Journal of
Water & Climate Change

IWA PUBLISHING

Journal of Water and Climate Change Vol 13 No 6, 2443 doi: 10.2166/wcc.2022.042

Downscaling model for rainfall based on the influence of typhoon under climate change

Shiu-Shin Lin [a,*], Yan-Li Hu[b] and Kai-Yang Zhu[a]

[a] Chung Yuan Christian University, Taoyuan, Taiwan
[b] Yilan City Office, Yilan, Taiwan
*Corresponding author. E-mail: linxx@cycu.edu.tw

S-SL, 0000-0002-9440-7690

ABSTRACT

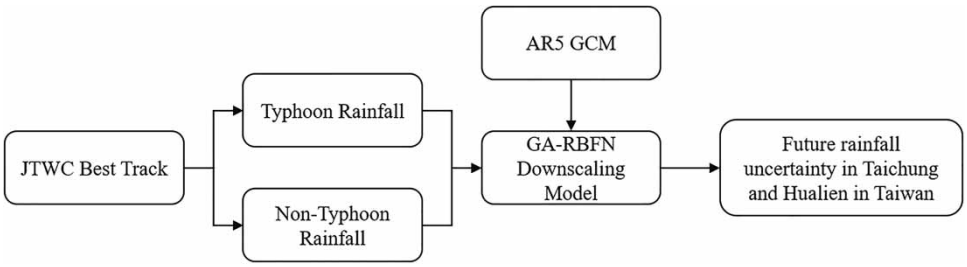
The effects of typhoon rainfall cannot be easily investigated to downscale the GCMs (general circulation models) data, because typhoons are short-term highly nonlinear processes. In this study, to explore the impact on the typhoon rainfall for downscaling models, the monthly rainfalls are divided into typhoon rainfall and non-typhoon rainfall. The GA-RBFN downscaling models, integrating genetic algorithm to optimize a radial basis function neural network, were established to enquire into the future rainfall under the effect of typhoons in Taichung and Hualien, Taiwan. The GCM data in this study included MRI-CGCM3 and CSIRO-Mk3.0 from the 5th Assessment Report of the Intergovernmental Panel on Climate Change. The historical monthly rainfall of Taichung and Hualien stations are collected from the Central Weather Bureau, Taiwan, and the best track of typhoons from the Joint Typhoon Warning Center. Principal component analysis (PCA) and the stepwise regression procedure (SRP) were adopted, respectively, to select input variables from the GCM data. The future rainfall trends and uncertainties are evaluated by the best GA-RBFN model, which is selected with the highest performance by applying holistic information criteria. The simulated results show that the model with variables transformed by PCA performs well in forecasting non-typhoon rainfalls, while the model with variables transformed by SRP performs well in forecasting monthly total rainfalls. According to the three classifications of future rainfalls in wet and dry seasons, the mid- and long-term rainfall amount are mainly low to normal for Taichung and normal to high for Hualien. For the long-term rainfall assessment and the probability analysis on exceeding historical rainfalls, the long-term rainfall variability of the two stations is higher than the mid-term rainfall variability. It demonstrates that the future rainfall estimation has high uncertainty under the force of typhoons. Finally, the rainfall variability in the dry period is higher than that in the wet season, and the average probability of the future dry season rainfalls exceeding historical rainfall is high.

Key words: artificial neural network, downscale, extremes, principal component analysis, stepwise regression procedure

HIGHLIGHTS

- The study estimated the influence of rainfall of typhoons for developing the statistical downscaling model using AR5 GCMs.
- Both at same northern latitudes point rainfall were compared.
- The methodology of integration of machine learning approach and the influence of typhoon to investigate the future trend of rainfall is new in this area.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Taiwan is an island located in the western Pacific in the northern semisphere. Typhoons make landfall in Taiwan every year and are accompanied by abundant rainfall. Rainfalls are profoundly influenced by tropical cyclones, typhoons, and their accompanying effects. Under the influence of typhoons, the monthly rainfall varies greatly and is different in wet and dry seasons. The temporal and spatial distribution of rainfalls is uneven, and the noise caused by typhoon rainfalls may influence the models to forecast future rainfalls.

In the Coupled Model Intercomparison Project Phase 5 (Fifth Assessment Report, AR5), the Working Group I of the Intergovernmental Panel on Climate Change (IPCC) developed the General Circulation Model (GCM) and forecasted the future climate. However, as the global-scale GCM data grid is too rough to analyze regional climate, it is necessary to downscale the data in the global-scale grid to show regional climate characteristics (Manor & Berkovic 2015). As a commonly used downscaling method, statistical downscaling is based on the statistical characteristics of past atmospheric, hydrological data and the future data from the GCM model. It has the ability to present ensemble results for different emission scenarios and global climate models.

In recent years, machine learning (ML), a manifold learning method, has been widely applied in various fields and led to good research results. In atmospheric hydrology, in order to improve forecast efficiency, ML has gradually replaced traditional statistical methods. They can process the nonlinearity of typhoons, reduce model errors, and better improve prediction (Shi *et al.* 2012, 2013; Kan *et al.* 2017; Huang *et al.* 2018; Hu *et al.* 2021). Common ML methods, such as support vector machine (SVM), self-organizing map (SOM), radial basis function neural network (RBFN), and deep learning, are mainly used for downscaling and hydrologic studies (Ahmed *et al.* 2015; Chuang *et al.* 2016; Zhang *et al.* 2016; Kundu *et al.* 2017; Li *et al.* 2020; Sulaiman *et al.* 2020; Kumar *et al.* 2021). The RBFN has the advantages of easy parameter setting and fast establishment. Therefore, this study chose the RBFN as the kernel of the statistical downscaling model. For models based on ML, inputs (also known as 'variables') are usually selected by multivariate analysis and statistical testing to reduce the dimensions of inputs, data noise, model complexity, and model efficiency. For example, Chuang *et al.* (2016) used the stepwise regression procedure (SRP) and principal components analysis (PCA) to select the input variables of RBFN to reduce the model complexity, respectively. Kan *et al.* (2017) extracted the data characteristics by nonlinear principal component analysis (NLPCA) and could explain more than 90% of the variability of the original data when three principal components were employed. Hu *et al.* (2021) used NLPCA to select the input variables of RBFN and build the statistical downscaling model. To reduce the complexity of RBFN, SRP, and PCA were used in this study to select or retain the combination of variables with better explanation. It is improved based on large amounts of historical and future climate data, as provided by the learning algorithm.

There have been many studies on the analysis of regional rainfall trends under climate change, but few studies have been published on the effects of typhoon rainfall on downscaling models. In order to explore the future rainfall trends of Taichung and Hualien, as well as the influence of typhoon rainfall on the establishment of downscaling models, this study established two downscaling models for scenarios with and without typhoon rainfall, respectively. AR5 GCM data and rainfall data were collected from Taichung and Hualien stations. Based on genetic algorithm (GA) optimization and RBFN downscaling models for scenarios with and without typhoon, rainfall were developed to assess the trends and uncertainties of future monthly rainfall influenced by typhoons.

2. METHODOLOGY

2.1. Radial basis function neural network

RBFN is a feedforward neural network first proposed by Hardy (1971). RBFN can be used to establish nonlinear models. As shown in Figure 1, the RBFN architecture consists of an input layer, a hidden layer, and an output layer.

RBFN substitutes the input vector (X) into the radial basis function (Φ) of the hidden layer to calculate the maximum error between the estimated and the target value. Regarding the radial basis function (Φ) of the hidden layer, the Gaussian function is taken as an example, as shown in Equation (1), where c_j is the center point of neuron j and σ is the standard deviation of the input data.

$$\Phi(\|x - c_j\|) = e^{-\frac{\|x - c_j\|}{2\sigma^2}} \tag{1}$$

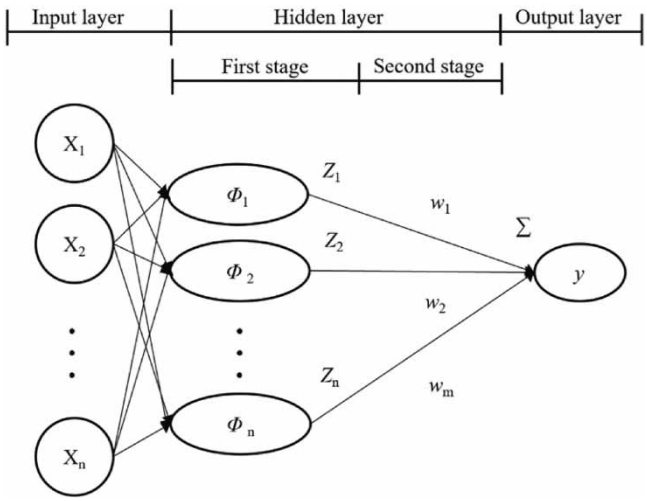


Figure 1 | Radial basis function neural network architecture.

