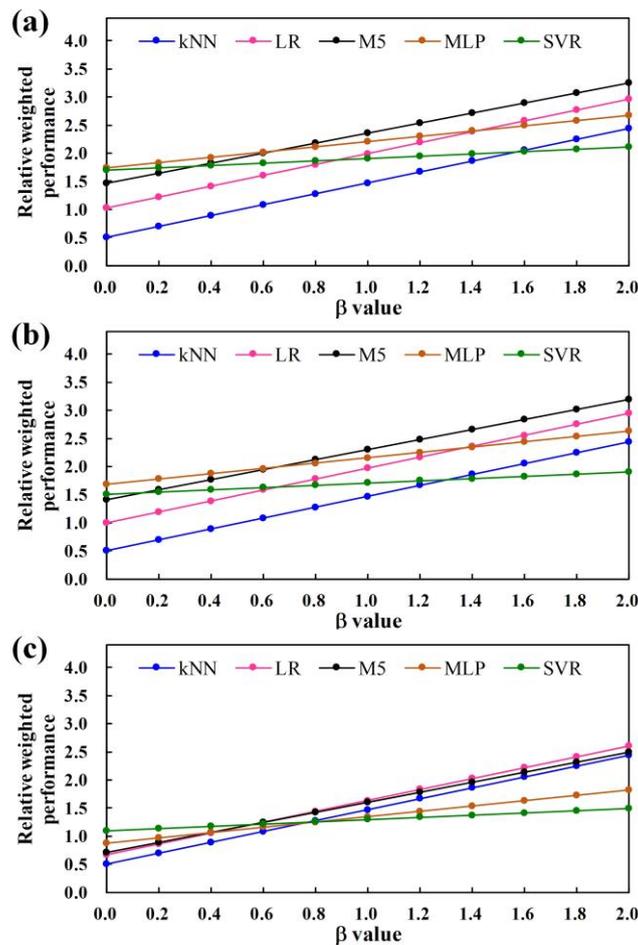


Figure 13. Relative weighted performance of various algorithms for lead time at: (a) 1 h, (b) 3 h, and (c) 6 h.

6. Conclusions

This paper reports a comparative study of PCA-derived data-driven wave-height prediction models during typhoon periods. The used data mining models are kNN, LR, M5, MLP, and SVR. The experimental site was the Longdong buoy, off the northeastern coast of Taiwan. The processing stages of model development were undertaken. First, the researcher performed PCA and used the results to reduce the potential attributes from the original data at the data preprocessing stage. On the basis of the PCA, the researcher designed four model cases: EV1, TV90, TV95, and ORI. Data concerning typhoons occurring during 2002–2011 and 2012–2013 were collected for training and testing, respectively. After calibrating the parameters of various model cases by using a training set, the simulation results made by a testing set were evaluated and discussed. The researcher obtained the following findings:

First, comparisons among four PCA cases: The results show that (1) when the number of attributes decreases, computing time decreases, and the prediction error increases. Reasons for this could be that although PCA performs the processing of data dimension reduction, pattern information is partially lost. (2) MLP and SVR result in the optimal ones compared with other models when averaging the RMSE measures of the four model cases.

Second, comparisons among the classified wave heights: The results show that (1) M5 provide the superior outcomes at small wavelet level compared with other models, (2) MLP has the optimum outcomes at the large wavelet and small/moderate wave levels compared with other models, and (3) SVR provides the optimal outcomes at the long wave and high/very high wave levels of all models.

Third, evaluation of various lead times: The results show that MLP and SVR achieved more favorable relative weighted performance without consideration of computational complexity; however, MLP and SVR might obtained lower performance when computational complexity is considered.

Acknowledgments: The support for this study under Grant No. MOST106-2111-M-019-001 provided by the Ministry of Science and Technology, Taiwan is greatly appreciated. The author acknowledges data provided by the Central Weather Bureau of Taiwan.

Author Contributions: Chih-Chiang Wei devised the experimental strategy and carried out this experiment and wrote the manuscript and contributed to the revisions.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Typhoons Affecting Offshore Northeast Taiwan 2002–2013

Table A1. Typhoons affecting offshore northeast Taiwan and typhoon intensity classification.

