

Incorporating large-scale atmospheric variables in long-term seasonal rainfall forecasting using artificial neural networks: an application to the Ping Basin in Thailand

M. S. Babel, T. A. J. G. Sirisena and N. Singhrattna

ABSTRACT

Understanding long-term seasonal or annual or inter-annual rainfall variability and its relationship with large-scale atmospheric variables (LSAVs) is important for water resource planning and management. In this study, rainfall forecasting models using the artificial neural network technique were developed to forecast seasonal rainfall in May–June–July (MJJ), August–September–October (ASO), November–December–January (NDJ), and February–March–April (FMA) and to determine the effects of climate change on seasonal rainfall. LSAVs, temperature, pressure, wind, precipitable water, and relative humidity at different lead times were identified as the significant predictors. To determine the impacts of climate change the predictors obtained from two general circulation models, CSIRO Mk3.6 and MPI-ESM-MR, were used with quantile mapping bias correction. Our results show that the models with the best performance for FMA and MJJ seasons are able to forecast rainfall one month in advance for these seasons and the best models for ASO and NDJ seasons are able to do so two months in advance. Under the RCP4.5 scenario, a decreasing trend of MJJ rainfall and an increasing trend of ASO rainfall can be observed from 2011 to 2040. For the dry season, while NDJ rainfall decreases, FMA rainfall increases for the same period of time.

Key words | artificial neural network, large-scale atmospheric variables, representative concentration pathways, seasonal rainfall forecasting

M. S. Babel (corresponding author)
T. A. J. G. Sirisena
 Water Engineering and Management,
 Asian Institute of Technology,
 P.O. Box 4,
 Klong Luang,
 Pathumthani 12120,
 Thailand
 E-mail: msbabel@ait.asia; msbabel@gmail.com

N. Singhrattna
 Department of Public Works and Town & Country
 Planning,
 Royal Thai Government,
 218/1, Rama 6 Road, Samsennai,
 Phrayathai,
 Bangkok 10400,
 Thailand

INTRODUCTION

Rainfall is one of the most important components of the hydrological cycle since rainfall is a major source of water on earth. Rainfall varies due to interactions among the physical processes of the atmosphere–ocean–cryosphere nexus that operate on a wide range of spatial and temporal scales. This complex physics of formation and forecast is yet to be understood in its entirety (Ramirez *et al.* 2005).

The methods used for rainfall forecasting depend on the time scale, the spatial scale, and the particular application of the study. This is due to the nonlinear time series of rainfall. Many studies have been conducted to forecast rainfall using different methods such as numerical weather prediction models (Ganguly & Bras 2003; Davolio *et al.* 2008; Diomede *et al.* 2008), statistical models (Singhrattna *et al.*

2005b; Munot & Kumar 2007; Tuleya *et al.* 2007; Li & Zeng 2008; Nayagam *et al.* 2008; Zaw & Naing 2008; Singhrattna & Babel (2011)) and soft computing based methods like artificial neural network (ANN), support vector regression and fuzzy logic models (Luchetta & Manetti 2003; Guhathakurta 2006; Ghiassi & Nangoy 2009; Srivastava *et al.* 2010; Afshin *et al.* 2011; Gadgay *et al.* 2011; Shukla *et al.* 2011; Kashid & Maity 2012; Donate *et al.* 2013). The applicability of ANN models for rainfall forecasting is well documented in the literature. The analysis of dynamic phenomena such as rainfall by intelligent systems like ANN has given better results than traditional methods such as regression, as proven in recent studies (Ghosh & Mujumdar 2006; Chattopadhyay *et al.* 2010; Srivastava

doi: 10.2166/nh.2016.212

et al. 2010; Wu *et al.* 2010; El-Shafie *et al.* 2011, 2012; Golabi *et al.* 2013; Wu & Chau 2013).

However, the capability of a forecasting model is limited, due to an incomplete understanding of trends of long- and short-term rainfall as well as of the effects of external factors on rainfall. The influence of such external factors as wind velocity, sea level pressure (SLP), temperature, aerosol concentration, land use change, and the El Niño Southern Oscillation (ENSO) on local rainfall has been studied in recent times using different prediction techniques (Krishnan *et al.* 2000; Yaremchuk & Qu 2004; Ramirez *et al.* 2005; Singhrattna *et al.* 2005a, 2012; Krishnamurthy & Shakla 2007; Chansaengkrachang *et al.* 2011; Singhrattna & Babel (2011); Bridhikitti 2012; Mekanik & Imteaz 2012). The combined effects of these variables on rainfall variability are more important than the individual effects of these variables on the same.

The inherent natural variability and anthropogenic forcing create uncertain climatic conditions over the world. The fluctuations in climatic occurrences vary from inter-annual-to-decadal time scales to as long as 50 years (Deser *et al.* 2012). To interpret such a climate system, climate model projection is a big challenge due to the different sources of uncertainties (Shepherd 2014). However, in recent years, the ability of general circulation models (GCMs) to predict climatic conditions at the continental and global levels has improved. Also, GCMs can predict conditions under different possible Special Report on Emission Scenarios. In 2008, a new approach called the Coupled Model Inter-comparison Project Phase 5 (CMIP5) was constituted, in order to integrate experiment sets that can simulate climate change and climate variability (Taylor *et al.* 2012). In CMIP5, the Representative Concentration Pathways (RCPs) are based on selected scenarios from modeling work on integrated assessment modeling, climate modeling, and the analysis of impacts. Four RCPs were defined according to the total radiative forcing in the 21st century. RCP8.5 is a radiative forcing scenario that increases to 8.5 W m^{-2} towards the end of the 21st century. Two intermediate scenarios, RCP6 and RCP4.5, are the balancing pathways, under which radiative forcing does not go beyond the peak level by the 21st century. Another scenario, RCP2.6, is the lowest peak-and-decay scenario. GCMs depict climate using a three-dimensional grid over the globe, and their resolution is coarser than the resolution

required in impact assessment (Salathe 2003). It cannot be properly modeled by any smaller scale physical process. Thus, their known properties must be averaged over the larger scale in a technique called parameterization which contributes to the uncertainty in GCM-based simulations of future climatic conditions (Trtanj & Houston 2014). Due to such uncertainties and biases in the output of climate models, a bias correction is often needed to make the output suitable for simulations.

According to the Köppen–Geiger climate classification, Thailand has a tropical climate with three classes: tropical monsoon, tropical savannah, and tropical rainforest (Rubel & Kottek 2010). The country suffers climate extremes; where, on the one hand, it experienced a drought in 2010 and on the other, a devastating flood occurred in 2011 due to unexpected heavy rainfall, mismanagement of reservoir operations, and inadequate drainage capacities. Hirabayashi *et al.* (2013) has stated that ongoing climate change may cause an increase in the occurrence of such massive floods. On the other hand, as per the future projections, the frequency of extreme droughts is increased (Hunukumbura & Tachikawa 2012). Such anomalous events create different problems among the economic sectors of the country. It was estimated that economic loss due to the 2011 flood was US\$45.7 billion (AON Corporation 2012). Thus, improving the management and planning of water resources has become a necessity. The forecasting models developed in the late 1990s for the Chao Phraya River Basin were mainly focused on real time and event-based simulation. Seasonal rainfall forecast was still missing. Singhrattna *et al.* (2005b) developed a rainfall forecast model for Thailand's summer monsoon using linear regression and non-parametric techniques. Singhrattna & Babel (2011) determined the effects of climate change on summer monsoon rainfall in the Ping River Basin by incorporating large-scale atmospheric variables (LSAVs) and developing the modified k-nearest neighbor (k-*nn*) model. Long-term prediction using traditional statistical methods is difficult (Kumarasiri & Sonnadara 2006) because rainfall dynamics are dependent on highly unpredictable physical parameters. Analysis of this kind of dynamical phenomena by an intelligent system like ANN has given better results than classical methods in the past recent years and they have become most popular in rainfall forecasting. Afshin

et al. (2011) obtained the best performance for two-year rainfall forecasting by an ANN model for Karoon Basin, Iran. In 2010, Chattopadhyay *et al.* suggested the suitability of ANN over other methods for rainfall forecasting in other regions in the world using the sea surface temperature (SST) anomaly (Chattopadhyay *et al.* 2010). The ANN model gave good results for the forecast of the monthly August rainfall from ocean-atmospheric indices in India with the use of fuzzy *c*-mean clustering (Srivastava *et al.* 2010). This approach was suggested to apply for forecasting rainfall of any period in regions where rainfall is significantly related to climate indices. There was better performance of rainfall forecasting on a yearly and monthly basis in Alexandria, Egypt using ANN than multi-layer regression (MLR) model (El-Shafie *et al.* 2011). Silva & Mendes (2015) stated that the ANN technique provides more ability to predict the present climate when compared to the MLR technique for CMIP5 data. Therefore, in the current study, the ANN technique was developed to forecast seasonal rainfall using LSAVs. Then, the developed model was incorporated with the projected LSAVs obtained from the GCMs of CMIP5 to determine the effects of climate change on rainfall in the Ping River Basin, Thailand.