The output of each neuron in the hidden layer is shown in Equation (2), where $\Phi(\|x_i - c_j\|)$ is the hidden layer neuron of variable i to center point j; w_j is the weighted value from neuron j in the hidden layer to the output layer; $\|x - c_j\|$ is the Euclidean distance between x and c_j; and m is the total number of neurons.

$$z_j(x) = \Phi(\|x - c_j\|), \quad j = 1,\ 2,\ \ldots,\ m \tag{2}$$

The output is the summation of weighted value of the hidden layer, as shown in Equation (3), where y is the network output value; w_j is the weight value from neuron j in the hidden layer to the output layer; and z_j is the output value of neuron j in the hidden layer.

$$y = \sum_{j=0}^{m} w_j \cdot z_j(x) \tag{3}$$

In the hidden layer, the radial basis function (Φ) is used to calculate the distance (Z) between the input vector and the center point of the hidden layer neuron. The number of neurons in the hidden layer should be determined for output. The output layer is a weighted linear combination of the output value (y) of the hidden layer. The radial basis function (Φ) of some hidden layer neurons is used to build the approximate mapping relationship between input and output variables.

Two stage algorithms, first and second stages, are employed to determine the hidden layer structure of RBFN. The center point c_j of the hidden layer is selected in the first stage, and parameters of RBFN are adjusted in the second stage. Within an allowable error range, RBFN successively selects the center points of input data that minimize the error of the simulation value, determines the number of neurons in the hidden layer, and establishes the network architecture. The network has proper complexity and initial parameters with the fewest center points. The learning algorithms to adjust parameters in the second stage can be divided into two categories: those that only modify the output weights, such as the least mean square algorithm (LMS), and those that modify the parameters of the entire network, including the gradient descent method and Newton's method, which are used to output the weights, neuron center points, and standard deviation. The proper initial positioning of the center point of the network can improve the convergence speed and training stability.

2.2. Genetic algorithm

The GA became popular due to the works of Holland (1992). With the concept from the theory of evolution by natural selection, this algorithm mimics the evolution of biological communities, meaning that it is an optimization algorithm to choose excellent offspring and eliminate poor parents by the trial and error method. The advantage of a GA is that it adopts the multi-point simultaneous search, which is faster than the single point search at finding the optimum solution. In addition, the

optimum solution may still be found in the multidimensional function (Reshma *et al.* 2018; Debbarma *et al.* 2019; Salih *et al.* 2020; Jahandideh-Tehrani *et al.* 2021).

2.3. Methods to process the input variables of models

This study applied two methods to process the input variables: PCA to reduce data dimension and retain principal components with large eigenvalues; SRP to select the variable combination with the maximum explanatory power. In order to ensure consistent variable scales, data were normalized before processing.

PCA can reduce data dimensions, and simplify and linearly transform data. The principal components and weights are calculated with the covariance matrix, where low-order principal components with large eigenvalues are retained, data dimensions are reduced, and maximum features are retained (Wold *et al.* 1987).

SRP was adopted to select the variable combination with the maximum explanatory of source data. SRP selects input variables from the variables that pass the *F*-test forward and deletes variables that fail the *F*-test backward, which is conducted according to the regression analysis of the output and input variables and the linear regression test of the variable combinations. After the combinations of the input variables pass the *F*-test of SRP, another *t*-test is performed on the causal relationship between the independent and dependent variables (Chuang *et al.* 2016; Li *et al.* 2020).

3. DATA AND ANALYSIS

3.1. Station data

This study selected the Taichung Weather Station (120 °41′02.8″E, 24 °08′44.5″N, 84.0 M, Taichung City, Taiwan) and Hualien Weather Station (121 °36′47.9″E, 23 °58′30.4″N, 16.1 M, Hualien City, Taiwan), which are located on either side of the Central Mountain Range of Taiwan and at a similar latitude, as shown in Figure 2. The hourly rainfall data from the two stations from January 1950 to December 2005 were employed in this study.

3.2. GCM data

This study collected historical and future GCM data from RCP4.5 and RCP8.5 scenarios. The historical scenario was from January 1950 to December 2005; the future scenario was from January 2006 to December 2100, and the data were recorded monthly. The GCM model is CSIRO-Mk3.0 (called CSMK3 in the study), developed by Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence, Australia, with grid resolution of 192(km) × 96(km); MRI-CGCM3 (called MRMC in this study), developed by the Meteorological Research Institute, Japan, with grid resolution of 320(km) × 160(km), respectively.

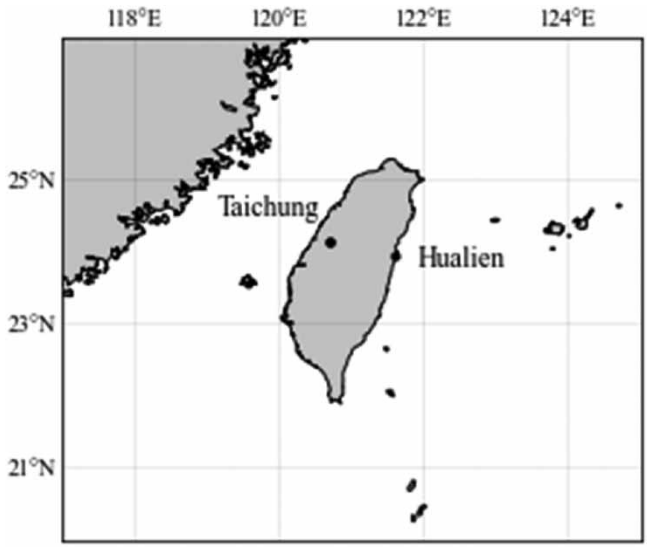


Figure 2 | Locations of Taichung and Hualien stations.

In addition to including the same as the AR4 GCM variables, some new GCM stratified variables in AR5, such as atmosphere, land, ocean, aerosol, and greenhouse gas, were selected as the GCM variables in this study. After missing values were removed, the GCM variables employed in the study are shown in Table 1.

3.3. Non-typhoon rainfall calculation

The research refers to Chang *et al.* (1993), Lee *et al.* (2006), Li (2006), and Huang (2014) for the estimation of typhoon rainfalls. The typhoon grids considered in the above studies are expressed in the same order. Chang *et al.* (1993) studied the influence of typhoons over Taiwan's Central Mountain Range on typhoon structure. Lee *et al.* (2006) established a typhoon rainfall climate model by considering topographic uplift and rainfall rate, which varies with radius. Li (2006) separated typhoon rainfall from main rainfall in Taiwan and only analyzed non-typhoon rainfall. Huang (2014) established three grids with different densities, each occupying 0.25°, 0.5°, and 1° latitude and longitude, to forecast rainfalls according to the radius of Category 7 typhoons, as defined by the Central Weather Bureau, Taiwan. Based on the typhoon grid established by Huang (2014), this study estimated typhoon rainfalls with each grid occupying 0.5° in latitude and in longitude.

According to Table 2 and Figure 3, typhoons that continue to move westward and circulate may still contribute to the rainfall in Taiwan. The typhoon grid of Li (2006) extends northward, hence, the average annual typhoon rainfall estimated by Taichung station increases by about 30 mm, and the typhoons that continue to move northward contribute less to the rainfalls. Finally, because its estimated typhoon rainfalls were the second highest, and its results were similar to other studies, the typhoon grid of Li (2006) was adopted in this study.

The non-typhoon rainfall in this study was estimated according to Li (2006), the typhoon path, and the longitude and latitude of the center point, which were the best track data of tropical cyclones from 1950 to 2005 from the Joint Typhoon Warning Center (JTWC) in the United States. The steps to estimate non-typhoon rainfall are as follows: (1) integrate the best track data of tropical cyclones from JTWC; (2) read the longitude and latitude of the typhoon center according to the date when a typhoon enters the grid; (3) read the total rainfall (hourly observed rainfall from the station) according to

