ORIGINAL PAPER

Efficacy of linear multiple regression and artificial neural network for long-term rainfall forecasting in Western Australia

Anirban Khastagir[1] · Iqbal Hossain[2] · A. H. M. Faisal Anwar[3]

Abstract

Precipitation is one of the most intrinsic resources for manifold industrial activities all over Western Australia; consequently, immaculate rainfall prediction is indispensable for flood mitigation as well as water resources management. This study investigated the performance of artificial neural networks (ANN) and Linear multiple regression (LMR) analysis to forecast long-term seasonal spring rainfall in Western Australia, using lagged El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as potential climatic phenomena. The ANN was developed in the form of multilayer perceptron using Levenberg–Marquardt algorithm and subsequently LMR was used with statistical significance for future spring rainfall forecast. The total climatic dataset has been divided into calibration and testing phases to determine the efficacy of the developed models. Different statistical skill tests such as root mean square error (RMSE), mean absolute error (MAE), and Willmott index of agreement 'd' were used to assess the efficacy of LMR and ANN modelling. In general, LMR has lower MAE and RMSE values as compared to ANN for most of the stations during calibration and testing periods, whereas ANN models performed better than LMR models based on 'd' values. The overall statistical analysis paradigm suggests the efficacy of LMR over ANN models for rainfall forecasting using more climatic variables. As a result, the developed LMR model, incorporated with lagged global climate indices, will facilitate the adequate preparedness for the risks associated with potential droughts in the study region.

1 Introduction

Australia is experiencing manifold extreme events such as bushfire, flood, and drought due to the climate change. Considerable spatial and temporal variations of climatic parameters (such as temperature, relative humidity, and wind speed) are observed in different parts of Australia. Khastagir et al. (2017) observed that mean temperature and relative humidity during summer months (Dec–Feb) are the most critical parameters for the occurrence of frequent bushfire events

✉ Anirban Khastagir
anirban.khastagir@rmit.edu.au

1   School of Vocational Engineering, Health and Sciences, RMIT University, Melbourne, VIC 3000, Australia

2   Department of Civil and Construction Engineering, Swinburne University of Technology, Hawthorn, VIC 3122, Australia

3   Department of Civil Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

in Victoria; consequently, northwest Victoria is more fire prone compared to other parts of Victoria, because of high temperature and relatively low relative humidity (Khastagir et al. 2018). In addition, spatial and temporal rainfall variation of rainfall was observed in Greater Melbourne, the capital of Victoria, Australia (Khastagir 2008). This also enhances the spatial and temporal variation in storm runoff and the streamflow (Islam et al. 2014). An incessant 11-year (1996–2007) drought was observed in Melbourne, when the cumulative rainfall was significantly below the long-term average (Khastagir and Jayasuriya 2008). Immaculate long-term prediction of rainfall is one of the main predicaments confronted by water resources manager due to the intricate ever-changing atmosphere. To facilitate water resources management as well as controlling excessive flooding, an accurate understanding of long-term rainfall prediction is intrinsic. Due to paucity of rainfall, there is a probability of detrimental effect on flora and fauna in the aquatic system.

Climatic parameters can play a significant role in planning and mitigation activities for several key areas of the rural and urban communities such as water supply, infrastructure, agriculture, and extreme event mitigation. A climate index

can be used effectually to illustrate the change of climate dynamics. Large-scale climate phenomena illustrated by the climate indices facilitates the identification of rainfall variability in most parts of the world. Climate phenomena namely El Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), Southern Annular Mode (SAM), and Madden–Julian oscillation (MJO) are widely known for their effect in Australian climate (Mekanik et al. 2013; Hossain et al. 2018b, 2020a). In addition, it is well established that the effects of ENSO on Australian seasonal rainfall is the strongest in the world. Moreover, IOD, a by-product of Indian Ocean sea surface temperature, significantly influences the Western Australian seasonal rainfall; therefore, ENSO and IOD may be used as potential predictors for seasonal rainfall forecasting. Numerous studies have identified that large-scale climatic phenomena are responsible for rainfall variability in Australia and beyond (Yilmaz et al. 2014; Hossain et al. 2018a, 2020a). Most of these studies investigated the influences of ENSO, IOD and IPO on seasonal and extreme rainfall.

A variety of methods can be used to derive the teleconnection of climate indices such as: Southern Oscillation Index (SOI), Dipole Mode Index (DMI), Niño 4, Niño 3.4, Niño 3, Niño 1.2, and the Inter-Decadal Pacific Oscillation (IPO) with Australian rainfall and their variabilities in different parts of Australia. Sea level pressure, sea surface temperature anomalies, geopotential height and precipitation are some of the important variables that need to be considered while using the above stated climate indices (Shams et al. 2018). In particular, several studies have been carried out in Australia and overseas to investigate the relationships between seasonal rainfall and climate phenomenon (Kirono and Kent 2011; Abbot and Marohasy 2012; Mekanik et al. 2013; Hossain et al. 2018a). Saha and Chattopadhyay (2020) carried out theory-based investigation pertaining the time series of rainfall in seasonal scale as well as yearly scale in the Himalayas during the summer monsoon. Pal et al. (2020) carried out study to investigate the behaviour of the time series of rainfall during the summer in northeast India. Bagirov et al. (2017) reported that numerous data-driven prediction models such as linear multiple regression (LMR), autoregressive integrated moving average (ARIMA), the K-nearest-neighbours (K-NN), artificial neural network (ANN), and support vector machines for regression (SVM-reg) are used for rainfall forecasting. Chattopadhyay (2007) investigated the efficacy of ANN model in predicting average summer-monsoon rainfall over India. Chattopadhyay and Chattopadhyay (2018) revealed the efficacy of conjugate gradient descent algorithm for multilayer ANN through Shannon-Fano coding. Chattopadhyay and Chattopadhyay (2009) used an autoregressive neural network (ARNN) model, and the neural network was trained as a multilayer perceptron with the extensive variable selection procedure.

Acharya et al (2013) carried out analysis for developing an artificial neural network based multi-model ensemble with a view to estimating the northeast monsoon rainfall over south peninsular India. In addition, the prediction of northeast monsoon rainfall of seven general circulation models (GCMs) was conducted by Acharya et al. (2011). Nonetheless, Hossain et al. (2020a) noted that the most popular and widely used models for rainfall forecasting in Australia are ANN and LMR.