STUDY AREA

The Chao Phraya River Basin is the largest basin in Thailand. It covers about one-third of the country's land area (i.e., 178,000 km²). The Ping River Basin (Figure 1), lying between 15°45'–19°45'N latitudes and 98°05'–100°09'E longitudes is a sub-basin of the Chao Phraya River Basin. It covers an area of 33,896 km² (i.e., 22% of the Chao Phraya River Basin's area). The Ping River is 740 km long and originates from the Pee Pan Nam mountain range in Chiang Mai Province. The basin's elevation varies from 2,500 to 500 m above mean sea level. About two-thirds of the land area (i.e., 24,200 km²) is covered with forests, and the irrigated area is estimated to be around 8,020 km². Rice is the most commonly grown crop here. The sources of water for irrigation are the Ping River and several reservoirs in the basin. The largest multi-purpose dam in this basin, the Bhumibol Dam, has a maximum storage capacity of 13,462 million m³. It receives an average annual inflow of 5,900 million m³ from the Ping River Basin.

The climate of the Ping River Basin is characterized by an average annual precipitation of 1,097 mm and average annual temperature of 26.3 °C. The climate is influenced by two monsoons: the southwest and the northeast monsoons and the depression in the South China Sea from July to September. The monthly rainfall distribution in the basin from 1970 to 2012 is shown in Figure 2. Rainfall distribution over the Ping River Basin which is influenced by the southwest monsoon shows that about 90% of monthly rainfall occurs from May to October (wet season), which can be further divided into two periods: the Wet-1 season from May to July (MJJ) and Wet-2 season from August to October (ASO). The dry period, which lasts from November to April, can also be divided into two periods: one is from November to January of the following year (NDJ) and the other is from February to April (FMA).

DATA COLLECTION

Rainfall data

Daily rainfall data were obtained from the Thailand Meteorological Department (TMD) and the Royal Irrigation Department (RID) of Thailand. Thirty rainfall stations were selected for the study by considering the data availability of each station: only those stations which had less than 5% incomplete data were selected (Figure 1). The total daily rainfall series in each station consists of the data from 1/1/1970 to 31/12/2012. The average rainfall over the basin was calculated using the Thiessen polygon method (Punma & Lal 1992). The total of daily rainfall for each three months (MJJ, ASO, NDJ, FMA) is calculated to find the seasonal rainfall over the basin.

LSAV data

Based on Pearson's correlation coefficient (*r*), statistical relationships between LSAVs and the average rainfall over the Ping River Basin have been developed using correlation maps (Singhrratna *et al.* 2005a, 2012; Singhrratna & Babel 2011). A correlation map is an online analysis provided by the Physical Sciences Division in the Earth System Research Laboratory (ESRL) of the National

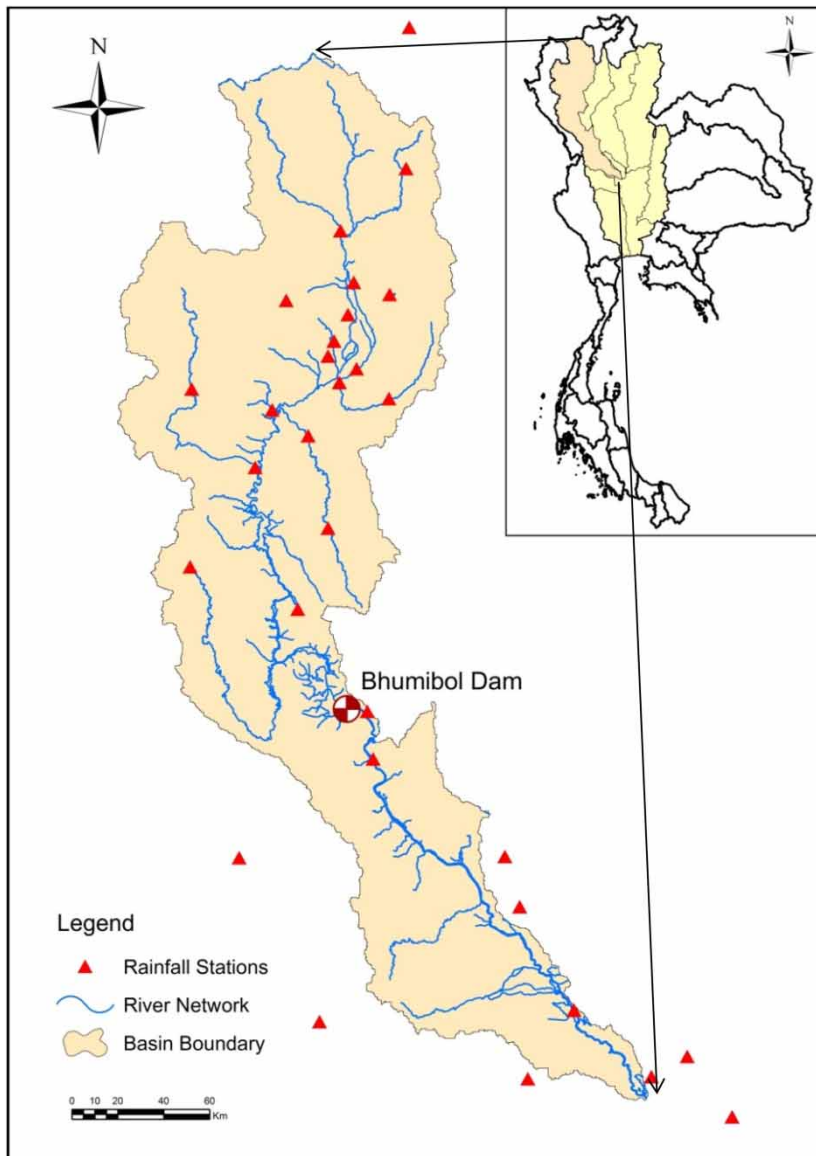


Figure 1 | The Ping River Basin (Thailand).

Centers for Environmental Prediction (NCEP) under the National Oceanic Atmospheric Administration (NOAA) (ESRL 2013). Based on a seasonal time scale, four LSAVs, surface air temperature (SAT), SLP, surface zonal wind (SXW), and surface meridian wind (SYW) were identified for each season of rainfall, i.e., MJJ, ASO, NDJ, and FMA (Singhrattna *et al.* 2012) and were used in a rainfall forecasting model by Singhrattna & Babel (2011). In this study, these identified variables, along with an additional three variables, precipitable water (PW), relative humidity

(RH), and SST were used with several lead times, ranging from 4 to 15 months prior to the start of the season. The predictors were identified over different regions (Figure 3) that are located close to the study basin, such as the South China Sea, and the Indian and Pacific Oceans. It was assumed that the same regions of LSAVs will be affected in the future as well. The data for the monthly observed LSAVs from 1948 to 2013 were obtained from the analysis data of NCEP/NOAA (<http://www.esrl.noaa.gov/psd/cgi-bin/data/timeseries/timeseries1.pl>).

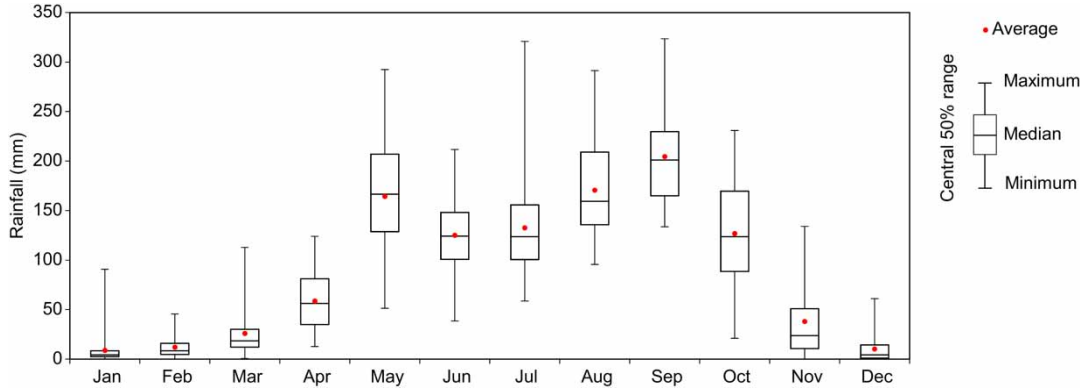


Figure 2 | Box plots of monthly rainfall in the Ping River basin from 1970 to 2012.