Table 1 | AR5 GCM variables used in this study

Model	Realm	Long name	Output variable name	Units
CSMK3 & MRMC	Atmosphere	Total Cloud Fraction	clt	%
		Evaporation	evspsbl	$kg \cdot m^{-2} \cdot s^{-1}$
		Specific Humidity	hurs	%
		Near-surface Specific Humidity	huss	1
		Precipitation	pr	$kg \cdot m^{-2} \cdot s^{-1}$
		Surface Air Pressure	ps	Pa
		Sea Level Pressure	psl	Pa
		Surface Downwelling Longwave Radiation	rlds	$W \cdot m^{-2}$
		Near-Surface Air Temperature	tas	K
		Eastward Near-Surface Wind	uas	$m \cdot s^{-1}$
		Northward Near-Surface Wind	vas	$m \cdot s^{-1}$
	Aerosol-Related	Total Emission Rate of SO_2	emiso2	$kg \cdot m^{-2} \cdot s^{-1}$
		Total Emission Rate of SO_4	emiso4	$kg \cdot m^{-2} \cdot s^{-1}$
		Surface Diffuse Downwelling Shortwave Radiation	rsdsdiff	$W \cdot m^{-2}$
	Ocean	Sea Surface Temperature	tos	K
MRMC	Ocean	Water Evaporation Flux Where Ice Free Ocean over Sea	evs	$kg \cdot m^{-2} \cdot s^{-1}$
		Northward Ocean Heat Transport	hfbasin	W
		Temperature Flux due to Evaporation Expressed as Heat Flux Out of the Sea Water	hfevapds	$W \cdot m^{-2}$
		Temperature Flux due to Rainfall Expressed as Heat Flux into Sea Water	hfrainds	$W \cdot m^{-2}$
		Ocean Heat *X* Transport	hfx	W
		Ocean Heat *Y* Transport	hfy	W
		Rainfall Flux where Ice Free Ocean over Sea	pr(Ocean)	$kg \cdot m^{-2} \cdot s^{-1}$
		Surface Net Downward Longwave Radiation	rlds (Ocean)	$W \cdot m^{-2}$
		Net Downward Shortwave Radiation at Sea Water Surface	rsntds	$W \cdot m^{-2}$

Table 2 | Grid ranges in studies on typhoons and average annual typhoon rainfalls at stations

Relevant studies	Longitude and latitude of typhoon grid	The number of times the typhoon center enters the typhoon grid	Average annual typhoon rainfall in Taichung (mm)	Average annual typhoon rainfall in Hualien (mm)
Chang *et al.* (1993)	117–125E, 19–27N	2,568	251.5	588.6
Lee *et al.* (2006)	118–126E, 19–27N	2,681	246.7	530.2
Li (2006)	117–125E, 19–28N	2,700	272.8	593.1
Huang (2014)	0.5° grid: 116–126E, 18–30N	4,282	309.7	710.8

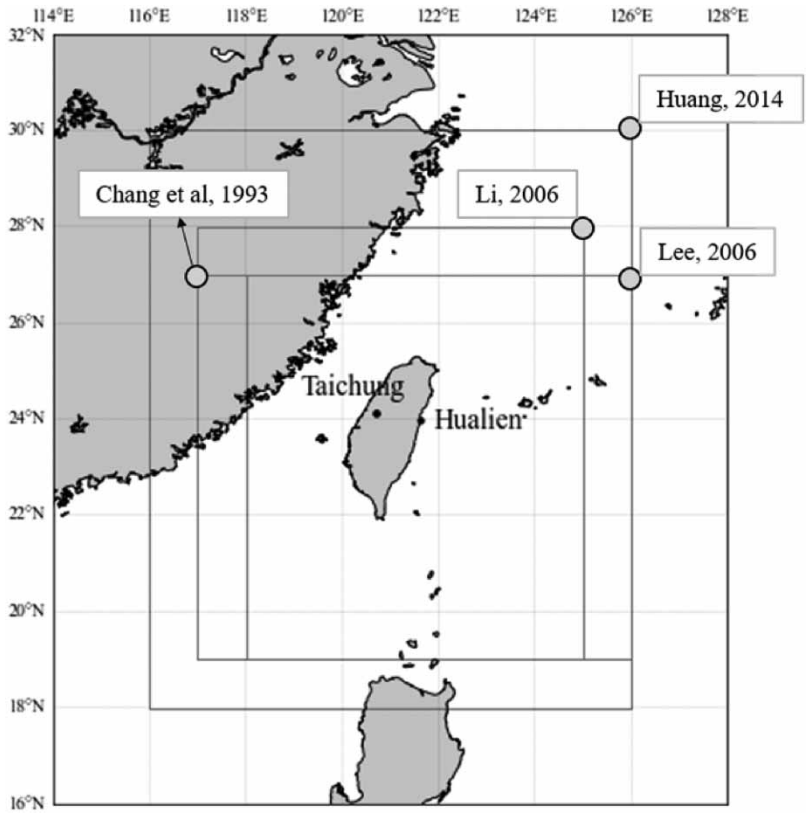


Figure 3 | Grid ranges in studies on typhoons.

the date when a typhoon enters the grid, the accumulation of the total rainfall between the typhoons located in that grid called typhoon rainfall, including tropical cyclones. The monthly typhoon rainfall is the typhoon rainfall accumulated for that month; (4) the monthly non-typhoon rainfall is the rainfall resulting from the monthly total rainfall disposing of monthly typhoon rainfall. According to Figure 4, the difference between monthly total rainfall and non-typhoon rainfall at Hualien station is larger than that at Taichung station.

4. ESTABLISHMENT OF GA-RBFN DOWNSCALING MODELS

The process to establish models in this study is shown in Figure 5. Firstly, the GCM variables related to AR5 atmospheric hydrology, the optimal path data of tropical cyclones and typhoons from JTWC in the United States, and hourly rainfall amounts of the Central Weather Bureau (CWB) were collected. Secondly, the data were preprocessed, and the estimated typhoon rainfall was separated from the monthly total rainfall according to the rainfall measured by the stations of the CWB and the optimal path

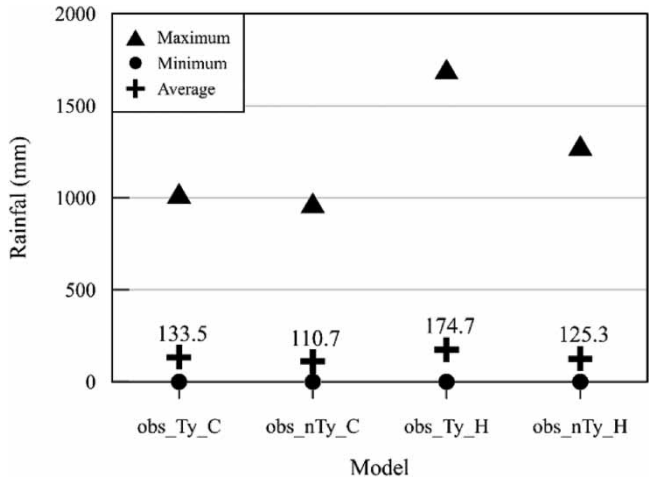


Figure 4 | Maximum, minimum, and average values of monthly total rainfalls and non-typhoon rainfalls from 1950 to 2005 at Hualien and Taichung stations.

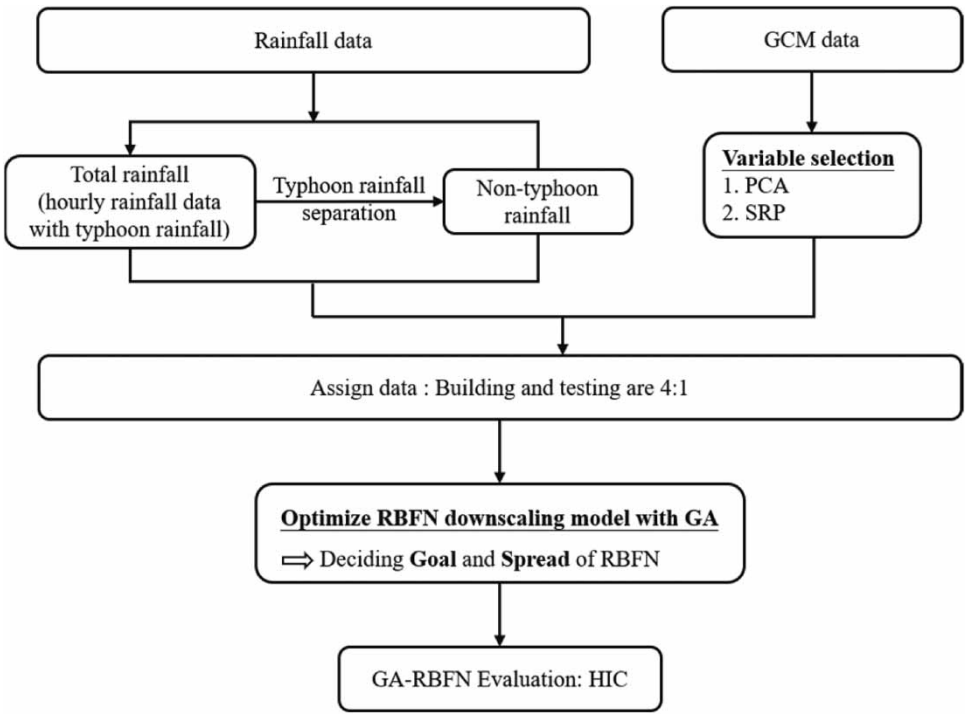


Figure 5 | Research process flowchart.

data from JTWC, in order to calculate the non-typhoon rainfall. GCM variables were transformed by PCA and SRP to reduce RBFN model complexity and improve efficiency. The models were then constructed by optimizing the hyperparameters of RBFN with GA. Finally, according to the holistic information criteria (HIC), the better models were assessed and selected, and an integrated model was obtained after ensembling these models to analyze the future rainfall changes at the two stations.