The non-linear ANN modelling technique is widely used not only in the field of hydrology but also in other areas. Acharya et al. (2014) *developed an ANN-based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India. Abbot and Marohasy (2017) used ANN to forecast the monthly rainfall in Murray–Darling basin, Australia using different climate indices such as SOI, Dipole Mode Index (DMI), Niño 4, Niño 3.4, Niño 3, Niño 1.2 and the Inter-Decadal Pacific Oscillation (IPO). Similarly, Abbot and Marohasy (2017) used ANN using several climate indices such as SOI, PDO, and Nino 3.4 to forecast monthly and seasonal rainfall in Queensland, Australia. Nonetheless, Hossain et al. (2020a) used both ANN and LMR to forecast spring rainfall of three rainfall stations in Western Australia. Chakraverty and Gupta (2008) carried out prediction of southwestern monsoon rainfall in India using ANN models. Bilgili and Sahin (2010) applied ANNs to predict the long-term monthly temperature and rainfall for different stations in Turkey. For the likelihood estimation of water resources and hydrological variables, LMR modelling technique is commonly used (Hossain et al. 2018b).

LMR is also widely used in different parts of the world for rainfall prediction (Hossain et al. 2020a). In this method, independent variables are used to predict one dependent variable using least squares method. Rasel et al. (2015) used LMR to reveal the influences of lagged ENSO and SAM as the potential climate predictors for the long-term rainfall forecasting in South Australia; similarly, Mekanik et al. (2013) examined the influence of lagged ENSO and IOD on Victorian rainfall using LMR technique. Several other studies such as Hossain et al. (2018b, 2020a), Mekanik et al. (2013) and Rasel et al. (2015) analysed either LMR or combined LMR with ANN model, using 3 months of lagged rainfall to forecast rainfall in different parts of Australia. Nonetheless, it is intrinsic to incorporate additional climatic variables to extend the forecasting capability of both the linear and non-liner models to determine the prediction efficacy.

It has been commonly observed that the forecasting capability of non-linear modelling approaches (e.g. ANN) are better than the linear modelling techniques in predicting seasonal rainfall (Adamowski et al. 2012; Mekanik et al. 2013; Djibo et al. 2015; Rasel et al. 2016a, b; Hossain et al. 2020a). Nevertheless, most of the seasonal rainfall predictive

models considered only two independent variables in developing the linear models (Mekanik et al. 2013; Rasel et al. 2016a, b; Hossain et al. 2018b, 2020a). Rainfall is a complex atmospheric global phenomenon, for which only two influential variables may not be sufficient to forecast accurately. Furthermore, non-linear ANN modelling approaches are black box models, which are intricate in nature and require special knowledge for their application. In addition, selection of number of hidden neurons is intrinsic in case of ANN models, for which no formal methods are currently available, other than the trial-and-error technique. As a result, the efficiency of both linear LMR and non-linear ANN models developed with multiple climatic variables needs to be investigated for practical use of the model output. Most of the previous research used only two variables to develop a particular linear model. However, in this study more than two variables were considered to develop a particular linear model. Consequently, it can be noted that if number of variables exceed two, the accuracy of the model can increase significantly. As a result, if the number of variables is more than two, three is a possibility that the prediction capability of linear model will be better than non-linear model. Accuracy of the ANN prediction depends on number of hidden layers; hence, in this research the best outcome was considered from 20 different hidden layers.

The main objective of this study is to investigate the efficacy of linear LMR and non-linear ANN modelling techniques using multiple climate variables and indices for spring rainfall forecasting in Western Australia. The multiple climate indices used in this study include Southern Oscillation Index (SOI), Niño4, Niño3, Niño3.4 and Dipole Mode Index (DMI) for monitoring ENSO and IOD respectively. These indices were used as predictors of monthly spring rainfall in LMR and ANN modelling.

2 Study area

Western Australia, the largest state of Australia, has been selected as the study area. It is bounded by the Indian Ocean to the north and west, and the Southern Ocean to the south. Perth is the fastest growing capital of Western Australia. Western Australia is vulnerable to several climatic catastrophes such as: floods, droughts, and bushfires. In the recent decades, agricultural water supplies have plummeted in South-West of Western Australia; moreover, the frequency of the most intense cyclones have augmented in Western Australia (DPIRD 2020). Significant increase in frequency of extreme fire events have been observed at Perth, Kalgoorlie, and Broome, respectively.

This study selects seven rainfall stations located across Western Australia (WA) as shown in Fig. 1. The selection of these stations was based on the data availability and their spatial distribution representing all climatic conditions in WA. The information of these rainfall stations is given in Table 1.

3 Methodology

3.1 Rainfall and climate indices data

To facilitate the construction and application of linear and non-linear models, two types of data (seasonal rainfall and climate indices) are required. Daily rainfall data (1965–2019) for the selected stations were collected from Australian Bureau of Meteorology (website: www.bom.gov.au/climate/data/). Monthly total rainfall was calculated from the daily data and the seasonal average was taken by averaging the monthly total of seasonal rainfall. The climate indices data were extracted from Climate Explorer website (http://climexp.knmi.nl). The climate indices can be explained by sea surface temperature and sea level pressure anomalies around the globe. They can be used to describe the state and changes in the global atmospheric phenomenon, e.g. seasonal rainfall. Statistical analysis (such as time series analysis; their averages, extremes, and trends) can be performed using the climate indices. Two different types of indicators represent the ENSO: Southern Oscillation Index (SOI) and sea surface temperature anomalies. Another climate index IOD is represented by Dipole Model Index (DMI).

All extracted data (seasonal rainfall, ENSO and IOD) were divided into two sets, for the construction and validation of LMR models. The LMR models were assembled using the data from 1965 to 2014. The performance and efficacy of the developed LMR models were tested using the data from 2015 to 2019. A brief description of climate indices used is given as follows:

(a) Southern Oscillation Index (SOI) is a measure of sea level pressure differences between Tahiti (149.23° W, 17.78° S) and Darwin (130.83° E, 12.45° S)

(b) Nino 3 is a measure of sea surface temperature anomalies which corresponds with the region (5° S–5° N, 150° W–90° W),

(c) Nino 4 corresponds with the region (5° S–5° N, 160° W–150° W)

(d) Nino3.4 corresponds with region (5° S–5° N, 170° W–120° W) in the equatorial Pacific Ocean.