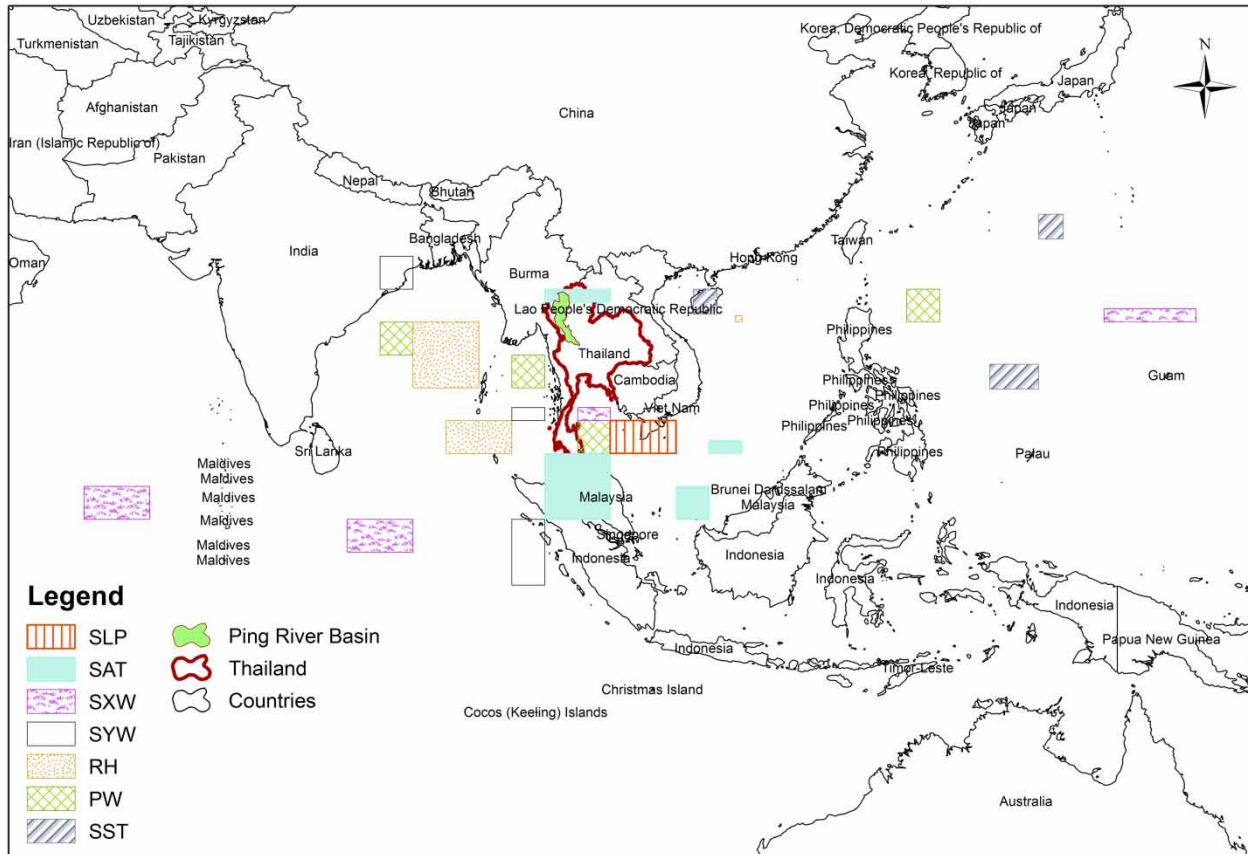


Figure 3 | The identified LSAVs collected from ESRL. SAT, surface air temperature; SLP, sea level pressure; SXW, surface zonal wind; SYW, surface meridional wind; PW, precipitable water; RH, relative humidity; SST, sea surface temperature.

GCM data

Different GCM data were selected by considering different aspects of the study. Singhrattna et al. (2012) selected

GFDL-R30 out of eight GCMs for the study of Ping River Basin based on the data availability. Additionally, for the same basin, the ECHAM4/OPY3 model's data were selected due to their easy accessibility and high temporal

(6 hourly) and spatial resolution among the six GCMs (Sharma et al. 2007). Both studies have recommended certain GCMs (ECHAM4/OPY3, CSIRO Mk2.0, HadCM3, and GFDL-R30) based on the statistical characteristics of data for studies of the Ping River Basin. Finer resolution (1.9° longitude and 1.9° latitude) data are provided by CSIRO Mk2.0 and ECHAM4 and coarser resolution (3.75° longitude and 2.5° latitude) data are provided by HadCM3 and GFDL-R30. Therefore, in this study, the two GCMs with finer resolution, namely, CSIRO Mk3.6 and MPI-ESM-MR, which are extended versions of CSIRO Mk2.0 and ECHAM4, were selected. The monthly data of large-scale atmospheric predictors from 1971 to 2040 under RCP4.5, which represents a balancing pathway of total radioactive forcing less than 4.5 Wm^{-2} in 2100, were used. The data were provided by the Earth System Grid Federation (<http://pcmdi9.llnl.gov/esgfwebfe/live;jsessionid=EFC2F48B4E3C53560E8064810317B168>).

METHODOLOGY

The main aim of this study was to develop seasonal rainfall forecasting models with the use of observed LSAVs data and estimate the future rainfall using LSAVs data obtained from GCMs to analyze the effects of climate change on seasonal rainfall in the Ping River Basin. Based on the relationship between LSAVs and seasonal rainfall, LSAVs have been used as predictors for forecasting models in this study. The ANN technique was used to develop the seasonal rainfall forecasting models using observed predictors. GCMs provided projected data of LSAVs under different climate change scenarios. After bias correction of the GCM data, the developed ANN models were used to forecast seasonal rainfall until 2040. The procedures are described in the following sections.

Selection of the most significant variables

The selection of input variables or independent variables ($x_1, x_2, x_3, \dots, x_n$) is a fundamental step in the development of data-driven models like ANN. All independent variables may not have the same level of importance, because some of them may be correlated with others or noisy or possess

no predictive powers (Bowden et al. 2005). The importance of variable selection and selection methodologies for ANN models have been summarized by Maier & Dandy (2000), Guyon & Elisseeff (2003), Bowden et al. (2005), May et al. (2011), Nourani & Parhizkar (2013) and Nourani et al. (2015). The most commonly used methods are variable ranking method, mutual information, Gamma test, self-organizing map, and ANN modeling. For this study, the variable ranking method was adopted. Table 1 shows that 88 potential independent variables relevant for this study were tested for each season of rainfall. The lead time was defined as the time difference in months between the last month of the forecasted season and the last month of the independent variable's season. For example, if $SST_{(t-4)}$ is a predictor for MJJ, it means the season for SST is JFM (January–February–March) and the lead time is four months. As a result, MJJ can be forecasted at the beginning of April, which is one month prior to the MJJ season.

First, the selection of independent variables (x) was done by correlation analysis using the variable ranking method (Guyon & Elisseeff 2003). The correlation matrices of Pearson's correlation coefficients (r) were developed for all dependent (y) and independent variables. The independent variables were short listed in the 85th percentile category based on an absolute correlation between that independent variable and a dependent variable when the correlation was greater than $2/\sqrt{n}$ (where n is the number of data points). All the screened variables, which were in the 85th percentile category, were cross-checked between two variables for linear dependence. If the correlation between two screened variables was more than 0.75 (i.e. $r > 0.75$), then that variable, which indicates a smaller r value between itself and the dependent variable, was eliminated. For example, let us suppose there are three independent variables, X, Y, and Z. X and Y are strongly

Table 1 | Potential independent variables for each season of rainfall

Season of rainfall	MJJ (t), ASO (t), NDJ (t), FMA (t)
Independent variables (x)	$SAT_{(t-a)}$, $SLP_{(t-a)}$, $SXW_{(t-a)}$, $SYW_{(t-a)}$, $PW_{(t-a)}$, $RH_{(t-a)}$, $SST_{(t-a)}$, $Rainfall_{(t-b)}$
Lead time of independent variable (month)	a = 4, 5, 6, 7, , 15 b = 6, 9, 12, 15

correlated. Among the two, if Y is also strongly correlated ($r > 0.75$) with Z, then Y is omitted and X and Z are selected. If neither X nor Y show a strong correlation with Z, none of them are omitted (Babel & Shinde 2011). Using this selection method, multi-collinearity was eliminated and those independent variables which were statistically most significant for the rainfall were filtered.

In this study, the independent variables include the identified LSAVs at different lead times, from 4 to 15 months. The dependent variables are seasonal rainfall in the study basin. For each season of rainfall, the most significant LSAVs with different lead times were selected, with a maximum number of 6–8 predictors out of 88 potential predictors. Then, all possible combination cases that were equal to $2^n - 1$ (where n is the number of potential predictors) were run with different hidden neurons (i.e., 5 to 20) in one hidden layer. Subsequently, the best performing networks with the best combination cases were identified based on three indices: r and root mean square error (RMSE). The technique applied for selecting the most significant variables is called variable ranking and the model-based embedded method.