Typhoon	Period (yy/mm/dd)	Intensity	Typhoon	Period (yy/mm/dd)	Intensity
Rammasun	2002/7/2–4	Category 2	Sepat	2007/8/17–19	Category 3
Nakri	2002/7/9–10	Tropical storm	Mitag	2007/11/26–27	Category 1
Sinlaku	2002/9/4–8	Category 1	Kalmaegi	2008/7/17–18	Category 1
Kujira	2003/4/21–24	Category 2	Fung-Wong	2008/7/27–29	Category 2
Nangka	2003/6/1–3	Tropical storm	Nuri	2008/8/19–21	Category 1
Soudelor	2003/6/16–18	Category 1	Sinlaku	2008/9/12–14	Category 3
Imbudo	2003/7/21–23	Category 2	Hagupit	2008/9/21–23	Category 2
Morakot	2003/8/2–4	Tropical storm	Jangmi	2008/9/27–29	Category 3
Vamco	2003/8/19–20	Tropical storm	Linha	2009/6/19–21	Tropical storm
Krovanh	2003/8/22–23	Category 1	Molave	2009/7/16–18	Tropical storm
Dujuan	2003/8/31–9/2	Category 2	Morakot	2009/8/5–10	Category 1
Melor	2003/11/2–3	Tropical storm	Parma	2009/10/3–6	Category 2
Conson	2004/6/7–9	Category 1	Namtheun	2010/8/30–31	Tropical storm
Mindulle	2004/6/30–7/2	Category 2	Lionrock	2010/8/31–9/2	Tropical storm
Kompasu	2004/7/14–15	Tropical storm	Meranti	2010/9/9–10	Tropical storm
Rananim	2004/8/10–13	Category 1	Fanapi	2010/9/19–20	Category 2
Aere	2004/8/24–26	Category 1	Megi	2010/10/21–23	Category 2
Haima	2004/9/12–13	Tropical storm	Aere	2011/5/9–10	Tropical storm
Meari	2004/9/26–27	Category 1	Songda	2011/5/27–28	Category 3
Nanmadol	2004/12/3–4	Category 1	Meari	2011/6/23–25	Tropical storm
Haitang	2005/7/17–19	Category 3	Muifa	2011/8/4–6	Category 2
Matsa	2005/8/3–6	Category 1	Nanmadol	2011/8/28–31	Category 3
Sanvu	2005/8/11–13	Tropical storm	Talim	2012/6/19–21	Tropical storm
Talim	2005/8/31–9/1	Category 3	Doksuri	2012/6/28–29	Tropical storm
Damrey	2005/9/21–23	Tropical storm	Saola	2012/7/31–8/3	Category 1
Longwang	2005/10/1–3	Category 3	Haikui	2012/8/6–7	Category 1
Chancho	2006/5/16–18	Category 2	Kai-Tak	2012/8/14–15	Tropical storm
Ewiniar	2006/7/7–9	Category 2	Tembin	2012/8/21–28	Category 2
Bilis	2006/7/12–15	Tropical storm	Jelawat	2012/9/27–28	Category 3
Kaemi	2006/7/24–26	Category 1	Soulik	2013/7/11–13	Category 3
Bopha	2006/8/8–9	Tropical storm	Trami	2013/8/20–22	Tropical storm
Saomai	2006/8/9–10	Category 2	Kong-Rey	2013/8/27–29	Tropical storm
Shanshan	2006/9/14–16	Category 2	Fitow	2013/10/4–7	Category 1
Pabuk	2007/8/7–9	Tropical storm	-	-	-