(e) DMI is a measure of difference in the average sea surface temperature anomalies from the tropical Eastern Indian Ocean (10° S to Equator, 90° E–110° E) to the tropical Western Indian Ocean (10° S–10° N, 50° E–70° E).
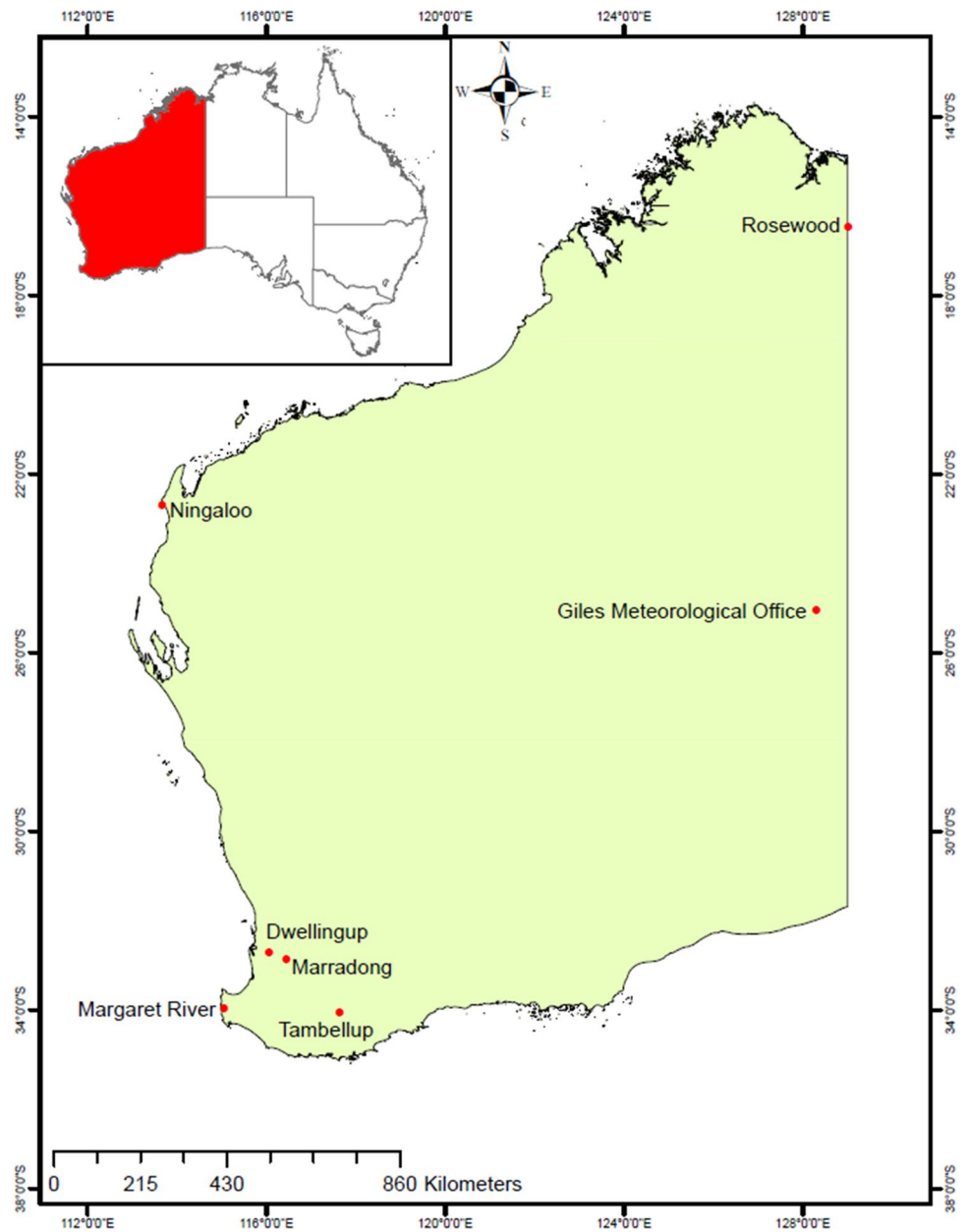
Fig. 1 Locations of the study area

Table 1 Specific information for the selected rainfall stations

Region	Station	Latitude	Longitude
Southern	Dwellingup	− 32.7103	116.0594
	Margaret River	− 33.9583	115.0636
	Marradong	− 32.8556	116.45
	Tambellup	− 34.0444	117.6419
Eastern	Giles Meteorological Office (GMO)	− 25.0341	128.301
Western	Ningaloo	− 22.6975	113.6744
Northern	Rosewood	− 16.46	129.01

3.2 Linear multiple regression modelling (LMR)

The LMR is a linear statistical modelling technique, which uses the least square method to find the best correlation between a variable (seasonal rainfall) and several other variables (climate indices). The general equation for LMR models can be expressed according to the following equation:

$$R_i = c_0 + a_1 X_{lag} + a_2 X_{lag} + e \qquad (1)$$

where R_i is the spring rainfall; X_{lag} and X_{lag} are the variables of the linear MR equations (ENSO and IOD in this study), a_1 and a_2 are the coefficients of the corresponding variables; c_0 is the constant and 'e' is the error of the linear MR analysis.

The effects of lagged climate indices were considered by adopting the spring rainfall of year 'n' and monthly (Dec$_{n-1}$, Jan$_n$ to Aug$_n$) values of the climate indices.

For any established model, evaluation is considered as an essential part to determine whether the initiative is worthwhile in terms of providing the anticipated outputs. The general tendency for the evaluation of empirical models is performed based on statistical correlation tests. In this study, the performances of the constructed LMR models were assessed by applying several error indices and statistical performance tests as given in Sect. 3.4.

3.3 Artificial neural network modelling (ANN)

Artificial neural network (ANN) is a statistical method designed to simulate the way biological human brain processes information. ANN is a data-driven mathematical model that was developed to imitate the structure of a human brain neural network and has been widely applied to solve problems such as prediction and discrimination (Mekanik et al. 2013). In general, an ANN refers to a multilayer perceptron structure. ANNs learn by detecting patterns and relationship within the provided input and desired output variables. ANN has been inspired by biological neural networks; it consists of simple neurons and connections that process information to find a relationship between inputs and outputs. The most common ANN architecture used by hydrologist is the multilayer perceptrons (MLP) which is a feedforward network that consists of three layers of neurons, the input layer, the hidden layers, and the output layer (Mekanik et al. 2013). The number of input and output neurons is based on the number of input and output data. The input layer only serves as receiving the input data for further processing in the network.