The GA-RBFN model in this study was named by four codes, namely, the GCM name, the variable selection method, the target output rainfall, and the station. In terms of GCM name, CSMK3 and MRMC present the GCM data of different research units, such as CSIRO-Mk3.0 and MRI-CGCM3. In the variable selection method, pP is the input variable of the PCA, and pS is the input variable selected by SRP. The target output rainfall consists of monthly rainfall excluding typhoon rainfall (non-typhoon rainfall), and monthly rainfall including typhoon rainfall (monthly total rainfall), which are hereinafter referred to as nTy and Ty. C and H represent Taichung and Hualien stations, respectively.

4.1. Mode input variable selection

The input variables in this study include the data transformed by PCA and the variable combinations selected by SRP. The number of principal components selected for each GCM was based on Li *et al.* (2020), and the principal components with the cumulative proportion of variation that explained more than 80% were selected. As shown in Figure 6, CSMK3 and MRMC for Taichung and Hualien selected four principal components as the input variables.

After the combinations of input variables passed the *F*-test of SRP, another *t*-test was performed on the causal relationship between the independent and dependent variables. In order to make the number and conditions of GA-RBFN models of non-typhoon rainfall or monthly total rainfall similar to those of the principal components selected by PCA, the model input variables were variable combinations that combined two target output rainfall tests. The predictive variables of the GCM data of CSMK3 and MRMC at the stations, as selected by SRP, are shown in Tables 3 and 4, respectively. According to the two tables, CSMK3 has fewer predictive variables than MRMC, which is possibly due to the poor resolution of CSMK3.

4.2. Model parameter optimization

In this study, the rainfall data from January 1950 to December 2005 were partitioned in proportions of 4:1 for training and testing. In order to determine the parameter ranges of the models, the parameters of GOAL and SPREAD of RBFN were

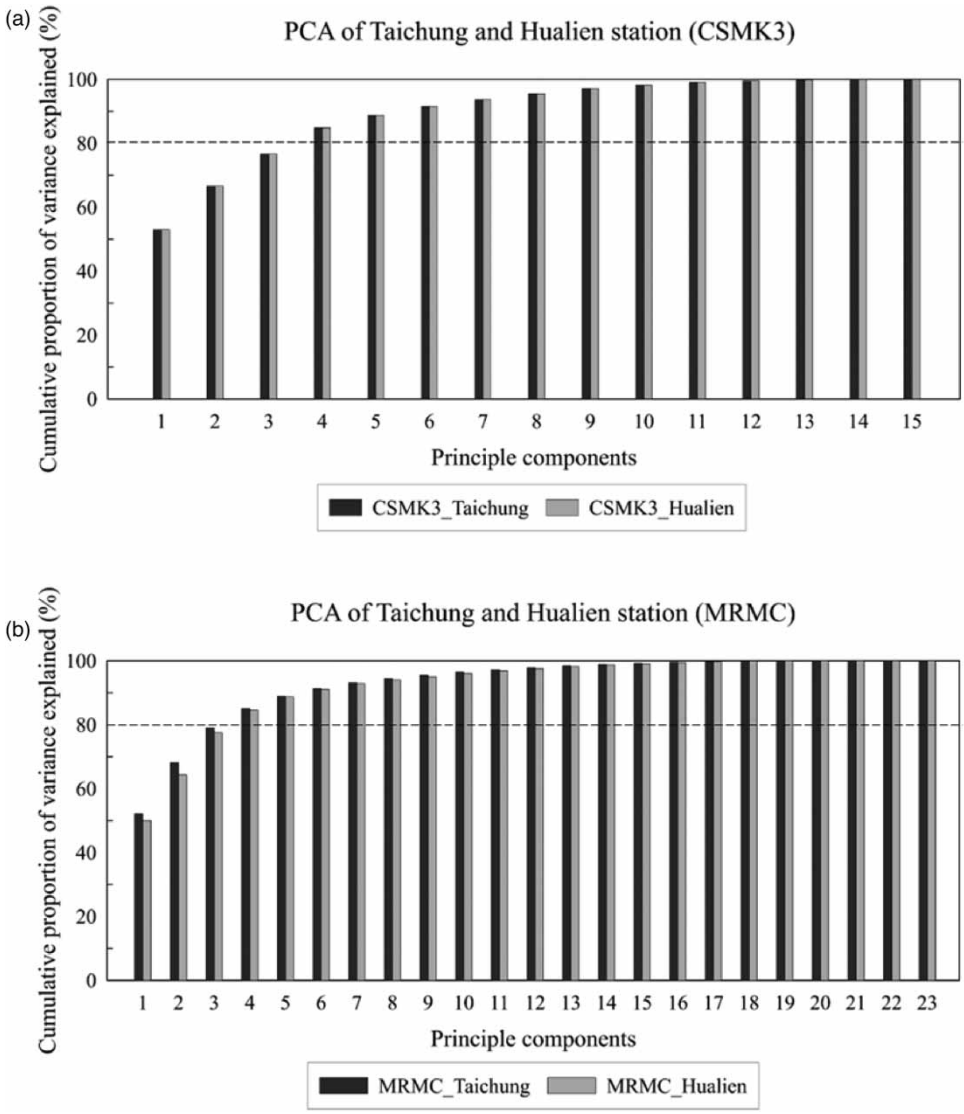


Figure 6 | Principal component analysis on CSMK3 and MRMC of Taichung and Hualien stations.

Table 3 | Variable statistics of CSMK3 at Taichung and Hualien stations for stepwise regression

		F-test		*t*-test	
Target output rainfall	**Variables selected**	***F*-value**	***P*-value**	***F*-value**	***P*-value**
(a) CSMK3 Taichung station					
Non-typhoon rainfall	1. emiso4	4.0052	0.0188	−2.2664	0.0238
	2. pr	4.0052	0.0188	−2.0291	0.0429
Total rainfall	None of the variables was selected.	–	–	–	–
(b) CSMK3 Hualien station					
Non-typhoon rainfall	1. uas	4.358	0.0373	−2.0876	0.0373
Total rainfall	1. evspsbl	5.6255	0.0181	2.3718	0.0181

Table 4 | Variable statistics of MRMC at Taichung and Hualien stations for stepwise regression

		F-test		*t*-test	
Target output rainfall	**Variables selected**	***F*-value**	***P*-value**	***F*-value**	***P*-value**
(a) MRMC Taichung station					
Non-typhoon rainfall	1. hfbasin	65.7654	1.989E−53	−2.5787	0.0102
	2. huss	65.7654	1.989E−53	6.8185	<0.0001
	3. rlds	65.7654	1.989E−53	−3.9404	0.0001
	4. rlds (Ocean)	65.7654	1.989E−53	3.0635	0.0023
	5. tos	65.7654	1.989E−53	−6.3267	<0.0001
Total rainfall	1. hfbasin	86.9508	7.74E−57	−2.1663	0.0307
	2. ps	86.9508	7.74E−57	−3.4533	0.0006
	3. rlds (Ocean)	86.9508	7.74E−57	3.8551	0.0001
	4. tas	86.9508	7.74E−57	3.1075	0.002
(b) MRMC Hualien station					
Non-typhoon rainfall	1. hfbasin	12.5706	5.95E−08	4.3783	<0.0001
	2. hurs	12.5706	5.95E−08	−3.7619	0.0002
	3. pr	12.5706	5.95E−08	2.3479	0.0192
Total rainfall	1. hurs	38.2017	4.04E−28	−7.3903	<0.0001
	2. rlds	38.2017	4.04E−28	11.1523	<0.0001
	3. rlds (Ocean)	38.2017	4.04E−28	−2.4078	0.0164
	4. vas	38.2017	4.04E−28	−5.9414	<0.0001

tested. The number of RBFN neurons was assumed to be between the number of training data (538 data were collected from January 1950 to October 1994) and the minimum number of neurons was assumed to be 2. In order to avoid overfitting or underfitting the models, this study selected GA to optimize the RBFN parameters according to the range of changes in the number of neurons as mentioned above, and established neural networks with sufficient neurons for modeling, rather than the maximum or minimum number of neurons.

The parameter range of GOAL was tested by the mean variance of normalized rainfall values. First, the range of GOAL was tested from 0, 2, 4, 6, and 8 multiplied by 10 to the power of negative three and negative two times the mean variance to 1, 2, 4, 6, and 8 multiplied by 10 to the power of negative one, zero, one, and two times the mean variance. In the process of GOAL testing, SPREAD was first fixed as the default value (radius of influence of a unit center point), and the range of GOAL fell between 538 and 2 after the neurons were tested.

