



Prediction of Seasonal Rainfall with One-year Lead Time Using Climate Indices: A Wavelet Neural Network Scheme

Meysam Ghamariadyan¹ · Monzur A. Imteaz¹

Received: 15 September 2020 / Accepted: 16 October 2021 / Published online: 20 October 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

This paper presents the development of the Wavelet Artificial Neural Networks (WANN) model to forecast seasonal rainfall in Queensland, Australia, using the Inter-decadal Pacific Oscillation (IPO), Southern Oscillation Index (SOI), and Nino3.4 climate indices as predictors. Eight input sets with different combinations of predictive variables from 1908 to 2016 were considered to develop forecast models for ten selected rainfall stations in Queensland, Australia. The outcomes of WANN modeling are compared with Artificial Neural Networks (ANN). Moreover, the skillfulness of the WANN in comparison to the current climate prediction system used by the Australian Community Climate Earth-System Simulator–Seasonal (ACCESS–S) and climatology forecasts are investigated. Besides, the WANN predictions are compared with two other conventional approaches like autoregressive integrated moving average (ARIMA) and multiple linear regression (MLR) for further investigations. The comparisons indicated that the WANN achieves the lower average root mean square error (RMSE) in all the stations with 112.2mm compared to ANN with 178.9mm, ACCESS-S with 281.8mm, climatology prediction with 279.7mm, MLR with 195.1mm, and ARIMA with 187.7mm. The WANN seasonal rainfall forecasts are more accurate than the ANN, ACCESS-S, Climatology, MLR, and ARIMA by 37%, 60%, 53%, 42%, and 40%, respectively. It was also found that the ACCESS-S underestimates the extreme seasonal rainfall during the testing period up to 80%, while it is limited to 21% for the WANN among the selected stations. The results show that the WANN model outperforms the MLR, ARIMA, climatology, ACCESS-S, and ANN forecasts in all the selected stations.

Keywords Climate indices · Seasonal rainfall forecasting · Wavelet artificial neural network

1 Introduction

Forecasting seasonal rainfall a few months in advance is potentially valuable for a vast range of decision-makers and users, including irrigators and urban and rural water managers, to devise risk management strategies. It is also essential for the regions with high

✉ Monzur A. Imteaz
mimteaz@swin.edu.au

¹ Department of Civil and Construction Engineering, Swinburne University of Technology, Melbourne, VIC 3122, Australia

spatial and temporal rainfall variability. However, forecasting rainfall is complex due to participating in many variables such as temperature, pressure, climate drivers, etc. In the past decades, Artificial Intelligence (AI) based methods were introduced to overcome the complexity of rainfall forecasting. ANNs are one of the AI-based methods that have been broadly used for forecasting rainfall (Pakdaman et al. 2020; Poonia and Tiwari 2020; Yadav and Sagar 2019). For example, Babel et al. (2015) developed three ANN models with various explanatory variables for forecasting rainfall in Mumbai, India. ANN was used for post-processing the monthly precipitation forecast by Pakdaman et al. (2020). They showed that the performance of ANN is superior compared to the multi-model ensemble (NMME) models for all months.

However, sometimes, the ANN is integrated with other techniques to improve the results and accuracy of the models. There have been many attempts to improve rainfall prediction by developing novel hybrid techniques. For example, recently, a novel post-processing technique was proposed by Yazdandoost et al. (2020) that can improve the conventional method's performance over the monthly and seasonal predictions. There is another successful effort in using hybrid techniques in forecasting rainfall. A new hybrid ANN was developed to improve the performance of the ANN in forecasting rainfall by Abdel-Kader et al. (2021). A Hybrid Adaptive Neuro-Fuzzy Inference System was suggested to evaluate temporal rainfall variability by Farrokhi et al. (2020). Diop et al. (2020) established a hybrid Multilayer Perceptron-Whale Optimization Algorithm (MLP-WOA) model for forecasting annual rainfall. They found that the proposed method improves the performance of the Multilayer Perceptron (MLP).