The non-linear ANN modelling technique was applied to predict the long-term seasonal rainfall considering the climate indices as the probable predictors. The same climate indices which were used to develop the LMR model were considered to develop the ANN model. The computer program MATLAB was used for the development of the ANN model. No standard method has been discovered yet to determine the number of nodes in the hidden layer of ANN modelling technique. As a result, common practice of hidden layer node detection is the trial-and-error method, which is also applied in this research. The ANN model has the danger to learn from the random fluctuation in the training data, which is called overfitting. Consequently, the performance of the network becomes very poor for unseen data during the testing phase. To prevent the overfitting problem in this research, early stop technique was employed which stop the training when the error in the testing data sets start to increase even though the error in the training continue to decrease.

For the application of ANN modelling approach, two activation functions are required for hidden layer and output layer, respectively. In this research, non-linear tan-sigmoid activation function is used for hidden layer and linear purelin function is applied for the output layer as recommended by Maier and Dandy (2001). These functions can be expressed according to the following equations for tan-sigmoid and purelin functions, respectively:

$$f_1 = \frac{2}{1 + e^{-2Y}} \tag{2}$$

$$f_2 = Y \tag{3}$$

where Y is the sum of all the inputs coming to the neuron.

3.4 Goodness of fit tests

To check the efficiency of the adopted linear and non-linear modelling techniques, the outputs of the developed models are assessed using statistical methods, such as Pearson correlation coefficients (R), root mean square error (RMSE), mean absolute error (MAE) and Willmott index of agreement (d):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_{\text{obs},i} - P_{\text{pred},i})^2}{n}} \tag{4}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|P_{\text{pred},i} - P_{\text{obs},i}\right|}{n} \tag{5}$$

$$d = 1 - \frac{\sum_{i=1}^{n}\left(P_{\text{obs},i} - P_{\text{pred},i}\right)^2}{\sum_{i=1}^{n}(\left|P_{\text{pred},i} - \overline{P_{\text{obs}}}\right| + \left|P_{\text{obs},i} - \overline{P_{\text{obs}}}\right|)^2} \tag{6}$$

where $P_{\text{obs},i}$ is the observed rainfall, $P_{\text{pred},i}$ is the predicted rainfall, $\overline{P_{\text{obs}}}$ is the mean observed rainfall, and n is the number of observation.

4 Results and discussion

4.1 Relationship between lagged climate indices and spring rainfall

To forecast West Australian seasonal rainfall, the ability of ENSO and IOD has been investigated in this study. It is apposite to investigate the influences of IOD and ENSO on Western Australian spring rainfall, using both linear as well non-linear modelling approaches. This is because of 2–3 months average values of climate indices having the highest correlation with Austral spring rainfall (Chiew et al.

1998). For this, it was decided to analyse the individual correlation between spring rainfall and monthly climate indices, e.g. DMI, Southern Oscillation index (SOI), and Nino3.4, respectively. The rainfall values which have statistically significant correlations with climate indices were further analysed using linear LMR modelling technique. The results of the correlation analysis are shown in Table 2. The outcomes of the correlation analysis reveal that only 4 months (May, June, July and August) of SOI and Nino3.4 have significant correlations with spring rainfall. Nonetheless, DMI values in April, May, June, July, August, and December have significant correlations with spring rainfall.

Table 2 illustrates that Western Australian spring rainfall significantly affected by DMI, SOI and Nino3.4 for most of the rainfall stations. In particular, for Rosewood station, the maximum significant correlation between spring rainfall and SOI was found 0.50 in July, whereas for the same station maximum significant correlation between spring rainfall and Nino 3.4 was found 0.50 in August. Although, Tambellup's spring rainfall is significantly affected by DMI and SOI,

there is no statistically significant correlation with Nino3.4. The maximum significant correlation between the rainfall and DMI was observed in August for the five rainfall stations namely, Dwellingup, Margaret River, Marradong, Tambellup, and Rosewood. The analysis verifies the intricate nature of atmospheric rainfall formation in West Australian rainfall stations; therefore, it can be postulated that single climatic driver is not effectual to predict long-term seasonal rainfall.

Since both, the ENSO and IOD have intense influence on Western Australia, combined effects of the indices DMI, SOI and Nino 3.4 were investigated. To assess this combined effect, the climate drivers with significant correlated months were further organised to apply LMR and ANN techniques. IOD-ENSO input sets were analysed as potential predictors of Western Australian spring rainfall for all seven rainfall stations, and the potential combined models are illustrated in Table 3. It should be noted that the climate indices, which have the significant correlation with the spring rainfall were considered to develop the models. Hence, input data sets of climate indices, $DMI_{Apr–Dec}$, $Nino3.4_{May–Aug}$, and $SOI_{May–Aug}$

Table 2 Correlation between lagged climate indices and spring rainfall

Lagged climate indices	Stations						
	Ningaloo	Dwellingup	Margaret River	Marradong	Tambellup	GMO	Rosewood
$DMI_{(Apr)}$	–	–	–	–	–	0.30^a	–
$DMI_{(May)}$	–	–	–	0.29^a	–	–	–
$DMI_{(Jun)}$	–	–	0.26^a	–	–	0.26^a	–
$DMI_{(July)}$	–	–	–	-0.29^a	–	–	–
$DMI_{(Aug)}$	–	-0.30^a	-0.29^a	-0.38^b	-0.32^a	–	-0.33^a
$DMI_{(Dec)}$	-0.28^a	–	–	–	–	–	–
$Nino3.4_{(Apr)}$	-0.27^a	-0.31^a	-0.31^a	–	–	–	–
$Nino3.4_{(May)}$	0.29^a	-0.29^a	-0.33^a	–	–	–	–
$Nino3.4_{(Jun)}$	-0.33^a	-0.26^a	-0.33^a	–	–	-0.29^a	-0.41^b
$Nino3.4_{(July)}$	-0.29^a	–	-0.29^a	-0.27^a	–	-0.27^a	-0.47^b
$Nino3.4_{(Aug)}$	–	–	-0.27^a	-0.31^a	–	-0.28^a	0.50^b
$SOI_{(May)}$	–	0.29^a	0.32^a	–	–	–	–
$SOI_{(Jun)}$	–	0.27^a	0.31^a	–	0.26^a	–	0.39^b
$SOI_{(July)}$	–	–	–	–	–	0.35^b	0.50^b
$SOI_{(Aug)}$	–	–	–	–	–	–	0.49^b

[a] Correlation is significant at the 0.05 level (2-tailed)

[b] Correlation is significant at the 0.01 level (2-tailed)

Table 3 Climate indices months considered to develop linear and non-linear models

DMI	Nino3.4				SOI			
Apr	May	Jun	Jul	Aug	May	Jun	Jul	Aug
May	May	Jun	Jul	Aug	–	Jun	Jul	Aug
Jun	May	Jun	Jul	Aug	–	Jun	Jul	Aug
Jul	May	Jun	Jul	–	May	–	–	–
Aug	May	–	–	–	–	–	–	–
Dec	May	Jun	Jul	Aug	May	–	Jul	–

were considered to develop both LMR and ANN models for all seven selected rainfall stations.