To test the parameter range of SPREAD, SPREAD was set between 0.1 and 5.0, and 0.1 was added to SPREAD for each test (Chuang *et al.* 2016; Li *et al.* 2020). Finally, the GOAL and SPREAD of neuron tests in which neurons fell between 538 and 2 were used as the upper and lower bounds of the GA optimized RBFN.

Regarding GA optimization setting, in addition to the range of GOAL and SPREAD of the RBFN parameters, the iteration conditions were generation size and population size. The termination condition was calculated according to the minimum mean absolute error between the predicted value and the observed value of the model. A GA-RBFN model can test three (generation size and population size were 10, 50, and 100) iteration conditions, thus, GA-RBFN calculated each iteration condition three times to obtain GCM, namely, the rainfall forecast model, including nine submodels. The GA-RBFN models were selected to simulate future rainfall.

The parameter ranges of all models are shown in Table 5. Compared with CSMK3, MRMC had a smaller upper limit of GOAL and a larger upper limit of SPREAD, indicating that the forecast error of MRMC is smaller than that of CSMK3.

4.3. Model efficiency assessment

This study assessed model efficiency by HIC in the test stage. HIC is an assessment indicator that integrates the coefficient of correlation (CC) and mean square error (MSE) (Hsieh 2007), as shown in Equations (4)–(6), where x_i is the observed value; $\bar{x}$ is the average observed value; y_i is the predicted value; $\bar{y}$ is the predicted value; and n is the number of data.

$$\mathrm{CC} = \frac{\sum_{i=1}^{n}[(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^{n}[(x_i - \bar{x})^2] \cdot \sum_{i=1}^{n}[(y_i - \bar{y})^2]}} \tag{4}$$

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \tag{5}$$

$$\mathrm{HIC} = \mathrm{MSE} - \mathrm{CC} \times \mathrm{MSE} \tag{6}$$

The GA-RBFN models used to forecast future scenarios are shown in the following table. By classifying the models according to the same GCM, target output rainfall, and stations, this study compared the efficiencies of different variable selection methods and selected the GA-RBFN model with smaller HIC to forecast future rainfall. According to Table 6, CSMK3 and MRMC employed SRP and PCA as variable selection methods to achieve high-efficiency GA-RBFN models.

Table 5 | Parameter ranges of GA-RBFN models of Taichung and Hualien stations

Model	Goal	Spread	Generation size	Population size
CSMK3_pS_nTy_C	[0.0177,0.0221]	[0.1,0.2]	50	50
CSMK3_pP_nTy_C	[0.0044,0.0221]	[0.1,0.9]	10	10
CSMK3_pS_Ty_C	[0.0211, 0.0264]	[0.1,0.2]	100	100
CSMK3_pP_Ty_C	[0.0053,0.0263]	[0.1,0.9]	10	10
MRMC_pS_nTy_C	[0.00004,0.0177]	[0.1,2.8]	100	100
MRMC_pP_nTy_C	[0.0001,0.0162]	[0.1,1.0]	10	10
MRMC_pS_Ty_C	[0.00005,0.01583]	[0.1,2.8]	10	10
MRMC_pP_Ty_C	[0.00005,0.0189]	[0.1,0.9]	100	100
CSMK3_pS_nTy_H	[0.0084,0.0105]	[0.1,0.3]	50	50
CSMK3_pP_nTy_H	[0.0011,0.0105]	[0.1,0.9]	10	10
CSMK3_pS_Ty_H	[0.0093,0.0116]	[0.1,0.4]	100	100
CSMK3_pP_Ty_H	[0.0012,0.0116]	[0.1,0.9]	10	10
MRMC_pS_nTy_H	[0.00002,0.01053]	[0.1,2.0]	10	10
MRMC_pP_nTy_H	[0.00008,0.0099]	[0.1,1.0]	10	10
MRMC_pS_Ty_H	[0.00002,0.0116]	[0.1,2.1]	10	10
MRMC_pP_Ty_H	[0.00002,0.0093]	[0.1,0.9]	10	10

Table 6 | GA-RBFN model efficiency assessment

Model	CC	MSE	HIC	Assessment results
CSMK3_pS_nTy_C	0.2379	0.0141	0.0107	yes
CSMK3_pP_nTy_C	0.3199	0.0262	0.0178	no
CSMK3_pS_Ty_C	0.2161	0.0233	0.0182	yes
CSMK3_pP_Ty_C	0.3092	0.0304	0.0210	no
MRMC_pS_nTy_C	0.4535	0.0209	0.0114	no
MRMC_pP_nTy_C	0.359	0.0175	0.0112	yes
MRMC_pS_Ty_C	0.4314	0.0392	0.0223	no
MRMC_pP_Ty_C	0.3617	0.0176	0.0113	yes
CSMK3_pS_nTy_H	0.0432	0.0116	0.0111	yes
CSMK3_pP_nTy_H	0.0365	0.0258	0.0249	no
CSMK3_pS_Ty_H	0.0928	0.0114	0.0104	yes
CSMK3_pP_Ty_H	0.1797	0.0176	0.0144	no
MRMC_pS_nTy_H	0.2033	0.0128	0.0102	no
MRMC_pP_nTy_H	0.2035	0.0122	0.0097	yes
MRMC_pS_Ty_H	0.2826	0.0153	0.0110	no
MRMC_pP_Ty_H	0.2653	0.0125	0.0092	yes

5. FUTURE RAINFALL ASSESSMENT

The IPCC's AR5 proposed four emission pathways at different concentrations, namely, RCP2.6, RCP4.5, RCP6, and RCP8.5. In this study, RCP4.5 had stable radiative force from now to 2100, while the radiative force of RCP8.5, which continued to increase from now to 2100, was selected as the future climate scenario to forecast and analyze the future rainfalls of Taichung and Hualien with and without typhoons by the GA-RBFN models.

In this study, the GA-RBFN model with high integration efficiency was used as the ensemble model for rainfalls with and without typhoons in Taichung and Hualien. The model was named future scenario target output rainfall station. The mid- and long-term future scenarios were from 2041 to 2060 and from 2061 to 2080, with the codes of M and L, respectively.

5.1. Three classifications and interval analysis of future rainfalls

This study analyzed the future rainfall trend of the GA-RBFN model by three-way classification for a climatic forecast of Taiwan's CWB. The three classifications of historical rainfalls were sorted in ascending order (1950/01–2005/12), with 30 and 70% of the rainfall as the upper and lower limits of the normal range, where rainfall more than or equal to the lower limit and less than or equal to the upper limit was normal, while rainfall less than the lower limit was low, and vice versa. Based on the foregoing statement, the upper and lower limits of three classifications of historical rainfalls in Taichung and Hualien were obtained, as shown in Figure 7.

The mid- and long-term rainfalls of Taichung and Hualien in the GA-RBFN models were classified according to the upper and lower limits of the three classifications of historical rainfalls, as shown in Figure 8. The results show that the rainfall amounts of Taichung and Hualien were low or normal in the mid-term wet season and high in the dry season, while the rainfall in the long-term wet season was similar to that in the mid-term. However, while the non-typhoon rainfall of Taichung mainly changed from high to normal, the rainfall amounts in the long-term dry season at the two stations were mainly high.

The rainfall interval forecasts of Taichung and Hualien with and without the influence of typhoons are shown in Figure 9. The rainfall of Taichung mainly ranges from 0 to 100 mm in the mid-term wet season, and from 100 to 200 mm in the dry season, while the rainfall of Hualien mainly ranges from 100 to 200 mm in wet and dry seasons. According to the forecast, the rainfall in long-term wet/dry seasons is similar to that in mid-term in Taichung and Hualien, and rainfall exceeding 400 mm in Hualien in the dry season tends to increase, as compared with that in the mid-term.

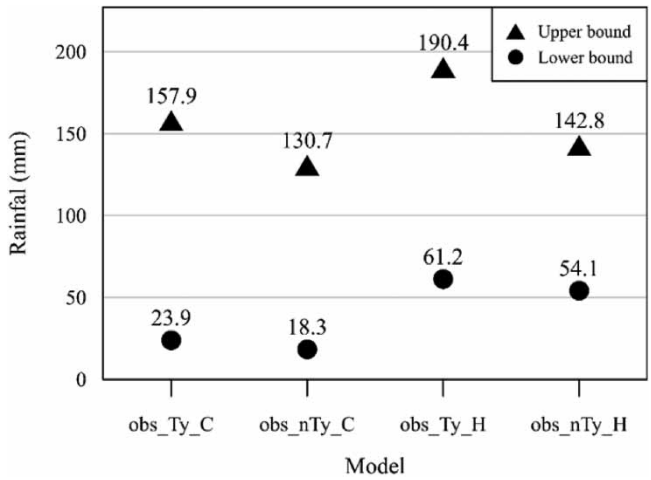


Figure 7 | Upper and lower limits of normal historical rainfall range at Taichung and Hualien stations.