The wavelet transform is one of the techniques that help researchers to investigate climatological events precisely and help them have better insight into these phenomena. This technique can capture useful hidden information of weather parameters such as rainfall and climate drivers, which cannot be easily seen in the main time series. He et al. (2015) conducted appreciable research to forecast monthly rainfall in Australia using a hybrid wavelet ANN. Their study included the processing and input selection methodologies. Sharghi et al. (2018) applied a wavelet ANN technique for forecasting daily and monthly rainfall in two watersheds. They showed that the WANN could enhance the prediction of the simple feed-forward neural network up to 50%. A model using the integration of wavelet transform and convolutional neural network (CNN) was developed for forecasting monthly and daily rainfall by Chong et al. (2020). They showed that their proposed model captures patterns of the rainfall time series satisfactorily. Recently, Ghamariadyan and Imteaz (2021a) developed a wavelet-ANN model for forecasting monthly rainfall in Queensland, Australia, with up to twelve-month lead time.

The Australian climate is variable compared to other parts of the world (Bagirov and Mahmood 2018). It has been changing in recent decades, and researches indicate that it has undergone higher temperature and less rainfall. The El Niño–Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) are the most significant anomalies that cause rainfall variability in Australia (Hasan and Dunn 2012). The ENSO is mainly measured using the southern oscillation index (SOI) (Abtey and Trimble 2010). In addition, IPO is another influential index that is often defined as exhibiting an “ENSO-like” decadal pattern (Power et al. 1999). The IPO is measured by variations in the sea surface temperature in the Pacific Ocean.

There are few studies on combined and lagged effects of climate indices on rainfall (Ghamariadyan and Imteaz 2021b; Mekanik et al. 2013; Schepen et al. 2012). It is noteworthy that understanding the relationships between rainfall and different climate indices can help to have a better insight to predict future rainfall.

The contribution of this study is to develop a coupled wavelet-ANN for predicting seasonal rainfall in Queensland. Also, results are compared with the conventional methods (such as MLR, ARIMA), the current prediction system of the Bureau of Meteorology (BOM) of Australia, and climatology forecasts.

2 Methodology

A summary of the procedure used in the study is shown in Fig. 1. A correlation analysis is conducted in the first step, and then the best input set is selected according to the correlation analysis. Next, the various models are developed, and the best input set for each model is identified based on the RMSE value. In the last step, the performance of WANN is compared with other methods using various evaluation criteria.

The methodology for forecasting seasonal rainfall in this study is mainly based on the WANN and ANN methods. To date, seasonal rainfall forecasting in Australia has been investigated for some regions (Mekanik et al. 2016; Tozer et al. 2017). Nevertheless, no study has been conducted to predict the seasonal rainfall in Queensland. Besides, there is no versatile model to forecast the seasonal rainfall one year ahead with reasonable accuracy.

2.1 Wavelet Analysis

The concept of the wavelet theorem, primarily, was introduced as a substitute for the Fourier transforms (FT). In wavelet analysis, the primary signal is split into various sub-signals with lower resolution using shifting and scaling processes. The wavelet is shifted to align with the feature we are looking for in the signal. Then, the similarity of the shifted or stretched signal through the scaling process is measured using a factor called the wavelet coefficient (Mallat 1999).

It is noteworthy that selecting the appropriate wavelet function (mother wavelet) affects the precision of the results. The proper wavelet function is acquired through the trial and

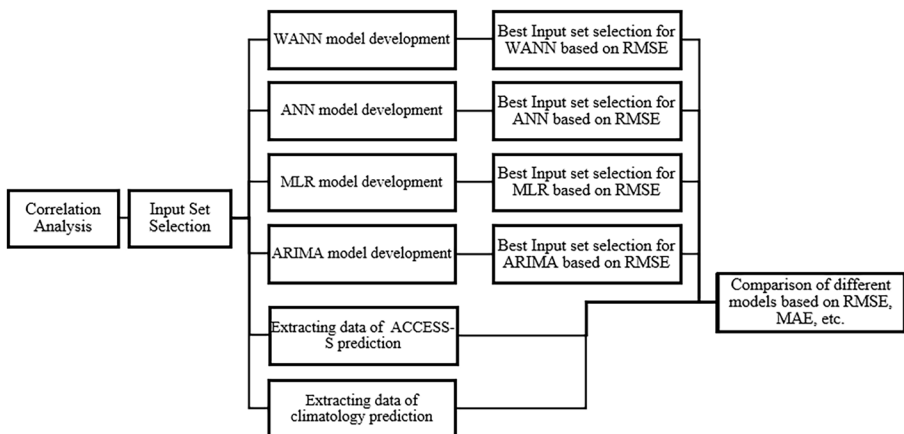


Fig. 1 Schematic diagram of different models development and comparison

error process (Nourani et al. 2012). In the current study, Dmey (Discrete Meyer) was chosen as the main wavelet function.