4.2 Goodness of fit tests

As mentioned earlier, the statistical performances of the developed LMR as well as ANN models were assessed with statistical errors such as RMSE and MAE. In addition, the index of agreement '*d*' was also used to assess the capability of LMR and ANN model to fit the observation. Table 4 illustrates the goodness of fit tests results during calibration (1965–2014) and testing period (2015–2019) for the developed LMR and ANN models for all stations, respectively. In general, the RMSEs were found lower for LMR technique compared to ANN models; therefore, the linear modelling technique seems to be more accurate than the ANN model in predicting the long-term seasonal rainfall of WA. It is evident from the results that the RMSE of the constructed linear LMR models are reasonably lower during the calibration period for all stations, although, RMSE of the testing period is considerably higher for Margaret river and Rosewood station. The results of MAE of the constructed models for LMR and ANN for calibration and validation shows that the higher MAE values were observed during the calibration period for the two stations (Marradong and Rosewood) for LMR model; although, during testing period, higher MAE values were observed for Margaret river and Rosewood station. Nevertheless, for the remaining stations, higher MAE values were found for ANN models both for calibration and testing periods. Hence, it can be postulated that both MAE and RMSEs are generally lower for LMR compared to ANN models, in case of West Australian spring rainfall prediction.

The precision of the developed linear and non-linear models can be evaluated through a statistical test, index of agreement '*d*' (Wilmott 1984), which demonstrates the efficacy of the models to fit the observations. Table 4 highlights that IOD-ENSO-based ANN models possess better agreement compared to LMR models during the calibration period, having '*d*' closer to 1 for most of the stations. In addition, during the validation period for GMO and Tambellup stations, '*d*' value has significantly plummeted, while using LMR model. In general, for both the calibration and testing periods the non-linear ANN models performed better than LMR models based on '*d*' values. The above stated information verifies that the LMR model may not be effectual to forecast the rainfall with utmost exactitude for all stations; hence, identification of an individual model to forecast rainfall, which is a complex global phenomenon is not pragmatic. As a result, in this study, it was decided to carry out the selection of the best predicted models, based on comparatively lower MAE and RMSE as well as '*d*' value relatively close to 1. It is important to note that the results obtained from this study

Table 4 Goodness of fit tests for the developed models during calibration and testing period

Stations	Goodness of fit test	Models	Calibration	Testing
Ningaloo	RMSE	LMR	3.88	2.25
		ANN	4.38	3.81
	MAE	LMR	2.94	1.29
		ANN	3.58	2.79
	d	LMR	0.60	0.42
		ANN	0.70	0.60
Dwellingup	RMSE	LMR	25.36	31.13
		ANN	49.23	30.15
	MAE	LMR	19.13	28.94
		ANN	35.57	29.43
	d	LMR	0.47	0.37
		ANN	0.17	0.38
Margaret River	RMSE	LMR	16.87	24.41
		ANN	13.38	13.89
	MAE	LMR	13.97	20.63
		ANN	20.63	12.84
	d	LMR	0.59	0.41
		ANN	0.58	0.95
Marradong	RMSE	LMR	12.57	18.30
		ANN	5.17	24.2
	MAE	LMR	9.92	16.92
		ANN	3.08	20.64
	d	LMR	0.53	0.47
		ANN	0.96	0.61
Tambellup	RMSE	LMR	12.05	14.26
		ANN	14.6	19.55
	MAE	LMR	9.69	12.867
		ANN	10.42	17.45
	d	LMR	0.61	0.26
		ANN	0.57	0.54
GMO	RMSE	LMR	13.44	14.51
		ANN	14.94	16.71
	MAE	LMR	11.02	11.05
		ANN	11.97	13.28
	d	LMR	0.66	0.26
		ANN	0.74	0.67
Rosewood	RMSE	LMR	15.46	23.02
		ANN	11.19	12.74
	MAE	LMR	11.86	20.49
		ANN	8.71	11.85
	d	LMR	0.77	0.42
		ANN	0.9	0.76

are not consistent with the similar studies carried out for different parts of Australia by Rasel et al. (2015), Hossain et al. (2018b), Hossain et al. (2020b) and Mekanik et al. (2013). The contrasting output in this present study is due

to incorporation of more critical variables for LMR model, unlike the above-mentioned studies.

Table 5 reveals the percentage improved by LMR models compared to ANN models; therefore, positive percentage in the table represents the prediction improvement over ANN models. For most of the stations, the linear MLR models performed better than the non-linear ANN models based on RMSE and MAE, although ANN models performed exceedingly well based on index of agreement '*d*' values.

4.3 Relationship between LMR and ANN outputs with observed rainfall

Comparisons of the linear MR modelling and ANN outputs with observed rainfall are plotted during the calibration period and shown in Figs. 2, 3 and 4. Figure 2 delineates the comparison between the observed rainfall and predicted rainfall obtained for different stations in four regions, using both LMR and ANN models. Compared to ANN model, LMR model could re-produce the observed rainfall with reasonable accuracy for most of the stations. The LMR model has demonstrated considerable deftness to predict the extreme rainfall during the calibration period. Figure 3 depicts the comparison between the observed rainfall