ANN

The model algorithm

ANN is a data-driven approach that has the ability to forecast rainfall on different spatial and temporal scales with all types of predictors. The development of ANN consists of data division and pre-processing significant input selection (as described above), network architecture, optimization, model testing, and performance evaluation. This study adopted the ANN technique that was developed using the MATLAB Version 7.12.0 (Sirisena 2013) to select optimal combination cases of independent variables and to forecast seasonal rainfall (i.e., FMA, MJJ, ASO, and NDJ) in the Ping River Basin.

The available data can be split into two groups for training and testing. Past research has concluded that the best trained models are not always good at testing due to over-training (ASCE 2000; Luk et al. 2000; Hung et al. 2009; Mekanik & Imteaz 2012). Cross validation is a highly recommended technique for training a network. Basically, the

training dataset is used to calibrate the weights of the network while the cross validation dataset monitors the progress of the training process. In addition, the testing dataset is used for the final evaluation of the model's performance as well. The dataset for cross validation is not a part of the training dataset. Therefore, in this study, the available dataset of years 1970 to 2007 was divided into three groups, 70%, 15%, and 15% for training, cross validation, and testing, respectively, using random indices.

The input data or independent variables, i.e., LSAVs, of the ANN model were normalized between the upper and lower bounds prior to modeling (Equation (1)) to transform all the variables to the same scale:

$$Data(i) = Amp(i) \times Data(i) + Offset(i) \quad (1)$$

where $Amp(i) = (Upper\ Bound - Lower\ Bound) / (Max(i) - Min(i))$ and $Offset(i) = Upper\ Bound - Amp(i) \times Max(i)$. $Max(i)$ and $Min(i)$ are maximum and minimum values of the input data. The upper and lower bounds are 1 and 0, respectively.

Basically, the network's architecture determines the way information flows through the network. The main goal here is to find a simple network with the best performance. Optimal network geometry or the number of hidden nodes and layers and connection weights is determined by trial and error. The feed forward neural network (FFN) has at least three layers: namely, the input layer, the hidden layer, and the output layer. The input layer consists of all possible independent variables. The nonlinear transformation is done from the input layer to the hidden layer. Then, the output of the hidden layers transfers to the output layer. In this study, FFN, having one hidden layer, was trained with 5 to 20 neurons. Network training or learning is a process of parameter estimation aimed at optimizing connection weights. The Levenberg–Marquardt (LM) optimization function was used to train the network and to update the weights and biases accordingly. LM optimization is a supervised learning algorithm, and it is the fastest back propagation algorithm known for minimizing the mean square error (MSE) of a network (Toth et al. 2000). Several hydrological applications have utilized supervised training (ASCE 2000); however, it must be noted that this algorithm is sensitive to initial conditions. Small random weight values were assigned for

every node like processing elements due to the absence of prior knowledge about initial conditions. The upper and lower bounds of the random weights varied from 0 to 0.5.

Sigmoid-type functions such as logistics and the hyperbolic tangent are commonly used transfer functions. As also implemented in Maier & Dandy (2000), the sigmoid transfer function (Equation (2)) was used in the hidden layers, and the linear transfer function was used in the output layer in this study. This is an advantage when extrapolation is needed:

$$f(x) = \frac{1}{1 + e^{-x}}, x \text{ is an activation function of neurons.} \quad (2)$$

The number of epochs is the maximum number of iterations to run (training cycles) before the simulation stops; in this study, the number was set at 100. The MSE function was applied to define the network's performance. This is the function most frequently used in the ANN model's training because it is an easy calculation function that lies close to normal distribution. The training automatically stops when the generalization does not improve, as indicated by the MSE of the cross validation sample.

The model's evaluation

The model's evaluation can be done by statistical and graphical methods. Both testing and training performances in this study were evaluated using the following indices (here X and Y are observed and simulated data, respectively).

I. RMSE that indicates the error in the units of the model's output.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad (3)$$

II. Correlation coefficient (r) that indicates how well the model can replicate the observed data. It ranges from 0 to 1.0 and 1.0 is the perfect match.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^{0.5} \left[\sum_{i=1}^n (Y_i - \bar{Y})^2 \right]^{0.5}} \quad (4)$$

Determination of the effects of climate change on seasonal rainfall

Bias correction of GCM data

Bias correction is the process of identifying and removing systematic biases between simulated and observed data. Several researchers have found that bias correction methods significantly improve the quality of GCM scenario data (Ines & Hansen 2006; Sharma et al. 2007; Piani et al. 2010; Acharya et al. 2012; Johnson & Sharma 2012; Teutschbein & Seibert 2012; Thrasher et al. 2012). Many bias correction methods have been applied for rainfall and temperature correction on different time scales. In this study, we adopted a bias correction method called quantile mapping. Historical data and GCM data were fitted by a two-parameter Gamma distribution, as shown in Equation (5). Then, the cumulative distribution (Equation (6)) of monthly GCM data was mapped to the cumulative distribution of historical data (Equation (7)). Two parameters of Gamma distribution, i.e., shape (α) and scale (β), for a given dataset, are defined as mean = $\alpha\beta$ and variance = $\alpha\beta^2$.

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad (5)$$

$$F(x; \alpha, \beta) = \int_{x_{\text{trunc}}}^x f(t) dt \quad (6)$$

$$F(x_{\text{GCM}}; \alpha, \beta|_{\text{GCM}}) \Rightarrow F(x_{\text{His}}; \alpha, \beta|_{\text{His}}) \quad (7)$$

All hydrological parameters, temperature, pressure, humidity, and PW, which were simulated by two GCMs, CSIRO Mk3.6 and MPI-ESM-MR, were bias corrected by fitting the Gamma distribution. Even though temperature is found to fit best with normal (Gaussian) distribution (Thom 1954; Schoenau & Kehrig 1990), it was fitted to Gamma distribution in this study. Mean and standard deviation (SD) of two distributions are equal in the raw GCM and observed data. When the shape parameter (α) in a gamma distribution is large, the distribution tends to become normal. The data of 1971 to 2000 was taken as the base period to calculate the required distribution parameters (α and β).

To evaluate the bias correction method, seasonal rainfall during the base period was simulated using two approaches: one, with raw GCM data and another, with bias corrected GCM data. The overall performance of both approaches was evaluated by mean, RMSE, and SD.

Rainfall forecasting under climate change scenarios

The calibrated and tested models of seasonal rainfall forecasting were applied to assess the impact of future climate changes on seasonal rainfall. Bias corrected GCM data for LSAVs from CSIRO Mk3.6 and MPI-ESM-MR were used for long-lead rainfall forecasting in the Ping River Basin during 2011–2040.

Trend analysis of forecasted rainfall

Trend analysis was performed to identify any significant changes in a climatic variable over a long time period. Mann–Kendall (Kendall 1975) and Sen's slope (Sen 1968) were the non-parametric test methods used to analyze the trends in climatic data.

Mann Kendall test. This test measures a monotonic relationship between each value of the time series and the remaining values in a sequential order. In the Mann–Kendall test, the null hypothesis H_0 is a series x_1, \dots, x_n that comes from a population where random variables are independent and identically distributed. The alternative hypothesis H_1 of a two-sided test is that the distributions of x_i and x_j are not identical for all $i, j \leq n$ with $i \neq j$. The test statistics Z ,

$$Z = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (8)$$

where,

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{if } (x_j - x_k) > 0 \\ 0 & \text{if } (x_j - x_k) = 0 \\ -1 & \text{if } (x_j - x_k) < 0 \end{cases}$$

The two-tailed test at different significance level is denoted as $\alpha = 0.1, 0.05, 0.01, \text{ and } 0.001$. At a certain

probability level H_0 is rejected in favor of H_1 if the absolute value of Z equals or exceeds $Z_{\alpha/2}$. In this study, the significant level (α) was selected to be 0.05.

Sen's slope. It estimates and compares the changes in the slope with time at each point and takes the median value to define the trend of the data. Sen's slope can be calculated using the following equation:

$$S = \text{median} \left[\frac{(x_j - x_k)}{j - k} \right] \text{ for } i = 1, 2, 3 \dots n \text{ data} \quad (9)$$

RESULTS AND DISCUSSION

Selected combination cases and model networks

The most significant ' n ' number of variables under the 85th percentile category was obtained using the variable ranking method. It was found that the dominant predictors for all the seasons were pressure, temperature, and wind. ASO and FMA rainfall are also correlated to RH and PW (Table 2).