In wavelet transform analysis, the signal is converted from the time domain into the time/frequency domain. It is mathematically defined as (Mallat 1999):

$$f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \varphi\left(\frac{t-b}{a}\right) dt \quad (1)$$

where, $f(t)$ stands for the main signal, $\varphi(t)$ is wavelet function with length (t), a and b are scale factors, and time-like translation variables, correspondingly. In the high-scale band, the wavelet is greatly stretched, while the wavelet is immensely compacted in the low-scale band.

The approximation and detail components are obtained after decomposition of the main signal at different levels (identifying levels number is based on trial and error) and are defined by

$$f(t) = D_1 + D_2 + D_3 + \dots + D_i + A_i \quad (2)$$

where, A and D stand for “approximation” and “detail”, respectively. i shows the level of decomposition. In this definition, the low-frequency signal (A) consists of a high-frequency and low-frequency signal. In other words, A_1 is a low-frequency signal that can create A_2 (a low-frequency signal) and D_2 (a high-frequency signal).

2.2 Artificial Neural Networks

ANNs are computing systems that try to imitate human brain properties by learning and training rules. An ANN comprises nodes connected by some weighted synaptic connections. An MPL neural network contains an input and output layer with one or more hidden layers. The hidden nodes' size is obtained through trial and error, and it relies on the complexity of the problem. An MLP is mathematically defined as (Kim and Valdés 2003):

$$\hat{y}_k = f_0 \left[\sum_{j=1}^M w_{kj} \cdot f_h \left(\sum_{i=1}^N w_{ji} x_i + w_{j0} \right) + w_{k0} \right] \quad (3)$$

where, w_{ji} is the weights of the hidden layer that links the i^{th} neuron in the input layer with the j^{th} neuron in the hidden layer; f_0 stands for the output layer's activation function; w_{j0} is the bias for the j^{th} hidden neuron; w_{kj} denotes the weight of the output layer that connects j^{th} neuron with the k^{th} neuron in the output layer; w_{k0} is the bias for the k^{th} output neuron; f_h is the activation function of hidden neurons and \hat{y}_k is the neural network's output.

For the hidden layer, the tan-sigmoid transfer function, and for the output layer, the linear purlin (g) is used. The use of sigmoidal and linear form transfer functions for hidden and output layers is recommended, respectively (Maier and Dandy 2000).

In this study, the training algorithm is the Levenberg–Marquardt (LM) algorithm. The early stop technique is used to eliminate overtraining. In this process, the training is stopped when the validation error begins to increase while the training set error is still declining. Hence, by using this procedure, overfitting does not happen. Moreover, an out-of-sample data set is selected to evaluate the performance of the model accurately.

2.3 Integrating the Wavelet with ANN

For connecting the ANN with the wavelet, the approximations (A) and details (D) acquired through the wavelet analysis at each level are used as inputs of the ANN. That means that the time series of the approximation and detail at each level is used as the input of the ANN. For instance, through the wavelet analysis, if there is only one time series as input set (model 1) and three decomposition levels are used, then the six sub-time series is applied as the network's input.

Figure 2 shows the process of connecting the wavelet and ANN. In this figure, 'A' and 'D' stand for the approximation and detail at each level. Also, 'm' and 'n' are numbers of inputs and levels, correspondingly.

2.4 MLR and ARIMA

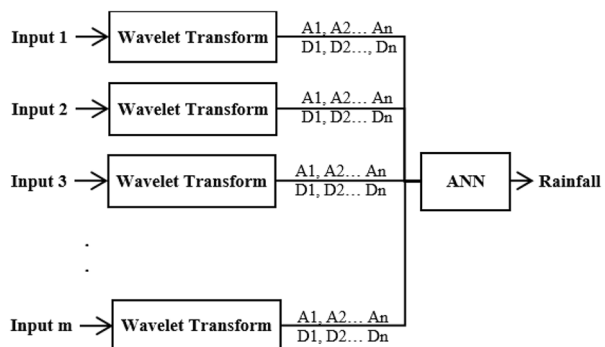
In this study, the classical methods of the ARIMA and MLR are also used for further comparison. The MLR is a linear method to create a correlation between the dependent and independent variables (Brown 2014). In MLR, the stepwise technique is used, and the F-test is employed to identify the best-fitted model (Galvao et al. 2008).