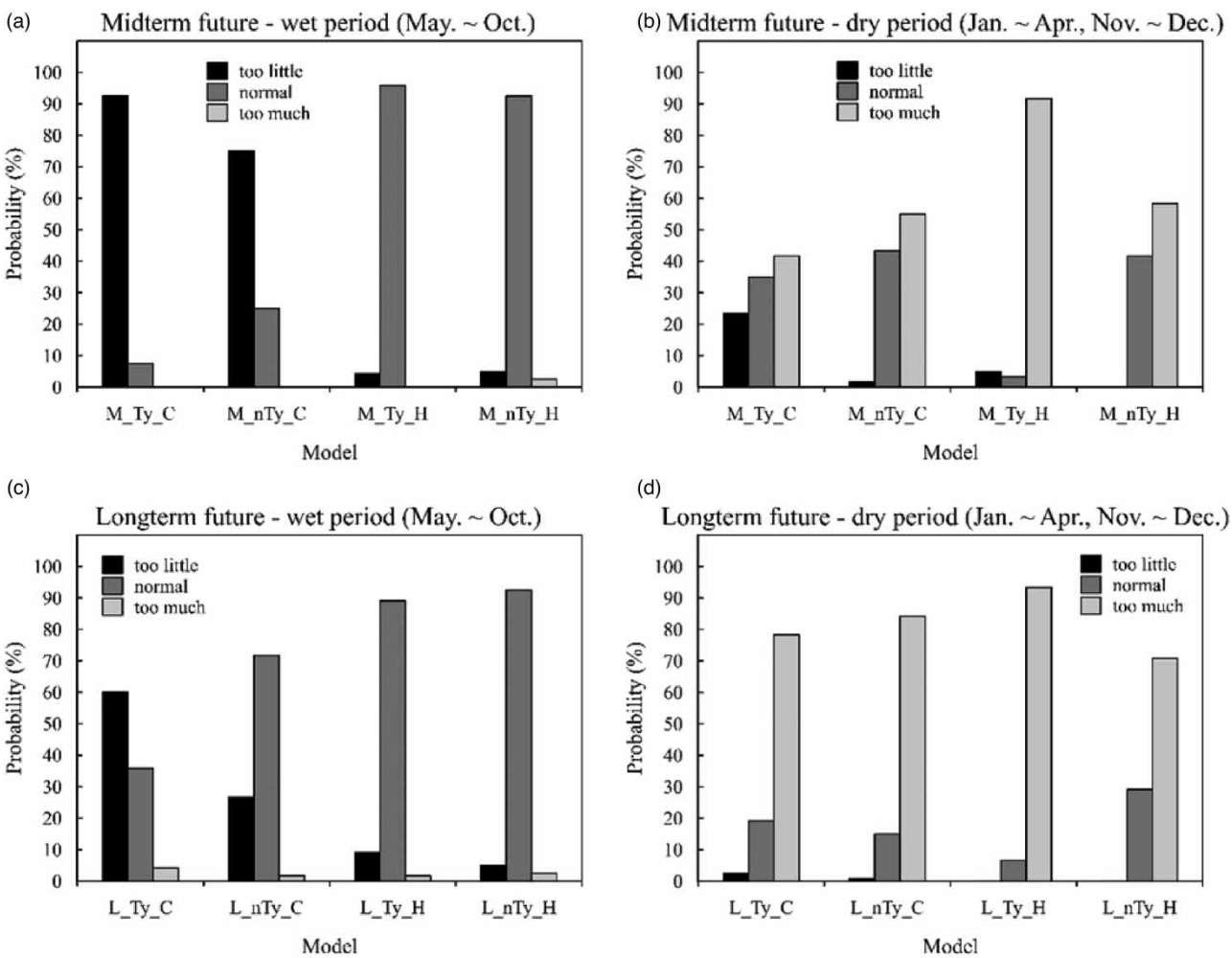


Figure 8 | Three classifications of rainfalls of Taichung and Hualien models in mid- and long-term wet/dry seasons.

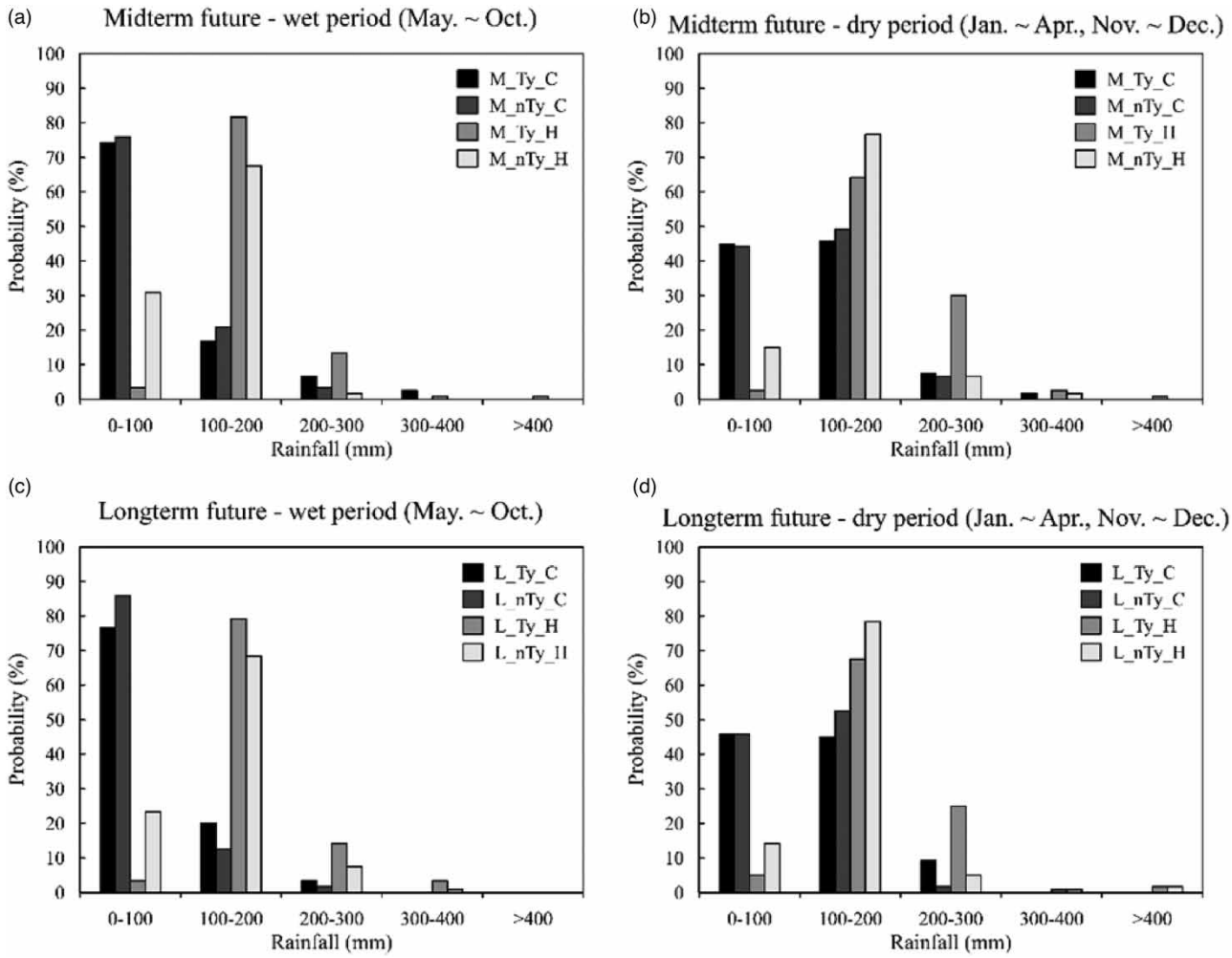


Figure 9 | Interval probabilities of rainfalls of Taichung and Hualien models in mid- and long-term wet/dry seasons.

5.2. Mid- and long-term future rainfall assessments

The future rainfall forecast for Taichung and Hualien is shown in Figure 10. According to the mid-term rainfall forecast for Taichung, in the two models, the rainfall amount is lower than the historical average in the wet season, higher than the historical average in the dry season, and the rainfall variability in the dry season is great. According to the long-term rainfall forecast for Taichung, the annual rainfall increases, as compared with the mid-term rainfall, and the annual rainfall variability is great. According to the mid-term rainfall forecast for Hualien, the rainfall variability in the dry season is greater than that in the wet season, and the model without the influence of typhoons is close to the historical rainfall trend. According to the long-term rainfall forecast for Hualien, the rainfall in March increases significantly, while the rainfall in November varies greatly. In conclusion, the mid-term rainfall variability of Taichung and Hualien in the dry season is greater than that in the wet season, while the long-term rainfall variability is greater than the mid-term rainfall variability.