The ANN technique was employed to develop a forecast model for each season. At the first stage, a trial and error method was used to select the best combination of input variables and the best network architecture. The best networks and their respective best variable combinations were selected according to the performance criteria: r and RMSE among $2^n - 1$ combination cases of ' n ' input variables. Then, the best network was checked for a testing

Table 2 | The most potential predictors for each season of rainfall

Season of rainfall			
FMA	MJJ	ASO	NDJ
$SYW_{(t-8)}$	$SYW_{(t-12)}$	$SLP_{(t-8)}$	$SXW_{(t-6)}$
$SYW_{(t-4)}$	$SXW_{(t-4)}$	$SLP_{(t-15)}$	$SYW_{(t-12)}$
$PW_{(t-4)}$	$SST_{(t-4)}$	$SLP_{(t-10)}$	$SYW_{(t-11)}$
$SLP_{(t-6)}$	$SYW_{(t-10)}$	$RH_{(t-5)}$	$SST_{(t-4)}$
$SLP_{(t-12)}$	$SAT_{(t-12)}$	$SYW_{(t-9)}$	$SXW_{(t-5)}$
$SXW_{(t-14)}$	$SLP_{(t-5)}$	$SAT_{(t-12)}$	$SAT_{(t-8)}$
$SST_{(t-5)}$			$SYW_{(t-10)}$
			$PW_{(t-8)}$

with ‘out of sample dataset’ (2008–2012). The best network architectures and corresponding combination cases for each n^c_r (where n is the total number of variables, and r is the number of variables used for each combination case) have been summarized for the case of ASO rainfall forecast in Table 3. The selected networks for further study are in bold letters. Following the same procedure, the selected networks and predictors for each season of rainfall were obtained and summarized in Table 4. The proposed models for MJJ and FMA are able to forecast one month prior to each season and the models for ASO and NDJ are able to do so two months prior to each season according to the lag time of the best predictors.

The model’s testing with ‘out of sample dataset’

For the model’s testing with ‘out of sample dataset’, the modeled results which were simulated using observed inputs for each season were compared with the observed data of each year from 2008 to 2012 (Figure 4). For the FMA season, underestimations and overestimations of rainfall vary from 17 to 55 mm. MJJ rainfall from the model is overestimated, except in 2009, and 11–86 mm is considered to be the model’s excess. Throughout this five-year period (2008–2012), ASO rainfall from the model is underestimated, except for the years 2009 and 2012, in which there are overestimations. The model overestimates rainfall in the NDJ

Table 3 | Best combination cases as an example for ASO rainfall season

n^c_r	Input variables	Hidd. N.	Training		Testing	
			RMSE/mm	r	RMSE/mm	r
1	<i>SLP</i> _(<i>t</i>-8)	5	67	0.51	37	0.68
2	<i>SLP</i> _(<i>t</i>-8) <i>SYW</i> _(<i>t</i>-9)	13	52	0.74	59	0.80
3	<i>SLP</i> _(<i>t</i>-8) <i>SLP</i> _(<i>t</i>-13) <i>RH</i> _(<i>t</i>-5)	9	50	0.77	70	0.80
4	<i>SLP</i> _(<i>t</i>-8) <i>SLP</i> _(<i>t</i>-13) <i>RH</i> _(<i>t</i>-5) <i>SYW</i> _(<i>t</i>-9)	7	41	0.75	38	0.75
5	<i>SLP</i> _(<i>t</i>-8) <i>SLP</i> _(<i>t</i>-13) <i>SLP</i> _(<i>t</i>-10) <i>RH</i> _(<i>t</i>-5) <i>SYW</i> _(<i>t</i>-9)	9	32	0.86	68	0.83
6	<i>SLP</i> _(<i>t</i>-8) <i>SLP</i> _(<i>t</i>-13) <i>SLP</i> _(<i>t</i>-10) <i>RH</i> _(<i>t</i>-5) <i>SYW</i> _(<i>t</i>-9) <i>SAT</i> _(<i>t</i>-12)	14	58	0.71	64	0.77

Table 4 | Summary of selected networks for further study

Season of rainfall	Network architecture (in-hidden-out)	Selected combination case
FMA	5–20–1	<i>SYW</i> _(<i>t</i>-8) <i>PW</i> _(<i>t</i>-4) <i>SLP</i> _(<i>t</i>-6) <i>SLP</i> _(<i>t</i>-12) <i>SXW</i> _(<i>t</i>-14)
MJJ	4–19–1	<i>SYW</i> _(<i>t</i>-12) <i>SXW</i> _(<i>t</i>-4) <i>SYW</i> _(<i>t</i>-10) <i>SAT</i> _(<i>t</i>-12)
ASO	5–9–1	<i>SLP</i> _(<i>t</i>-8) <i>SLP</i> _(<i>t</i>-13) <i>SLP</i> _(<i>t</i>-10) <i>RH</i> _(<i>t</i>-5) <i>SYW</i> _(<i>t</i>-9)
NDJ	4–8–1	<i>SXW</i> _(<i>t</i>-6) <i>SYW</i> _(<i>t</i>-11) <i>SXW</i> _(<i>t</i>-5) <i>SYW</i> _(<i>t</i>-10)

season, except in 2008. Model excess varies from 2 to 34 mm.

Evaluation of bias correction during the base period (1971–2000)

The monthly LSAVs from GCM data are not accurate when simulating rainfall at the basin level. During the period of 1971–2000, it was observed that CSIRO Mk3.6 and MPI-ESM-MR over- or under-predicted mean monthly temperature, humidity, PW, pressure, and wind. Thus, the direct application of GCM outputs of LSAVs is the cause of uncertainty when forecasting seasonal rainfall. In addition, uncertainty is also introduced by the ANN model itself. Wind (i.e., SXW and SYW) is a vector. The quantile mapping bias correction method was not applicable in this study because it cannot correct for the direction of the wind. Therefore, wind bias correction was not done; instead, raw GCM data were directly applied to the ANN model.

The performance of the bias correction method applied to both GCM datasets was evaluated by mean, SD, and RMSE with respect to the observed data during 1971–2000 (Table 5). When compared to the raw GCM data, the monthly mean and SDs of all corrected variables were found to be closer to deduced values for field observations. The highest RMSE of PW and RH lies between 4.3 and 2.5 after bias correction. Based on the statistical parameters, both bias corrected GCMs’ data can be used for future rainfall projections in the study area.

The proposed ANN models were evaluated using raw GCM and corrected GCM data under the RCP4.5 scenario during the base period (1971–2000). The model

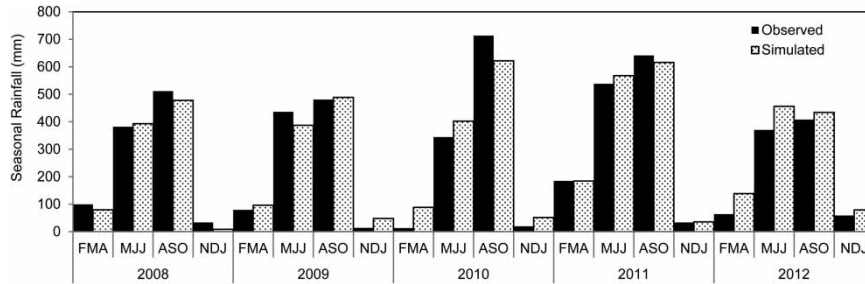


Figure 4 | Model testing with 'out of sample dataset' from 2008 to 2012.

Table 5 | Performance of quantile mapping bias correction method on monthly data

Season	Statistics	Variable	Observed	Raw		Corrected	
				CSIRO Mk3.6	MPI-ESM-MR	CSIRO Mk3.6	MPI-ESM-MR
FMA season	Mean	PW (kg/m ²)	42.9	46.2	44.2	42.9	42.9
		SLP (mbar)	1,012.9	1,013.2	1,012.5	1,012.9	1,013.0
	SD	PW (kg/m ²)	5.8	12.6	10.9	5.8	5.8
		SLP (mbar)	1.1	1.3	1.4	1.0	1.0
	RMSE	PW (kg/m ²)		9.0	7.2	4.3	4.0
SLP (mbar)			1.4	1.5	1.2	1.1	
MJJ season	Mean	SAT (°C)	21.3	25.0	23.2	21.3	21.3
	SD		2.3	5.4	4.2	2.4	2.4
	RMSE			4.3	4.6	1.2	1.1
ASO season	Mean	SLP (mbar)	1,010.1	1,007.4	1,010.5	1,010.1	1,010.1
		RH (%)	77.0	83.5	79.0	77.0	77.0
	SD	SLP (mbar)	3.0	4.8	4.2	3.0	3.0
		RH (%)	5.0	4.2	5.8	5.0	4.9
	RMSE	SLP (mbar)		3.0	2.8	1.4	1.3
		RH (%)		7.2	3.8	2.5	2.6

performance in terms of the statistical parameters (i.e., mean, SD, and RMSE) is given in Table 6. It is clear that the model's performance is better when the simulation incorporates observed variables than when it incorporates GCM data. The simulation of wet seasonal rainfall using MPI-ESM-MR inputs shows better performance than CSIRO Mk3.6 inputs. On the other hand, the simulation of dry season rainfall is more accurate when using CSIRO Mk3.6 inputs rather than MPI-ESM-MR data.