The ARIMA comprises autoregressive (AR) and moving average (MA) models (Box et al. 2015). This method is a widely used and popular model in time series prediction. However, sometimes identifying the typical inherent parameters [p, d, q] is not straightforward (Islam and Imteaz 2020). In addition, The ARIMA cannot be applied directly to the non-stationary time series, and it must be converted into a stationary time series by using the time series differencing technique. In the current study, since the time series is stationary, applying the differencing technique is unnecessary. The ARIMA modeling in this study is conducted using the SPSS software. The "expert modeler" is used as a feature to determining the order of the parameters and the best model. Thereby, the values of ARIMA parameters are different for each station. Further background and information about ARIMA are available in (Box et al. 2015).

3 Data and Study Area

Queensland, in the northeast of Australia, is the second-largest and third-most-populous state of Australia. The location of the study area, along with selected rainfall stations, is shown in Fig. 3. The data of rainfall and the SOI are obtained from the Australian BOM

Fig. 2 A multi-level Wavelet ANN architecture with various input sets



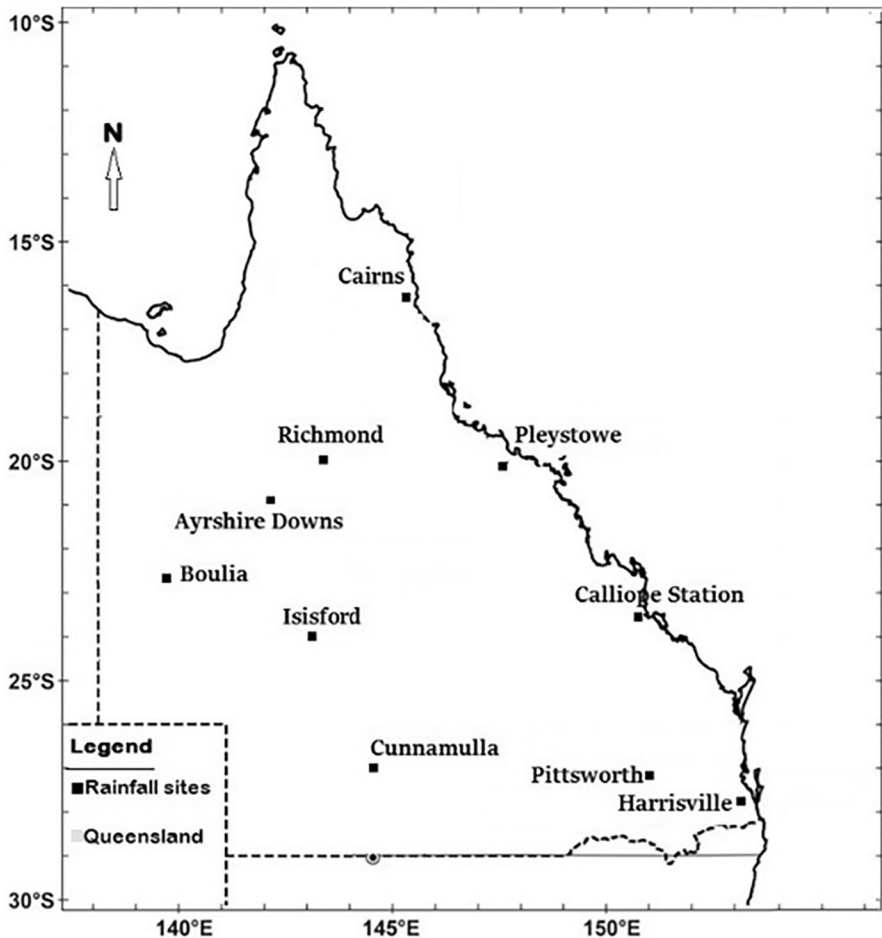


Fig. 3 Schematic map of Queensland State and weather stations

website (<http://www.bom.gov.au/climate/data/>). Ten weather stations with a long record of data were chosen. The details of the selected weather stations are shown in Table 1. The Nino3.4 data is collected from the Royal Netherlands Meteorological Institute (KNMI) Climate Explorer (<http://climexp.knmi.nl>). Also, the data related to IPO is obtained from Chris Folland, Met Office, Hadley Centre, UK (<https://www.metoffice.gov.uk>).