References

1. Wei, C.C. Improvement of typhoon precipitation forecast efficiency by coupling SSM/I microwave data with climatologic characteristics and precipitation. *Weather Forecast.* **2013**, *28*, 614–630. [[CrossRef](#)]
2. Chen, W.B.; Liu, W.C.; Hsu, M.H. Predicting typhoon-induced storm surge tide with a two-dimensional hydrodynamic model and artificial neural network model. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 3799–3809. [[CrossRef](#)]
3. Tsai, H.C.; Lee, T.H. Maximum covariance analysis of typhoon surface wind and rainfall relationships in Taiwan. *J. Appl. Meteorol. Climatol.* **2009**, *48*, 997–1016. [[CrossRef](#)]
4. Wu, C.C.; Kuo, Y.H. Typhoons affecting Taiwan: Current understanding and future challenges. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 67–80. [[CrossRef](#)]
5. Chang, C.P.; Yeh, T.C.; Chen, J.M. Effects of terrain on the surface structure of typhoons over Taiwan. *Mon. Weather Rev.* **1993**, *121*, 734–752. [[CrossRef](#)]
6. Tsay, C.Y. Orography effects on the structure of typhoons: Analyses of two typhoons crossing Taiwan. *Terr. Atmos. Ocean. Sci.* **1994**, *5*, 313–333. [[CrossRef](#)]
7. Wei, C.C. Surface wind nowcasting in Penghu Islands based on classified typhoon tracks and the effects of the Central Mountain Range of Taiwan. *Weather Forecast.* **2014**, *29*, 1425–1450. [[CrossRef](#)]
8. Wu, C.C.; Yen, T.H.; Kuo, Y.H.; Wang, W. Rainfall simulation associated with Typhoon Herb (1996) near Taiwan. Part I: The topographic effect. *Weather Forecast.* **2002**, *17*, 1001–1015. [[CrossRef](#)]
9. Chiu, F.C.; Huang, W.Y.; Tiao, W.C. The spatial and temporal characteristics of the wave energy resources around Taiwan. *Renew. Energy* **2013**, *52*, 218–221. [[CrossRef](#)]
10. Feng, X.; Zheng, J.; Yan, Y. Wave spectra assimilation in typhoon wave modeling for the East China Sea. *Coast. Eng.* **2012**, *69*, 29–41. [[CrossRef](#)]
11. Zamani, A.; Solomatine, D.; Azimian, A.; Heemink, A. Learning from data for wind-wave forecasting. *Ocean Eng.* **2008**, *35*, 953–962. [[CrossRef](#)]
12. Aminzadeh-Gohari, A.; Bahai, H.; Bazargan, H. Simulation of significant wave height by neural networks and its application to extreme wave analysis. *J. Atmos. Ocean. Technol.* **2009**, *26*, 778–792. [[CrossRef](#)]
13. Chang, H.K.; Liou, J.C.; Liu, S.J.; Liaw, S.R. Simulated wave-driven ANN model for typhoon waves. *Adv. Eng. Softw.* **2011**, *42*, 25–34. [[CrossRef](#)]
14. Clarizia, M.P.; Ruf, C.S. Bayesian wind speed estimation conditioned on significant wave height for GNSS-R ocean observations. *J. Atmos. Ocean. Technol.* **2017**, *34*, 1193–1202. [[CrossRef](#)]
15. Makarynskyy, O.; Pires-Silva, A.A.; Makarynska, D.; Ventura-Soares, C. Artificial neural networks in wave predictions at the west coast of Portugal. *Comput. Geosci.* **2005**, *31*, 415–424. [[CrossRef](#)]
16. Londhe, S.N. Soft computing approach for real-time estimation of missing wave heights. *Ocean Eng.* **2008**, *35*, 1080–1089. [[CrossRef](#)]
17. Etemad-Shahidi, A.; Mahjoobi, J. Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Eng.* **2009**, *36*, 1175–1181. [[CrossRef](#)]
18. Tsai, J.C.; Tsai, C.H. Wave measurements by pressure transducers using artificial neural networks. *Ocean Eng.* **2009**, *36*, 1149–1157. [[CrossRef](#)]
19. Deka, P.C.; Prahlada, R. Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Eng.* **2012**, *43*, 32–42. [[CrossRef](#)]
20. Asma, S.; Sezer, A.; Ozdemir, O. MLR and ANN models of significant wave height on the west coast of India. *Comput. Geosci.* **2012**, *49*, 231–237. [[CrossRef](#)]
21. Monahan, A.H. Nonlinear principal component analysis by neural networks: Theory and applications to the Lorenz system. *J. Clim.* **2000**, *13*, 821–835. [[CrossRef](#)]
22. Lu, W.Z.; Wang, W.J.; Wang, X.K.; Yan, S.H.; Lam, J.C. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environ. Res.* **2004**, *96*, 79–87. [[CrossRef](#)] [[PubMed](#)]
23. Ture, M.; Kurt, I.; Akturk, Z. Comparison of dimension reduction methods using patient satisfaction data. *Expert Syst. Appl.* **2007**, *32*, 422–426. [[CrossRef](#)]
24. Wei, C.C. RBF neural networks combined with principal component analysis applied to quantitative precipitation forecast for a reservoir watershed during typhoon periods. *J. Hydrometeorol.* **2012**, *13*, 722–734. [[CrossRef](#)]