Effects of climate change on seasonal rainfall

With the use of projected LSAVs by GCMs and the proposed ANN models, seasonal rainfall in the Ping River Basin was forecasted until 2040. The base period data from 1971 to 2000 and the climatology period data from

2011 to 2040 were analyzed to discover the trends in observed and forecasted rainfall. Table 7 presents the results of the regression analysis, and the Mann–Kendall test, including Sen's slope. While none of the trends is significant at 95% confidence level ($\alpha = 0.05$), the directions of these trends are similar in the three analyses for all the seasons.

The results are shown in terms of standardized anomalies, which are the measure of the distance between a data value and its mean. Standardized anomaly removes the influences of location and space from the data. Rainfall's standardized anomalies from 2011 to 2040 were calculated with respect to observed rainfall during the base period (1971–2000) using Equation (10). Figure 5 illustrates the annual standardized anomalies of observed rainfall in the Ping River Basin during 1971–2012 and the annual standardized anomalies of forecasted rainfall using bias corrected

Table 6 | ANN model's performance during the base period (1971–2000)

Parameter	Observed	With observed inputs	With raw GCM inputs		With bias corrected inputs	
			Sim_A	Sim_B	Sim_A	Sim_B
FMA season						
Mean (mm)	97	100	107	103	91	92
SD (mm)	41	43	66	51	36	44
RMSE (mm)		28	76	72	61	68
MJJ season						
Mean (mm)	415	395	310	241	339	384
SD (mm)	93	97	44	64	63	55
RMSE (mm)		77	142	210	122	101
ASO season						
Mean (mm)	495	502	817	687	499	499
SD (mm)	69	49	54	111	48	50
RMSE (mm)		54	335	246	71	105
NDJ season						
Mean (mm)	61	56	85	86	85	86
SD (mm)	44	29	34	42	34	42
RMSE (mm)		41	53	62	53	62

data of the two GCMs.

$$\text{Standardized anomaly} = \frac{(\text{Forecasted value} - \text{Mean value})}{\text{Standard Deviation}} \quad (10)$$

During the base period (1971–2000) of the MJJ season, maximum and minimum rainfall values were recorded as 605 and 247 mm in 1978 and 1997, respectively. Rainfall anomalies varied from 2.0 to –1.8 during the base period (Figure 5(a)). From both GCMs under the RCP 4.5 scenario, it can be seen that MJJ rainfall lies below the normal level with respect to observed rainfall during the base period. According to Singhrattna & Babel (2011), under A2 and B2 scenarios, MJJ rainfall in 2011–2100 will decrease on average by 0.1 and 0.17 mm yr⁻¹, respectively. In contrast, the multi-model dataset (MMD), as reported by the IPCC (2007), indicates that an increasing trend is expected in June, July, and August over Thailand, Cambodia, Laos, Malaysia, and Indonesia. The anomalies in ASO rainfall vary from 2.5 to –1.8 (Figure 5(b)). The post-1980s ASO

Table 7 | Summary of results of trend analysis

Season	Regression analysis	Mann-Kendall test (Z)	Sen's slope (S)	At 95% of confidence level	
				Smin	Smax
Observed data (1971–2000)					
MJJ	–1.10	–0.46	–1.21	–5.469	3.175
ASO	–2.20	–1.43	–2.13	–5.274	1.063
NDJ	–1.20	–0.86	–0.73	–3.078	1.078
FMA	1.25	0.75	0.76	–1.128	2.795
CSIRO Mk3.6 (2011–2040)					
MJJ	–1.10	–0.89	–1.16	–3.958	1.127
ASO	1.08	0.43	0.42	–2.209	3.138
NDJ	–0.43	–1.28	–0.44	–1.567	0.197
FMA	–1.32	–0.93	–1.25	–3.340	0.984
MPI-ESM-MR (2011–2040)					
MJJ	–1.80	–1.46	–1.88	–4.208	0.448
ASO	0.90	0.36	0.45	–2.964	4.183
NDJ	–0.61	–0.86	–0.04	–0.289	0.070
FMA	0.92	0.86	1.11	–1.323	3.329

rainfall was below normal; however, above-normal rainfall has been observed after 2009. As a result, compared to a decreasing trend of historical rainfall, the effect of climate change on ASO rainfall in the Ping River Basin causes an increasing trend in forecasted rainfall with both GCMs considered in the study. In addition, Bhaskaran & Mitchell (1998) concluded that the beginning of the monsoon season in Thailand and nearby regions such as Laos, Cambodia, and Vietnam will tend to get delayed by 10–15 days without there being any changes in the duration of the season during the 2030–2070s under the A1B scenario of climate change. However, decreasing trends at 1.09 and 6.16 mm yr⁻¹ in 2011–2100 ASO rainfall are estimated under A2 and B2 scenarios, respectively, from GFDL-R30 data (Singhrattna & Babel 2011). The NDJ rainfall anomalies vary between 2.5 and –1.3 during the base period (Figure 5(c)). As mentioned earlier, NDJ rainfall was forecasted using raw GCM data of SXW and SYW without any bias correction. It is also to be noted that the effects of climate change on NDJ rainfall cause a slower decreasing trend compared to the trend of observed rainfall. The standardized anomalies of the FMA season show almost normal conditions with both GCMs during the base period

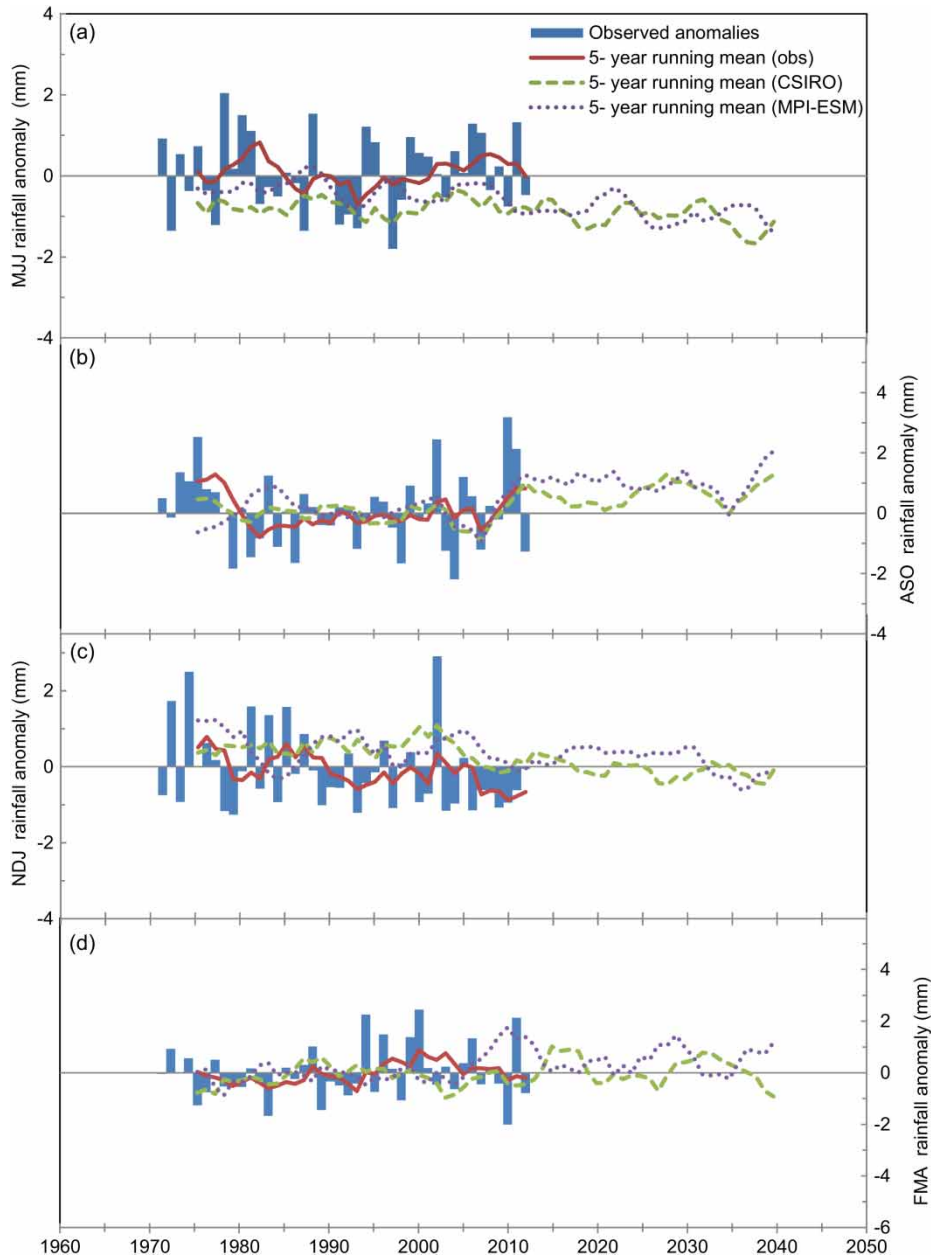


Figure 5 | Annual standardized anomalies of rainfall during (a) MJJ, (b) ASO, (c) NDJ, and (d) FMA.