Queensland's rainfall is spatially and temporally variable (Knip et al. 2011). However, the wet season is considered from November to March in north Queensland (Sumner and Bonell 1988). In southeastern and central Queensland, wet and dry seasons are considered from October to March and September to April, respectively (Wilson et al. 2013). Moreover, most of the flood events in Queensland occur in summer. In this study, rainfall of summer (December to February) as the wet season is considered.

Total 109 years of data from 1908 to 2016 is used in this study. For training the models, data of the first 94 years (1908 to 2001) are used, and the remaining 15 years (2002–2016)

Table 1 Geographical location of selected rainfall sites of the study

Region	Site Number	Site Name	Latitude	Longitude	Summer Mean Rainfall (mm)
Tropical East Coast	31036	Cairns	16.82° S	145.64° E	1039.91
	39020	Calliope	24.02° S	150.97° E	417.57
	33060	Pleystowe	21.14° S	149.04° E	846.99
Central Interior	37001	Ayrshire Downs	21.97° S	142.72° E	229.01
	30045	Richmond	20.73° S	143.14° E	299.14
	38003	Boulia	22.91° S	139.90° E	125.77
	36026	Isisford	24.26° S	144.44° E	193.73
South East	40094	Harrisville	27.81° S	152.67° E	342.02
	44026	Cunnamulla	28.07° S	145.68° E	141.06
	41082	Pittsworth	27.72° S	151.63° E	266.57

are assigned for verifying the models to avoid any potential overfitting. In other words, the testing data set are new data that were not used previously during the training process.

In this study, eight different model sets comprising rainfall and climate indices are defined to identify the best model for each station (Table 2). Notably, using a single climate driver might be unsuitable for forecasting rainfall accurately because of the complications of the relationships between climate drivers and rainfall. Thereby, combined inputs comprised of various effective climate indices plus rainfall values are employed. Also, the individual input sets (Rainfall, IPO, SOI, and Nino3.4) are defined to investigate better the impact of a single predictor on the prediction of summer rainfall in Queensland.

The models' performances during calibration and testing are evaluated using statistical parameters such as RMSE, mean absolute error (MAE), refined index of agreement (d_r) and correlation coefficient (R). According to Willmott et al. (2012), this index is more efficient for evaluating the model's accuracy than other available statistical parameters. Moreover, the Nash-Sutcliffe coefficient of efficiency (NSE) is another practical factor that reasonably measures the hydrological models' prediction capabilities (Nash and Sutcliffe 1970). It is bounded between $-\infty$ to 1 and is presented to different forms of mathematical models (Gupta and Kling 2011).

For further assessment of the models, a statistical score is used to assess the skills of the predicted model against a reference forecast like climatology prediction. The skill score, which is used for ACCESS-S, WANN, and ANN is as follows (Hawthorne et al. 2013)

Table 2 The input sets defined for each station

NO	Model (input set)
1	Rainfall
2	Rainfall+Nino
3	Rainfall+SOI
4	Rainfall+IPO
5	IPO
6	SOI
7	Nino3.4
8	Rainfall+IPO+SOI+Nino

$$\text{Skillscore} = \frac{RMSE_{Ref.} - RMSE_{For.}}{RMSE_{Ref.}} * 100\% \quad (4)$$

where, $RMSE_{For.}$ and $RMSE_{Ref.}$ stand for score (RMSE) of the model forecast and a reference model (climatology), respectively. The 100% skill score shows a perfect prediction, while a zero skill score indicates that the skills of the forecasted model are the same as the climatology (reference forecast). A score skill less than zero denotes that the prediction skill score is worse than the climatology.

4 Results and Discussion

The correlation of summer rainfall with each monthly climate index is investigated to understand better the influence of an individual climate index on summer rainfall in each station. Inputs of the final models are selected based on their correlations with summer rainfall. For example, according to the correlation analysis for Cairns station, the summer rainfall is correlated with Nino3.4 from January to April (Table 3). For the remaining months, the correlation is insignificant. Therefore, the input sets comprised of Nino3.4 from January to April (four inputs for Nino3.4) are considered input values. It is apparent that correlation is insignificant between different monthly Nino3.4 with summer rainfall for Ayrshire Downs, Boulia, and Cunnamulla. However, Pittsworth's summer rainfall correlation with Nino3.4 covers a broader range of months compared to other stations, though the correlations are not too strong.