Regarding the mid- and long-term rainfall forecast of the two stations, the probability of model-forecasted rainfall exceeding historical rainfall is calculated by Weber's formula, as shown in Figure 11. While the probability of rainfall amounts in mid- and long-term dry seasons exceeding historical rainfall is higher than that in the wet season in Taichung and Hualien, in Hualien, the probability of exceeding historical rainfall in the mid- and long-term without the influence of a typhoon is high in June.

6. CONCLUSION

By selecting variables through PCA and SRP as the GCM meteorological factors, this study established GA-RBFN models with and without typhoon rainfall to forecast future rainfall amounts, respectively, and draws the following conclusions:

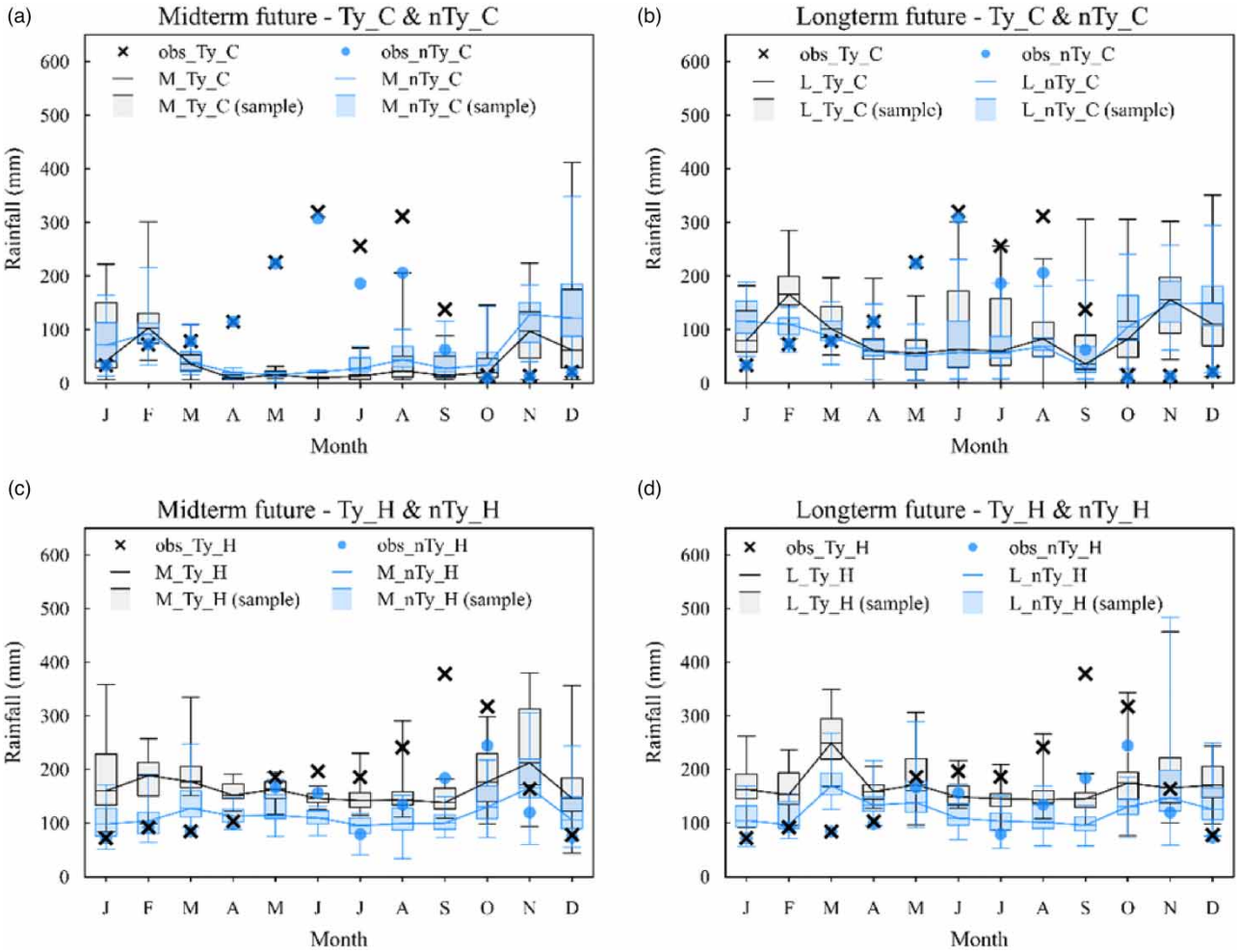


Figure 10 | Future forecast (mid- and long-term) of Taichung and Hualien stations with and without typhoon rainfall.

1. CSMK3 and MRMC of Taichung and Hualien selected four variables by PCA as the GCM meteorological factors; and MRMC selected more GCM meteorological factors by PCA than CSMK3. According to the efficiency assessment, the model with variables transformed by PCA performed well in forecasting non-typhoon rainfall, while the model with variables transformed by SRP performed well in predicting monthly total rainfall.
2. According to the three classifications of future rainfalls in wet and dry seasons, the rainfall amounts in the mid-term wet season are low and normal in Taichung and Hualien, with and without the influence of typhoons; the rainfall in the mid-term dry season is mainly normal and high in Taichung with and without the influence of typhoons, respectively. In Hualien, the rainfall is mainly high with the influence of typhoons and mainly normal and high without the influence of typhoons. In Hualien, the rainfall in the long-term wet season is similar to that in the medium term. The rainfall amounts in the long-term dry season are mainly high in Taichung and Hualien with or without typhoon rainfalls.
3. According to the future forecast rainfall distribution in wet and dry seasons, in the mid-term wet season, the rainfall of Taichung mainly ranges from 0 to 100 mm with or without the influence of typhoons, while the rainfall of Hualien mainly ranges from 100 to 200 mm with or without typhoon rainfalls. In the mid-term dry season, the rainfall of Taichung mainly ranges from 0 to 100 mm and from 100 to 200 mm with and without the influence of typhoons, respectively, and the rainfall of Hualien mainly ranges from 100 to 200 mm with or without the influence of typhoons. The rainfall in the long-term dry season is similar to that in the mid-term dry season in Taichung and Hualien, with or without the influence of typhoons. However, rainfall exceeding 400 mm in Hualien in the long-term dry season tends to increase slightly, as compared with that in the mid-term dry season.