(Figure 5(d)). As compared to the base period's climatology, FMA rainfall will tend to decrease based on CSIRO Mk3.6 and increase based on MPI-ESM-MR. These findings are inconsistent with the MMD in the IPCC report (2007) that shows only increasing trends of seasonal rainfall by 7% during 2080–2099 over Thailand and Southeast Asia.

Finally, there are limitations and uncertainties inherent in the predictors, ANN model and GCM data, thus quality

of forecasting was limited. It is generally accepted that ANN performs best when extrapolation is within the range of data used for calibration (Flood & Kartam 1994). Thus, rainfall can be underestimated for extreme cyclonic events. Also, this study is limited due to the lead time of SST as 4–15 months, but ENSO (index calculated from SST) effects vary from two to seven years. Therefore, model predictors could incorporate such variables to develop the ANN

model. Bias in GCM data is corrected by quantile mapping which is limited in capturing extremes beyond the range of historical data.

CONCLUSIONS

Long-term rainfall forecast is a crucial issue for several water-related sectors such as industry, agriculture, health, and transportation. In this study, the most potential predictors of LSAVs were identified using the variable ranking method. Pressure, temperature, and wind were found to be the most significant predictors among the eight identified predictors for seasonal rainfall forecast in the Ping River Basin. Seasonal rainfall over the Ping River Basin is influenced by atmospheric variables over the regions near Thailand, such as the Pacific and Indian Oceans and the South China Sea. The proposed forecasting models of MJJ and FMA rainfall were able to forecast rainfall one month in advance and the forecasting models for ASO and NDJ rainfall were able to do so two months in advance. The efficiency of the model to forecast dry season rainfall can be improved with the incorporation of local climate conditions. A bias correction method, namely quantile mapping, also can improve GCM simulations, except for the wind components. Under the RCP4.5 scenario, a decreasing trend in MJJ rainfall and an increasing trend in ASO rainfall from 2011 to 2040 have been predicted. In the same period, NDJ rainfall is predicted to decrease and FMA rainfall to increase linearly.

Moreover, the Ping River Basin will receive above-normal ASO rainfall and below-normal NDJ rainfall as compared to the rainfall in 1971–2000. The analysis results and previous studies show uncertainty in rainfall patterns in the study region and nearby areas. Given such rainfall variability in the future, adaptation measures will have to be planned for water management in order to deal with possible water crises: scarcity or excess. The study results can be coupled with a hydrological model to simulate stream flow so as to analyze the uncertainty of water availability in the basin. The operations of the Bhumibol reservoir and other water storage facilities can be managed and optimized to cope with anomalous weather events (wet or dry). With the aim of decreasing potential damage, the responsible

governmental agency can set up an initial long-term plan to manage water resources and to allocate water to different users appropriately under different hydrological conditions in the basin.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Sutat Weesakul and Dr Sylvain Perret for their valuable comments. We are grateful to the Royal Thai Government and the WEM project for their financial support to the second author of the paper for her master's degree at the Asian Institute of Technology, Thailand. The authors also thank the TMD and the RID for providing rainfall data for this study. Our special thanks go to the anonymous reviewers and the editor for their valuable contributions to the manuscript.

REFERENCES

- Acharya, N., Chattopadhyay, S., Mohanty, U. C., Dosh, S. K. & Sahoo, L. N. 2012 [On the bias correction of general circulation model output for Indian summer monsoon](#). *J. Meteorol. Appl.* **20**, 349–356.
- Afshin, S., Fahmi, H., Alizadeh, A., Sedghi, H. & Kaveh, F. 2011 Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in Karoon basin. *J. Sci. Res. Ess.* **6**, 1200–1208.
- American Society Civil Engineers 2000 [Artificial neural networks in hydrology. I: preliminary concepts](#). *J. Hydrol. Eng.* **5**, 115–123.
- AON Corporation 2012 [2011 Thailand Floods Event Recap Report](#). Impact Forecasting March 2012, 38. Available at: <http://thoughtleadership.aonbenfield.com/Pages/Home.aspx?ReportYear=2012> (accessed 30 March 2013).
- Babel, M. S. & Shinde, V. R. 2011 [Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok](#). *Water Res. Manage.* **25**, 1653–1676.
- Bhaskaran, B. & Mitchell, J. F. B. 1998 [Simulated changes in Southeast Asian monsoon precipitation resulting from anthropogenic emissions](#). *Int. J. Climatol.* **18**, 1455–1462.
- Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 [Input determination for neural network models in water resources applications. Part 1 – background and methodology](#). *J. Hydrol.* **301**, 75–92.
- Bridhikitti, A. 2012 [Associations of ENSO/IOD, Aerosols and Land Cover Changes with Thailand Rainfall Anomalies and Implication to Rainfall Forecast](#). APEC Climate Center,

- Korea. Available at: http://www.apcc21.net/eng/common/file_dn.jsp (accessed 21 August 2012).
- Chansaengkachang, K., Ascharyapho, N., Humphries, U., Wngwongchai, A. & Wongwises, P. 2011 Empirical orthogonal function analysis of rainfall over Thailand and its relationship with Indian Ocean Dipole. In: *Proceedings of Chiangmai University International Conference*, Thailand, 1, 47–54.
- Chattopadhyay, G., Chattopadhyay, S. & Jain, R. 2010 Multivariate forecast of winter monsoon rainfall in India using SST anomaly as a predictor: neuro-computing and statistical approaches. *J. Com. Ren. Geosci.* **342**, 755–765.
- Davolio, S., Miglietta, M. M., Diomede, T., Marsigli, C., Morgillo, A. & Moscatello, A. 2008 Ameteo-hydrological prediction system based on a multi-model approach for precipitation forecasting. *Nat. Hazard. Earth Syst. Sci.* **8**, 143–159.
- Deser, C., Knuti, R., Solomon, S. & Phillips, A. S. 2012 Communication of the role of natural variability in future North American climate. *Nat. Clim. Change* **2** (11), 775–779.
- Diomede, T., Davolio, S., Marsigli, C., Miglietta, M. M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi, A. & Malguzzi, P. 2008 Discharge prediction based on multi model precipitation forecasts. *Meteorol. Atmos Phys.* **101**, 245–265.
- Donate, J. P., Cortez, P., Sánchez, G. G. & Miguel, A. S. 2013 Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble. *Neuro-com.* **109**, 27–32.
- El-Shafie, A. H., El-Shafie, A., El Mazoghi, H. G., Shehata, A. & Taha, M. R. 2011 Artificial neural network technique for rainfall forecasting applied to Alexandria, Egypt. *Int. J. Phys. Sci.* **6**, 1306–1316.
- El-Shafie, A., Noureldin, A., Taha, M. R. & Hussain, A. 2012 Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia. *Hydrol. Earth Syst. Sci.* **16**, 1151–1169.
- ESRL 2013 Interactive plotting and analysis: Linear monthly/seasonal correlations. Available at: <http://www.esrl.noaa.gov/psd/data/correlation/> (accessed 27 May 2013).
- Flood, I. & Kartam, N. 1994 *Neural networks in civil engineering. I: principles and understanding*. *J. Com. Civil Eng.* **8** (2), 131–148.
- Gadgay, B., Kulkarni, S. & Chandrasekhar, B. 2011 A neural network model to predict rainfall in large Indian cities. *World J. Sci. Tech.* **1**, 6–11.
- Ganguly, A. R. & Bras, R. L. 2003 Distributed quantitative precipitation forecasting (DQPF) using information from radar and numerical weather prediction models. *J. Hydrometeor.* **4**, 1168–1180.
- Ghiassi, M. & Nangoy, S. 2009 Dynamical artificial neural network model for forecasting nonlinear processes. *Com. Indust. Eng.* **57**, 287–297.
- Ghosh, S. & Mujumdar, P. P. 2006 Future rainfall scenario over Orissa with GCM projections by statistical downscaling. *J. Curr. Sci.* **90**, 396–404.
- Golabi, M. R., Radmanesh, F., Akhondali, A. & Kashefipoor, M. 2013 Prediction of seasonal precipitation using artificial neural networks (case study: selected stations of Iran) Khozestan province). *J. Basic Appl. Sci Res.* **3**, 589–595.
- Guhathakurta, P. 2006 Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network. *Current Sci.* **90**, 773–779.
- Guyon, I. & Elisseeff, A. 2003 An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182.
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H. & Kanae, S. 2013 Global flood risk under climate change. *Nature Clim. Change* **3**, 816–821.
- Hung, N. Q., Babel, M. S., Weesakul, S. & Tripathi, N. K. 2009 An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *J. Hydrol. Earth Syst. Sci.* **13**, 1413–1425.
- Hunukumbura, P. B. & Tachikawa, Y. 2012 River discharge projection under climate change in the Chao Phraya River basin. *J. Met. Soc. Japan Ser.* **90A**, 137–150.
- Ines, A. V. M. & Hansen, J. W. 2006 Bias correction of daily GCM rainfall for crop simulation studies. *J. Agr. Forest Met.* **138**, 44–53.
- IPCC (Working Group I) 2007 *IPCC fourth Assessment Report: Climate Change 2007*. Inter Governmental Panel on Climate Change, Geneva, Switzerland.
- Johnson, F. & Sharma, A. 2012 A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *J. Water Resour. Res.* **48**, W01504.
- Kashid, S. S. & Maity, R. 2012 Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming. *J. Hydrol.* **454–455**, 26–41.
- Kendall, M. G. 1975 *Rank Correlation Methods*, 4th edn. Charles Griffin, London, UK.
- Krishnamurthy, V. & Shakla, J. 2007 Inter-seasonal and seasonal persisting patterns of Indian monsoon rainfall. *J. Clim.* **20**, 3–20.
- Krishnan, R., Zhang, C. & Sugi, M. 2000 Dynamics of breaks in the Indian summer monsoon. *J. Atmos. Sci.* **57**, 1354–1372.
- Kumarasiri, A. D. & Sonnadara, D. U. J. 2006 Rainfall forecasting: an artificial neural network approach. *Proceedings of the Technical Sessions, Institute of Physics, Sri Lanka* **22**, 1–16.
- Li, F. & Zeng, Q. 2008 Statistical prediction of East-Asian summer monsoon rainfall based on SST and sea ice concentration. *J. Meteorol. Soc. Jpn* **86**, 237–243.
- Luchetta, A. & Manetti, S. 2003 A real time hydrological forecasting system using a fuzzy clustering approach. *J. Com. Geosci.* **29**, 1111–1117.
- Luk, K. C., Ball, J. E. & Sharma, A. 2000 A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *J. Hydrol.* **227**, 56–65.
- Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environ. Modell. Softw.* **15**, 101–124.
- May, R., Dandy, G. C. & Maier, H. 2011 Review of input variable selection methods for artificial neural networks. In Tech