Moreover, the effect of SOI on the summer rainfall of Queensland is investigated in Table 3. It is evident that the summer rainfall is correlated with SOI in all stations, at least in one of the months. In Cairns, Richmond, Boulia, Isisford, and Cunnamulla, the summer rainfalls are correlated with SOI only in February.

Nonetheless, the correlation coefficients are significant for most of the months for Calliope and Pittsworth. The highest correlation value can be found in Pittsworth in April with 0.33. Similarly, the correlation analysis is done for IPO and summer rainfall in Table 3. It is evident that the correlations cover a broader range of months compared to Nino3.4 and SOI. It is seen that only in one station (Boulia), the summer rainfall is not correlated with IPO. However, for other stations, correlations are significant at least at three months. The observed summer rainfall with a one-year lead time is allocated to forecast the next year's seasonal rainfall for target values. In other words, the target values are considered as the next year's observed summer rainfall time series in the calibration process.

Various input sets are applied to the models to assess their performance. Table 4 compares the input matrix for ANN and WANN for Cairns station as a sample. The structure of each input set for each station depends on the correlation analysis and the number of levels in the WANN method. As in this study, one lag time is considered; two time series are generated (one main time series plus one lagged time series). After considering the lagged time series, the total number of time series applied to the network would be 12, with 93 values for each one during the calibration process. For instance, in model 8, 108 time series are generated through the wavelet analysis, while only 18 time series is applied to the ANN model. This technique provides very meticulous information for training the network, which is not identifiable by ANN solely.

Table 3 Pearson correlation coefficient of Nino 3.4, SOI, and IPO antecedents and summer rainfall

Month	Climate Index	Station									
		Cairns	Calliope	Pleystowe	Ayrshire Downs	Richmond	Boulia	Isisford	Harrisville	Cunnamulla	Pittsworth
Jan	Nino3.4	-0.21**	-	-	-	-0.21**	-	-	-	-	-0.21**
	SOI	-	-	0.19**	-	-	-	-	-	-	-
	IPO	-	-0.24**	-0.22**	-0.22**	-0.24**	-	-0.21**	-0.21**	-	-0.25*
Feb	Nino3.4	-0.22**	-	-	-	-0.19**	-	-	-	0.19**	-0.20**
	SOI	0.29*	-	0.21**	0.29*	0.31*	0.19**	0.22**	0.21**	0.22**	0.22**
	IPO	-0.19**	-	-0.27*	-0.22**	-0.26*	-	-0.21**	-0.20**	-	-0.23**
Mar	Nino3.4	-0.20**	-	-0.20**	-	-	-	-	-0.19**	-	-0.21**
	SOI	-	0.19**	-	-	-	-	-	0.31*	-	0.28*
	IPO	-0.22**	-	-0.29*	-0.24**	-0.27*	-	-0.19**	-	-	-0.22**
Apr	Nino3.4	-0.21**	-0.26*	-0.20**	-	-0.19**	-	-	-0.19**	-	-0.24**
	SOI	-	-	0.23**	-	-	-	-	0.23**	-	0.33*
	IPO	-0.20**	-0.25*	-0.32*	-0.21**	-0.26*	-	-0.20**	-0.22**	-	-0.24**
May	Nino3.4	-	-0.33*	-0.22**	-	-	-	-	-0.24**	-	-0.26*
	SOI	-	0.21**	0.19**	-	-	-	-	-	-	-
	IPO	-	-0.29*	-0.27*	-	-0.23**	-	-	-0.24**	-	-0.25*
Jun	Nino3.4	-	-0.27*	-	-	-	-	-	-	-	-0.24**
	SOI	-	0.21**	-	-	-	-	-	-	-	-
	IPO	-	-0.30*	-0.24**	-	-0.19**	-	-	-0.22**	-0.21**	-0.31*
Jul	Nino3.4	-	-0.24**	-	-	-	-	-	-0.20**	-	-0.19**
	SOI	-	0.22**	-	-	-	-	-	-	-	0.21**
	IPO	-	-0.30*	-0.23**	-	-	-	-	-0.26*	-	-0.26*
Aug	Nino3.4	-	-0.27*	-	-	-	-	-	-	-	-0.22**
	SOI	-	0.23**	0.24**	-	-	-	-	-	-	0.26*
	IPO	-	-0.28*	-	-	-	-	-	-0.20**	-	-0.23**