25. Toussaint, G.T. Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining. *Int. J. Comput. Geom. Appl.* **2005**, *15*, 101–150. [[CrossRef](#)]
26. Quinlan, R.J. Learning with Continuous Classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; pp. 343–348.
27. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
28. Liu, P.C.; Tsai, C.H.; Chen, H.S. On the growth of ocean waves. *Ocean Eng.* **2007**, *34*, 1472–1480. [[CrossRef](#)]
29. Liu, R.X.; Kuang, J.; Gong, Q.; Hou, X.L. Principal component regression analysis with SPSS. *Comput. Methods Programs Biomed.* **2003**, *71*, 141–147. [[CrossRef](#)]
30. Saegusa, R.; Sakano, H.; Hashimoto, S. Nonlinear principal component analysis to preserve the order of principal components. *Neurocomputing* **2004**, *61*, 57–70. [[CrossRef](#)]
31. Pearson, K. On lines and planes of closest fit to systems of points in spaces. *Philos. Mag.* **1901**, *2*, 559–572. [[CrossRef](#)]
32. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [[CrossRef](#)]
33. Richards, J.E. Recovering dipole sources from scalp-recorded event-related-potentials using component analysis: Principal component analysis and independent component analysis. *Int. J. Psychophysiol.* **2004**, *54*, 201–220. [[CrossRef](#)] [[PubMed](#)]
34. Chae, S.S.; Warde, W.D. Effect of using principal coordinates and principal components on retrieval of clusters. *Comput. Stat. Data Anal.* **2006**, *50*, 1407–1417. [[CrossRef](#)]
35. Liaw, Y.C.; Wu, C.M.; Leou, M.L. Fast k-nearest neighbors search using modified principal axis search tree. *Digit. Signal Process.* **2010**, *20*, 1494–1501. [[CrossRef](#)]
36. Bremner, D.; Demaine, E.; Erickson, J.; Iacono, J.; Langerman, S.; Morin, P.; Toussaint, G. Output-sensitive algorithms for computing nearest-neighbor decision boundaries. *Discret. Comput. Geom.* **2005**, *33*, 593–604. [[CrossRef](#)]
37. Atkeson, C.G.; Moore, A.W.; Schaal, S. Locally weighted learning. *Artif. Intell. Rev.* **1997**, *11*, 11–73. [[CrossRef](#)]
38. Finley, A.O.; McRoberts, R.E. Efficient k-nearest neighbor searches for multi-source forest attribute mapping. *Remote Sens. Environ.* **2008**, *112*, 2203–2211. [[CrossRef](#)]
39. Blaker, H. Minimax estimation in linear regression under restrictions. *J. Stat. Plan. Inference* **2000**, *90*, 35–55. [[CrossRef](#)]
40. Wei, C.C. Comparing single- and two-segment statistical models with a conceptual rainfall-runoff model for river streamflow prediction during typhoons. *Environ. Model. Softw.* **2016**, *85*, 112–128. [[CrossRef](#)]
41. Holmes, G.; Hall, M.; Frank, E. Generating Rule Sets from Model Trees. In Proceedings of the Twelfth Australian Joint Conference on Artificial Intelligence, Sydney, Australia, 6–10 December 1999; pp. 1–12.
42. Wang, Y.; Witten, I.H. Induction of Model Trees for Predicting Continuous Classes. In *Poster Papers of the 9th European Conference on Machine Learning*; Springer: Prague, Czech Republic, 1997.
43. Javadi, A.A. Estimation of air losses in compressed air tunneling using neural network. *Tunn. Undergr. Space Technol.* **2006**, *21*, 9–20. [[CrossRef](#)]
44. Lin, C.; Wei, C.C.; Tsai, C.C. Prediction of the influential operational compost parameters for monitoring composting process. *Environ. Eng. Sci.* **2016**, *33*, 494–506. [[CrossRef](#)]
45. Patra, J.C.; van den Bos, A. Auto-calibration and -compensation of a capacitive pressure sensor using multilayer perceptrons. *ISA Trans.* **2000**, *39*, 175–190. [[CrossRef](#)]
46. Iphar, M. ANN and ANFIS performance prediction models for hydraulic impact hammers. *Tunn. Undergr. Space Technol.* **2012**, *27*, 23–29. [[CrossRef](#)]
47. Wei, C.C.; Hsu, N.S.; Huang, C.L. Rainfall-runoff prediction using dynamic typhoon information and surface weather characteristics considering monsoon effects. *Water Resour. Manag.* **2016**, *30*, 877–895. [[CrossRef](#)]
48. Kaiser, H.F. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151. [[CrossRef](#)]
49. Trenn, S. Multilayer perceptrons: Approximation order and necessary number of hidden units. *IEEE Trans. Neural Netw.* **2008**, *19*, 836–844. [[CrossRef](#)] [[PubMed](#)]
50. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **2001**, *106*, 7183–7192. [[CrossRef](#)]
51. Hand, D.; Mannila, P.; Smyth, P. *Principles of Data Mining*; MIT Press: Cambridge, MA, USA, 2001.

52. Shafiullah, G.M. Hybrid renewable energy integration (HREI) system for subtropical climate in Central Queensland, Australia. *Renew. Energy* **2016**, *96*, 1034–1053. [[CrossRef](#)]
53. Ali, A.B.M.S.; Smith, K.A. On learning algorithm selection for classification. *Appl. Soft Comput.* **2006**, *6*, 119–138. [[CrossRef](#)]
54. Shafiullah, G.M.; Amanullah, M.T.O.; Ali, A.B.M.S.; Jarvis, D.; Wolfs, P. Prospects of renewable energy—A feasibility study in the Australian context. *Renew. Energy* **2012**, *39*, 183–197. [[CrossRef](#)]



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).