- Open Publishers. Available at: <http://cdn.intechopen.com/pdfs-wm/14882> (accessed 8 July 2013).
- Mekanic, F. & Imteaz, M. A. 2012 A multivariate artificial neural network approach for rainfall forecasting. In: *Proceedings of the World Congress on Engineering and Computer Science 2012*, Case study of Victoria, Australia. Vol I. Available at: http://www.iaeng.org/publication/WCECS2012/WCECS2012_pp557-561 (accessed 19 April 2013).
- Munot, A. A. & Kumar, K. K. 2007 Long range prediction of Indian summer monsoon rainfall. *J. Earth Syst. Sci.* **116** (1), 73–79.
- Nayagam, L. R., Janardanan, R. & Mohan, H. S. R. 2008 An empirical model for the seasonal prediction of south west monsoon rainfall over Kerala, a meteorological sub division of India. *Int. J. Climatol.* **28**, 823–831.
- Nourani, V. & Parhizkar, M. 2013 Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall-runoff modeling. *J. Hydroinform.* **15** (3), 829–848.
- Nourani, V., Khanghah, T. R. & Baghanam, A. H. 2015 Application of entropy concept for input selection of wavelet-ANN based rainfall-runoff modeling. *J. Environ. Inform.* **26** (1), 52–70.
- Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S. & Haerter, J. O. 2010 Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.* **395**, 199–215.
- Punna, B. C. & Lal, P. B. B. 1992 *Irrigation and Water Power Engineering*, 12th edn. Laxmi Publications, New Delhi, India, pp. 108–109.
- Ramirez, M. C. V., Velho, H. F. C. & Ferreira, N. J. 2005 Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *J. Hydrol.* **301**, 146–152.
- Rubel, F. & Kottke, M. 2010 Observed and projected climate shifts 1901–2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorol. Z.* **19**, 135–141.
- Salathe, E. P. 2003 Comparison of various precipitation downscaling methods for the simulation of stream flow in a rain shadow river basin. *Int. J. Climatol.* **23**, 887–901.
- Schoenau, G. J. & Kehrig, R. A. 1990 A method for calculating degree-days to any base temperature. *Energy Build.* **14** (4), 299–302.
- Sen, P. K. 1968 Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Ass.* **63**, 1379–1389.
- Sharma, D., Gupta, A. D. & Babel, M. S. 2007 Spatial disaggregation of bias corrected GCM precipitation for improved hydrologic simulation: Ping river basin, Thailand. *Hydrol. Earth Syst. Sci.* **11**, 1373–1390.
- Shepherd, T. G. 2014 Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7** (10), 703–708.
- Shukla, R. P., Tripathi, K. C., Pandey, A. C. & Das, I. M. L. 2011 Prediction of Indian summer monsoon rainfall using Niño indices: a neural network approach. *Atmos. Res.* **102**, 99–109.
- Silva, G. A. M. & Mendes, D. 2015 Refinement of the daily precipitation simulated by the CMIP5 models over the north of the Northeast of Brazil. *Front. Environ. Sci.* **3**, 29.
- Singhrratna, N. & Babel, M. S. 2011 Changes in summer monsoon rainfall in Upper Chao Phraya River Basin, Thailand. *Climate Res.* **49**, 155–168.
- Singhrratna, N., Rajagopalan, B., Kumar, K. K. & Clark, M. 2005a Inter-annual and inter-decadal variability of Thailand summer monsoon season. *J. Climate.* **18**, 1697–1708.
- Singhrratna, N., Rajagopalan, B., Clark, M. & Kumar, K. K. 2005b Seasonal forecasting of Thailand summer monsoon rainfall. *Int. J. Climatol.* **25**, 649–664.
- Singhrratna, N., Babel, M. S. & Perret, S. R. 2012 Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large scale atmospheric variables. *J. Hydrol. Sci.* **57**, 26–41.
- Sirisena, T. A. J. G. 2013 Forecasting of Long-term Seasonal Rainfall Using Artificial Neural Network: An Application to the Ping River Basin, Thailand. Master thesis, Asian Institute of Technology, Thailand.
- Srivastava, G., Panda, S. N., Mondal, P. & Liu, J. 2010 Forecasting of rainfall using ocean-atmospheric indices with Fuzzy neural technique. *J. Hydrol.* **395**, 190–198.
- Taylor, K. E., Stouffer, R. J. & Meehl, G. A. 2012 An overview of CMIP5 and the experiment design. *Am. Meteorol. Soc.* **93**, 485–498.
- Teutschbein, C. & Seibert, F. 2012 Bias correction of regional climate model simulations for hydrological climate change impact studies: review and evaluation of different methods. *J. Hydrol.* **29**, 456–457.
- Thom, H. C. S. 1954 The rational relationship between heating degree days and temperature. *Mon. Weather Rev.* **82** (1), 1–6.
- Thrasher, B., Maurer, E. P., McKellar, C. & Duffy, P. B. 2012 Technical note: bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol. Earth Syst. Sci.* **16**, 3309–3314.
- Toth, E., Brath, A. & Montanari, A. 2000 Comparison of short term rainfall prediction models for real time flood forecasting. *J. Hydrol.* **239**, 132–147.
- Trtanj, J. M. & Houston, T. G. 2014 Climate variability and change data and information for global public health. In: *Global Climate Change and Public Health* (K. E. Pinkerton & W. N. Rom, eds). Springer, New York, USA.
- Tuleya, R. E., Demaria, M. & Kuligowski, R. J. 2007 Evaluation of GFDL and simple statistical model rainfall forecasts for U.S. land falling tropical storms. *Weather Forecast.* **22**, 56–70.
- Wu, C. L. & Chau, K. W. 2013 Prediction of rainfall time series using modular soft computing methods. *J. Eng. Appl. Artif. Intell.* **26**, 997–1007.
- Wu, C. L., Chau, K. W. & Fan, C. 2010 Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *J. Hydrol.* **389**, 146–167.
- Yaremchuk, M. & Qu, T. 2004 Seasonal variability of the large-scale currents near the coast of Philippines. *J. Phys. Oceanogr.* **34**, 844–855.
- Zaw, W. T. & Naing, T. T. 2008 Empirical statistical modeling of rainfall prediction over Myanmar. *World Acad. Sci. Eng. Tech.* **2**, 492–495.

First received 28 October 2015; accepted in revised form 18 May 2016. Available online 5 July 2016