Table 3 (continued)

Month	Climate Index	Station										
		Cairns	Calliope	Pleystowe	Ayrshire Downs	Richmond	Boulia	Isisford	Harrisville	Cunnamulla	Pittsworth	
Sep	Nino3.4	-	-0.26*	-0.20**	-	-	-	-0.26*	-	-	-0.28*	
	SOI	-	0.22**	-	-	-	-	-	-	-	0.21**	
	IPO	-	-0.27*	-0.20**	-	-	-	-0.19**	-	-	-0.23**	
Oct	Nino3.4	-	-0.20**	-	-	-	-	-	-	-	-	
	SOI	-	0.27*	0.21**	-	-	-	-	-	-	0.26**	
	IPO	-	-0.22**	-	-	-	-	-	-	-	-0.21**	
Nov	Nino3.4	-	-	-	-	-	-	-	-	-	-	
	SOI	-	0.25*	0.26*	-	-	-	-	-	-	0.23*	
	IPO	-	-	-	-	-	-	-0.20**	-	-	-	
Dec	Nino3.4	-	-	-	-	-	-	-	-	-	-	
	SOI	-	0.23**	-	-	-	-	-	-	-	0.24**	
	IPO	-	-0.20**	-	-	-	-	-0.19**	-	-	-0.22**	

Table 4 Input structure of WANN and ANN for Cairns station during the calibration

Model	Input set	WANN		ANN	Output
		Input matrix	Level	Input matrix	
Model 1	(Summer Rain) _t	12×93	3	2×93	(Summer Rain) _{t+1}
Model 2	(Summer Rain + Nino 3.4 _{Jan. to Apr.}) _t	60×93	3	10×93	(Summer Rain) _{t+1}
Model 3	(Summer Rain + SOI _{Feb.}) _t	24×93	3	4×93	(Summer Rain) _{t+1}
Model 4	(Summer Rain + IPO _{Feb. to Apr.}) _t	48×93	3	8×93	(Summer Rain) _{t+1}
Model 5	(IPO _{Feb. to Apr.}) _t	24×93	2	6×93	(Summer Rain) _{t+1}
Model 6	(SOI _{Feb.}) _t	8×93	2	2×93	(Summer Rain) _{t+1}
Model 7	(Nino 3.4 _{Jan. to Apr.}) _t	48×93	3	8×93	(Summer Rain) _{t+1}
Model 8	(Summer Rain + Nino 3.4 _{Jan. to Apr.} + SOI _{Feb.} + IPO _{Feb. to Apr.}) _t	108×93	3	18×93	(Summer Rain) _{t+1}

4.1 Comparisons of WANN with ANN

The prediction skills of the ANN and WANN models during the calibration and testing periods are presented in Table 5. It is found that compared to the other stations, Cairns and Pleystowe show high RMSE values for both the ANN and WANN methods. However, the lower RMSEs are observed for Cunnamulla, Boulia, and Isisford through ANN and WANN, respectively. During the testing process, the RMSE values through the ANN are 78.8mm and 95.6mm for Cunnamulla and Boulia, while the RMSE values for the Cunnamulla and Isisford stations are 55.3mm and 54.6mm for WANN. It is evident that RMSE values are remarkably higher for the tropical east coast region (Cairns, Calliope, and Pleystowe weather stations), resulting from a high magnitude of summer mean rainfall. The average RMSE values through the ten stations are obtained 169.5mm and 178.9mm using ANN during calibration and testing periods, respectively. Nonetheless, the corresponding average RMSE values are 108.7mm and 112.2mm, using the WANN during calibration and testing periods. The WANN improves the prediction accuracy of the ANN by 37% over the ten selected stations in terms of RMSE values. Moreover, it is worthwhile noting that the average MAE through ANN during the testing period is 141.7mm, while it is 90.5mm through the WANN.