Table 5 Percentage improved by LMR models compared to ANN models

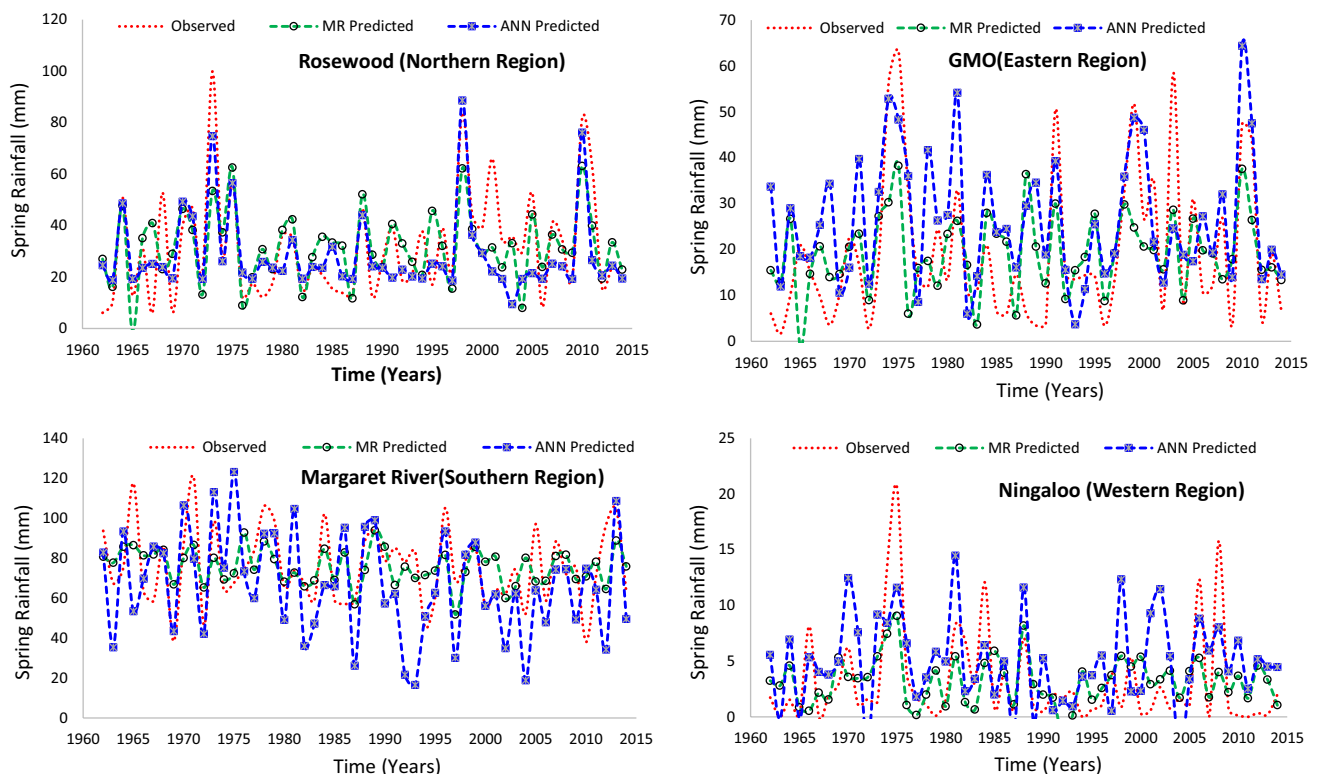| Stations	RMSE		MAE		d	
	Calibration	Testing	Calibration	Testing	Calibration	Testing
Ningaloo	11.4	41.0	17.9	53.7	− 14.7	− 30.5
Dwellingup	48.5	− 3.2	46.2	1.7	175.3	− 0.5
Margaret River	− 26.1	− 75.7	32.3	− 60.7	1.0	− 56.7
Marradong	− 143.0	24.4	− 222.0	18.0	− 45.2	− 23.8
Tambellup	17.5	27.0	7.0	26.3	6.7	− 52.4
GMO	10.1	13.1	7.9	16.8	− 11.5	− 61.5
Rosewood	− 38.2	− 80.7	− 36.2	− 72.9	− 14.1	− 44.6