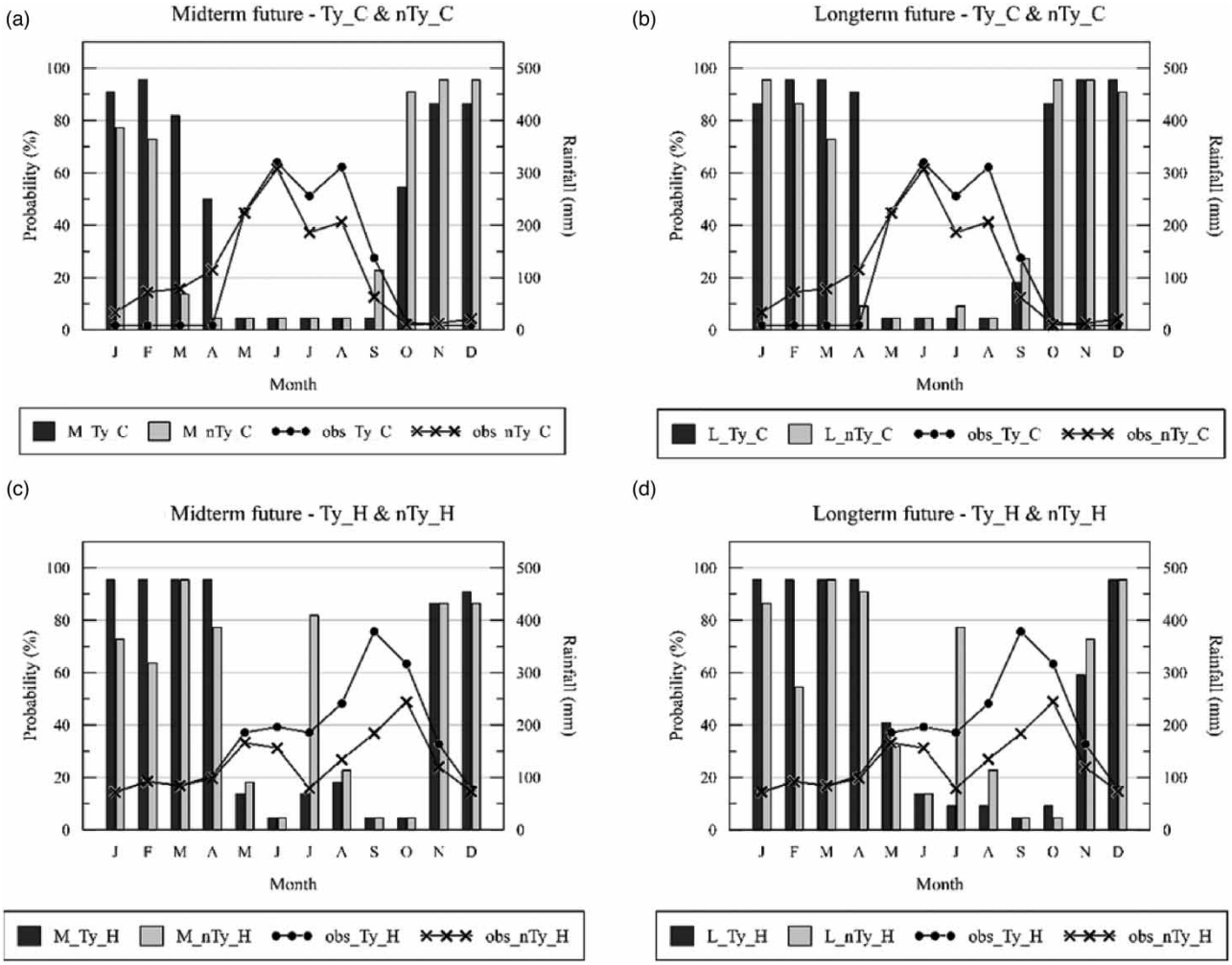


Figure 11 | Probabilities of monthly rainfall exceeding historical monthly average rainfall (mid- and long-term) of Taichung and Hualien with and without typhoon rainfall in the future.

4. According to the future rainfall assessment, the rainfall variability of models with or without the influence of typhoons in the mid-term dry season is high in Taichung and Hualien, and the monthly rainfall variability of models with or without the influence of typhoons in the long term is higher than that in the medium term in Taichung. The rainfall of models with and without the influence of typhoons in March increases significantly in Hualien.
5. The probability of the future rainfall, as forecasted by the GA-RBFN models, exceeding historical rainfall was obtained by comparison, and there was a high probability that the rainfall of Taichung and Hualien in mid- and long-term dry seasons would exceed the historical rainfall.

ACKNOWLEDGEMENTS

The support from Project No. MOST 109-2625-M-033-001 under the Ministry of Science and Technology for research projects is greatly appreciated for the completion of this study.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Ahmed, K., Shahid, S. & Harun, S. 2015 Statistical downscaling of rainfall in an arid coastal region: a radial basis function neural network approach. In: *Applied Mechanics and Materials*, Vol. 735. Trans Tech Publications Ltd, pp. 190–194. https://doi.org/10.4028/www.scientific.net/AMM.735.190.

Chang, C. P., Yeh, T. C. & Chen, J. M. 1993 Effects of terrain on the surface structure of typhoons over Taiwan. *Monthly Weather Review* **121** (3), 734–752. https://doi.org/10.1175/1520-0493(1993)121<0734:EOTOTS>2.0.CO;2.

Chuang, J. M., Lin, S. S., Kan, P. H., Li, C. Y. & Hu, Y. L. 2016 Applying bootstrap and radial basis function neural networks developing a climate change statistical downscaling model. *Taiwan Water Conservancy* **64** (4), 48–58.

Debbarma, N., Choudhury, P. & Roy, P. 2019 Identification of homogeneous rainfall regions using a genetic algorithm involving multi-criteria decision making techniques. *Water Supply* **19** (5), 1491–1499. https://doi.org/10.2166/ws.2019.018.

Hardy, R. L. 1971 Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research* **76** (8), 1905–1915. https://doi.org/10.1029/JB076i008p01905.

Holland, J. H. 1992 Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press, Cambridge, MA.

Hsieh, W. W. 2007 Nonlinear principal component analysis of noisy data. *Neural Networks* **20** (4), 434–443. https://doi.org/10.1016/j.neunet.2007.04.018.

Hu, J. N., Lin, S. S. & Zhu, K. Y. 2021 Integrating nonlinear principal component analysis and neural networks to develop a downscaling model evaluating future rainfall of Taichung and Hualien. *Journal of Taiwan Agricultural Engineering* **67** (4), 78–90. https://doi.org/10.29974/JTAE.202112_67(4).0005.

Huang, C. W. 2014 *Forecasting Spatial Distribution of 24-Hour Rainfall by Heuristic Model in Tamsui River*. Unpublished Master's Thesis, Feng Chia University.

Huang, Y., Jin, L., Zhao, H. S. & Huang, X. Y. 2018 Fuzzy neural network and LLE algorithm for forecasting precipitation in tropical cyclones: comparisons with interpolation method by ECMWF and stepwise regression method. *Natural Hazards* **91** (1), 201–220. https://doi.org/10.1007/s11069-017-3122-x.

Jahandideh-Tehrani, M., Jenkins, G. & Helfer, F. 2021 A comparison of particle swarm optimization and genetic algorithm for daily rainfall-runoff modelling: a case study for Southeast Queensland, Australia. *Optimization and Engineering* **22** (1), 29–50. https://doi.org/10.1007/s11081-020-09538-3.

Kan, P. H., Lin, S. S., Hu, Y. L. & Sung, J. H. 2017 Integrating nonlinear principal component analysis and GA-fuzzy model to develop a typhoon rainfall forecasting model. *Taiwan Water Conservancy* **65** (4), 42–53.

Kumar, B., Chattopadhyay, R., Singh, M., Chaudhari, N., Kodari, K. & Barve, A. 2021 Deep learning-based downscaling of summer monsoon rainfall data over Indian region. *Theoretical and Applied Climatology* **143** (3), 1145–1156. https://doi.org/10.1007/s00704-020-03489-6.

Kundu, S., Khare, D. & Mondal, A. 2017 Future changes in rainfall, temperature and reference evapotranspiration in the central India by least square support vector machine. *Geoscience Frontiers* **8** (3), 583–596. https://doi.org/10.1016/j.gsf.2016.06.002.

Lee, C. S., Huang, L. R., Shen, H. S. & Wang, S. T. 2006 A climatology model for forecasting typhoon rainfall in Taiwan. *Natural Hazards* **37** (1), 87–105. https://doi.org/10.1007/s11069-005-4658-8.

Li, Y. C. 2006 *The Analysis of Climatological Characteristics of Non-Typhoon Rainfall in Taiwan. Unpublished Master's Thesis*, National Taiwan University. https://doi.org/10.6342/NTU.2006.00413.

Li, C. Y., Lin, S. S., Chuang, C. M. & Hu, Y. L. 2020 Assessing future rainfall uncertainties of climate change in Taiwan with a bootstrapped neural network-based downscaling model. *Water and Environment Journal* **34** (1), 77–92. https://doi.org/10.1111/wej.12443.

Manor, A. & Berkovic, S. 2015 Bayesian inference aided analog downscaling for near-surface winds in complex terrain. *Atmospheric Research* **164**, 27–36.

Reshma, T., Venkata Reddy, K., Pratap, D. & Agilan, V. 2018 Parameters optimization using Fuzzy rule based multi-objective genetic algorithm for an event based rainfall-runoff model. *Water Resources Management* **32** (4), 1501–1516. https://doi.org/10.1007/s11269-017-1884-2.

Salih, S. Q., Sharafati, A., Ebtehaj, I., Sanikhani, H., Siddique, R., Deo, R. C., Bonakdari, H., Shahid, S. & Yaseen, Z. M. 2020 Integrative stochastic model standardization with genetic algorithm for rainfall pattern forecasting in tropical and semi-arid environments. *Hydrological Sciences Journal* **65** (7), 1145–1157. https://doi.org/10.1080/02626667.2020.1734813.

Shi, X., Huang, X., Jin, L. & Huang, Y. 2012 An neural network ensemble approach based on PSO algorithm and LLE for typhoon intensity. In: *2012 Fifth International Joint Conference on Computational Sciences and Optimization*. IEEE, pp. 877–880. https://doi.org/10.1109/CSO.2012.204.

Shi, X. M., Jin, L. & Huang, X. Y. 2013 A new neural network ensemble forecast method based on KPCA for typhoon intensity. *Journal of the Meteorological Sciences* **2013** (2), 184–189.

Sulaiman, N. A., Shaharudin, S. M., Zainuddin, N. H. & Najib, S. A. 2020 Improving support vector machine rainfall classification accuracy based on Kernel parameters optimization for statistical downscaling approach. *International Journal* **9** (1.4). https://doi.org/10.30Pl534/ijatcse/2020/9191.42020.

Wold, S., Esbensen, K. & Geladi, P. 1987 Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2** (1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

Zhang, H., Singh, V. P., Wang, B. & Yu, Y. 2016 CEREF: A hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system. *Journal of Hydrology* **540**, 246–256. https://doi.org/10.1016/j.jhydrol.2016.06.029.

First received 29 January 2022; accepted in revised form 5 May 2022. Available online 18 May 2022