Fig. 2 Comparison of LMR modelling with ANN for the selected stations from different regions during calibration
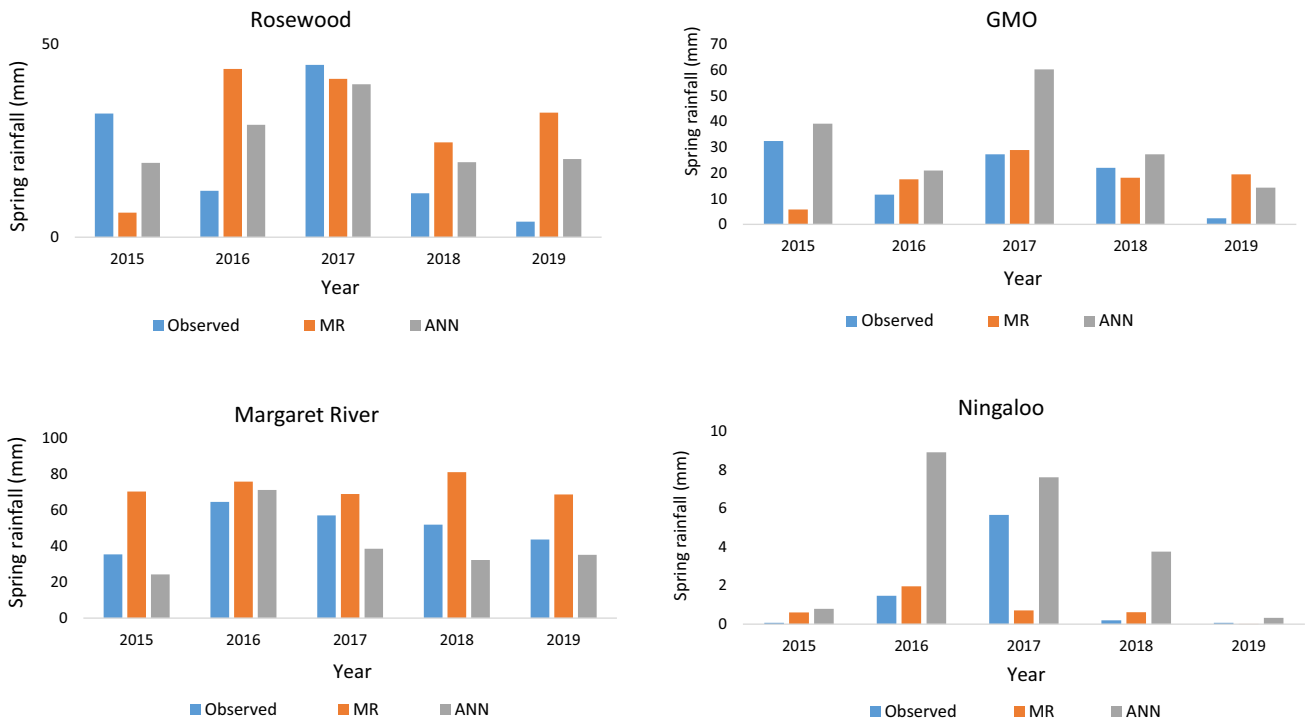
Fig. 3 Comparison of the models' outputs with the observed rainfalls for the selected stations from different regions during the testing period
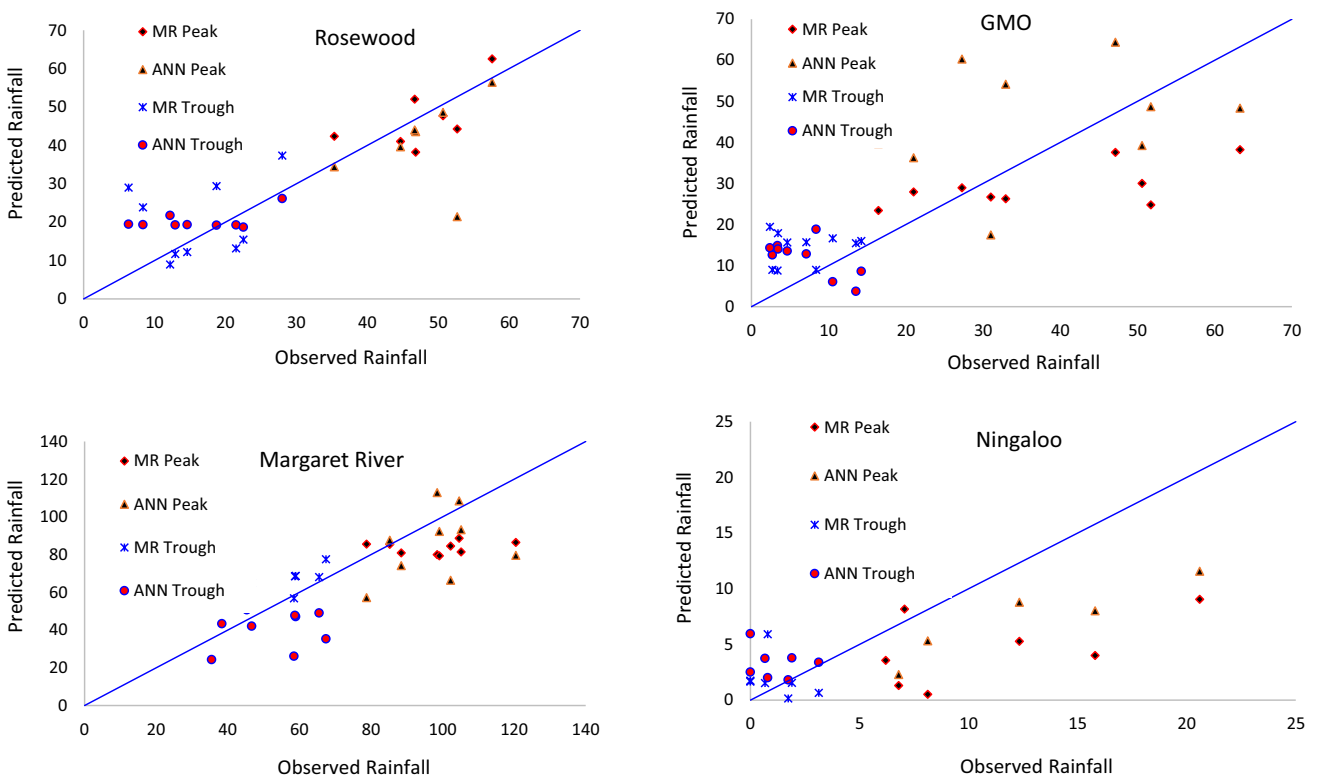

Fig. 4 Comparison of the ANN modelling outputs with MLR models outputs in regards to peaks and troughs for the selected stations

and predicted rainfall obtained for the selected stations, using both LMR and ANN models for the testing period (2015–2019). Although, on some occasions, LMR model outputs slightly deviate from the actual observed rainfall, overall, LMR model performs relatively well during testing phase compared to ANN model for several stations. The understandable reason behind these divergences of both LMR and ANN model on rainfall prediction is due to existence of climatic drivers as well as local climatic condition of Western Australia. The predictive capability of the models (both LMR and ANN) was also assessed in producing the rainfall peaks and troughs throughout the time series periods. Hence, additional information regarding developed LMR and ANN models can be extracted from the peaks and troughs; thus, plotted results of the peaks and troughs are illustrated in Fig. 4. In general, for all the four regions, LMR model was able to forecast the troughs and peaks with reasonable accuracy compared to ANN model.

5 Conclusions and recommendations

In this study, artificial neural network and LMR techniques were used to forecast monthly Spring rainfall in seven different rainfall stations spreading all around Western Australia. Climate indices representing ENSO and IOD, namely DMI, SOI, and Nino3.4 were used as predictors. Nino3.4 and Southern Oscillation Index (SOI) were used as ENSO indicators and Dipole Mode Index (DMI) was chosen as IOD indicator. It is important to note that the predictors were lagged by 4–6 months to provide prediction lead time. Both ANN and LMR models performed with some level of precision to predict the monthly rainfall for Western Australia.

The correlation coefficients of past values of the climate indices with spring rainfalls for the seven stations were determined. As mentioned earlier, climate indices, which have the significant correlation with the spring rainfall were considered to develop both LMR and ANN models. The outcomes of correlation analysis disclose that only 4 months (May, June, July and August) of SOI and Nino3.4 have significant correlations with spring rainfall; however, DMI values in April, May, June, July, August, and December have significant correlations with spring rainfall. Both the LMR and ANN models were analysed to explore the predictive potential of seasonal rainfall; subsequently, manifold statistical evaluation parameters namely: RMSE, MAE and d were used to determine the efficacy of these two models. In general, the RMSEs of the analysis disclose that LMR model is more effectual compared to the ANN models in predicting the long-term seasonal rainfall of WA. RMSE of the constructed LMR models are reasonably low during the calibration period for the study area, although, considerably higher RMSEs were observed for Margaret River and

Rosewood station during the testing period. Nonetheless, for most of the stations except Marradong and Rosewood higher MAE values were observed for ANN models in comparison with LMR models. ANN models possess better agreement compared to LMR models during the calibration period, having 'd' closer to 1 for most of the stations; similarly, during the validation period for Giles Meteorological Office (GMO) and Tambellup stations, 'd' value has significantly plummeted, while using LMR model. Based on the above discussion, it can be concluded that the errors (RMSE and MAE) are generally lower for LMR compared to ANN models, although, during the calibration and testing periods, ANN models performed better than LMR models based on 'd' values.

Overall, the constructed LMR models are suitable for most of the stations while predicting the extreme rainfall with reasonable exactitude, compared to ANN models. Due to considerable spatial and temporal variability of rainfall in Western Australia, further investigation using both linear and non-linear modelling techniques is required to recommend a generalised model for seasonal rainfall prediction. To recapitulate, this study revealed the possibility of seasonal rainfall forecasting using ANN as well as LMR models for the study area. Immaculate prediction of the spring rainfall in Western Australia will facilitate water resources management to safeguard flood mitigation and provide adequate strategies to withstand potential droughts. Hence, the developed LMR model can be considered as an effectual alternative, in addition to the prevalent physically based forecasting models.

Declarations

Conflict of interest Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abbot J, Marohasy J (2012) Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Adv Atmos Sci 29:717–730

Abbot J, Marohasy J (2017) Application of artificial neural networks to forecasting monthly rainfall one year in advance for locations within the Murray Darling basin, Australia. Int J Sustain Dev Plan 12:1282–98

Acharya N, Kar S, Kulkarni M, Sahoo L (2011) Multi-model ensemble schemes for predicting northeast monsoon rainfall over peninsular India. J Earth Syst Sci 120(5):795–805

Acharya N, Shrivastava NA, Panigrahi BK, Mohanty UC (2013) Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Clim Dyn 43(5):1303–1310

Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and non-linear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour Res 48:25

Bilgili M, Sahin B (2010) Prediction of long-term monthly temperature and rainfall in Turkey. Energy Sour Part A 32:60–71

Chattopadhyay S (2007) Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India. Acta Geophys 55(3):369–382

Chattopadhyay S, Chattopadhyay G (2009) Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. CR Geosci 342:100–107

Chattopadhyay S, Chattopadhyay G (2018) Conjugate gradient descent learned ANN for Indian summer monsoon rainfall and efficiency assessment through Shannon-Fano coding. J Atmos Solar Terr Phys 179:202–205

Chakraverty S, Gupta P (2008) Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Comput Appl 17:187–192

Chiew F, Piechota TC, Dracup J, McMahon T (1998) El Nino/Southern oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. J Hydrol 24:138–149

Djibo A, Karambiri H, Seidou O, Sittichok K, Philippon N, Paturel J, Saley HJC (2015) Linear and non-linear approaches for statistical seasonal rainfall forecast in the Sirba watershed region (Sahel). Climate 3:727–752

DPIRD (2020) Cliamte trends in Western Australia. https://www.agric.wa.gov.au/climate-change/climate-trends-western-australia. Accessed on 18 Sept 2020

Hossain I, Esha R, Alam Imteaz M (2018a) An attempt to use non-linear regression modelling technique in long-term seasonal rainfall forecasting for Australian Capital Territory. Geosciences 8:282

Hossain I, Rasel H, Imteaz MA, Mekanik F (2018b) Long-term seasonal rainfall forecasting: efficiency of linear modelling technique. Environ Earth Sci 77:280

Hossain I, Rasel HM, Imteaz MA, Mekanik F (2020a) Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorol Atmos Phys 132:131–141

Hossain I, Rasel H, Imteaz MA, Mekanik F (2020b) Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorol Atmos Phys 132:131–141

Islam SA, Bari MA, Anwar AHMF (2014) Hydrologic impact of climate change on Murray-Hotham catchment of Western Australia: a projection of rainfall–runoff for future water resources planning. Hydrol Earth Syst Sci 18:3591–3614

Khastagir A, Jayasuriya N, Bhuyian M (2017) Use of regionalisation approach to develop fire frequency curves for Victoria, Australia. Theoret Appl Climatol 134:849–858

Khastagir A, Jayasuriya N, Bhuyian MA (2018) Assessment of fire danger vulnerability using McArthur's forest and grass fire danger indices. Nat Hazards 94:1277–1291

Khastagir A, Jayasuriya N (2008) Role of rainwater tanks in managing demand during droughts. In: Corinne C (ed) Enviro 08 Proceedings, Australia, 05–07 May 2008

Khastagir A (2008) Optimal use of rainwater tanks to minimize residential water consumption, Masters by Research, Civil, Environmental and Chemical Engineering, RMIT University

Kirono DG, Kent DM (2011) Assessment of rainfall and potential evaporation from global climate models and its implications for Australian regional drought projection. Int J Climatol 31:1295–1308

Maier HR, Dandy GC (2001) Neural network based modelling of environmental variables: a systematic approach. Math Comput Modell 33:669–682

Mekanik F, Imteaz MA, Gato-Trinidad S, Elmahdi A (2013) Multiple linear regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. J Hydrol 503:11–21

Pal S, Dutta S, Nasrin T, Chattopadhyay S (2020) Hurst exponent approach through rescaled range analysis to study the time series of summer monsoon rainfall over northeast India. Theoret Appl Climatol 142:581–587

Rasel HM, Imteaz MA, Mekanik F (2016a) Investigating the Influence of remote climate drivers as the predictors in forecasting South Australian spring rainfall. Int J Environ Res 10:1–12

Rasel, HM, Imteaz, MA, Mekanik, F (2015) Evaluating the effects of lagged ENSO and SAM as potential predictors for long-term rainfall forecasting. In: Proceedings of the international conference on water resources and environment (WRE 2015), Beijing, China, 25–28 July, Taylor & Francis Group, London/Miklas Scholz (ed), pp 125–129

Rasel H, Esha RI, Imteaz MA, Klaas D (2016b) Long-term rainfall prediction using large scale climate variables through linear and non-linear methods. In: 37th hydrology and water resources symposium 2016b: water, infrastructure and the environment, 2016b. Engineers Australia, pp 236

Saha S, Chattopadhyay S (2020) Exploring of the summer monsoon rainfall around the Himalayas in time domain through maximization of Shannon entropy. Theoret Appl Climatol 141:133–141

Shams MS, Lamb AHMF, Bari K (2018) Relating ocean-atmospheric climate indices with Australian river streamflow. J Hydrol 556:294–309

Willmott CJ (1984) On the validation of models. Phys Geogr 2:184–194

Yilmaz AG, Hossain I, Perera BJC (2014) Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne. Hydrol Earth Syst Sci 18:4065–4076