The correlation coefficients (R) through the WANN are not less than 0.72 (Boulia station), whereas those values range between 0.14 and 0.61 through the ANN. It can be seen that for Calliope, the forecasted rainfall using WANN is well correlated with observed values for the fourth model (R=0.91). It is evident that the best model through the WANN is developed using the IPO index, which is applicable in all the stations except Pleystowe and Boulia. It is evident that the IPO is a determinative factor in predicting the summer rainfall in this region. However, the ANN does not follow a specific trend regarding the suitable input set for each station.

It is observed that the values of d_r for WANN are relatively higher (average d_r=0.68) than the values for ANN (average d_r=0.51) in all stations. According to the RMSE and MAE values, the superior performance of the WANN compared to the ANN is notable in all districts of Queensland.

Figure 4 shows the comparisons between WANN, ANN, and observed summer rainfall for the selected three regions of Queensland through testing sets. It is observed that the WANN models can predict the summer rainfall accurately, while the ANN is unable to forecast the observed values with the same accuracy. In Pleystowe station, the predicted

Table 5 Best ANN and WANN models performance

ANN									
Station	Model	Calibration				Testing			
		RMSE	MAE	d _r	R	RMSE	MAE	d _r	R
Cairns	8	383.8	262.9	0.64	0.70	398.6	317.4	0.54	0.36
Calliope	6	180.3	133.5	0.57	0.57	230.7	185.2	0.61	0.61
Pleystowe	5	400.6	295.5	0.58	0.46	418.4	343.7	0.47	0.14
Ayrshire Downs	8	107.2	82.9	0.62	0.66	100.1	69.82	0.53	0.53
Richmond	4	124.4	99.0	0.58	0.62	130.8	115.5	0.52	0.36
Boulia	3	101.2	77.0	0.51	0.37	95.6	88.7	0.3	0.23
Isisford	8	88.3	70.8	0.60	0.58	98.4	73.6	0.49	0.35
Harrisville	7	135.7	102.4	0.55	0.37	131.3	96.2	0.55	0.20
Cunnamulla	3	81.9	68.2	0.49	0.40	78.8	55.2	0.60	0.56
Pittsworth	6	91.2	44.3	0.70	0.64	106.4	71.5	0.57	0.41
Average	-	169.5	123.6	0.58	0.53	178.9	141.7	0.51	0.37

WANN									
Station	Model	Calibration				Testing			
		RMSE	MAE	d _r	R	RMSE	MAE	d _r	R
Cairns	4	321.4	217.4	0.70	0.77	307.9	255.4	0.63	0.75
Calliope	4	110.2	60.2	0.80	0.85	130.0	96.0	0.76	0.91
Pleystowe	1	219.5	170.1	0.76	0.88	215.1	174.9	0.73	0.87
Ayrshire Downs	8	49.4	21.9	0.89	0.93	67.7	51.6	0.66	0.84
Richmond	4	68.1	34.8	0.85	0.89	85.8	62.9	0.74	0.87
Boulia	1	53.7	42.1	0.73	0.88	52.7	44.6	0.64	0.72
Isisford	4	51.9	31.0	0.82	0.87	54.6	46.4	0.67	0.84
Harrisville	4	74.3	34.0	0.85	0.86	85.4	72.3	0.66	0.76
Cunnamulla	8	71.5	58.8	0.56	0.63	55.3	46.5	0.70	0.80
Pittsworth	4	67.3	50.8	0.66	0.73	67.8	54.6	0.67	0.80
Average	-	108.7	72.1	0.76	0.82	112.2	90.5	0.68	0.81

rainfall through ANN shows a smooth trend, and it is hardly able to predict the seasonal rainfall values correctly. Nonetheless, the results obtained by WANN ascertain that this method can calculate extreme summer rainfall with excellent accuracy. The extreme summer rainfall throughout the testing period for Ayrshire station is obtained as 302.6mm through the ANN, while it is acquired as 423.9mm by WANN when the observed value was 572.6mm. Such accuracy benefits researchers and water resources managers to plan appropriate solutions before the occurrence of extreme seasonal rainfall.

4.2 Comparisons of WANN's Forecast Accuracy with ACCESS-S, Climatology, ARIMA, and MLR Predictions

The accuracy of the WANN, ACCESS-S, and climatology are presented in Table 6. It is evident that ACCESS-S and climatology forecasts produce much higher RMSE and MAE values for all the stations. The highest RMSE value among all stations is related to Cairns

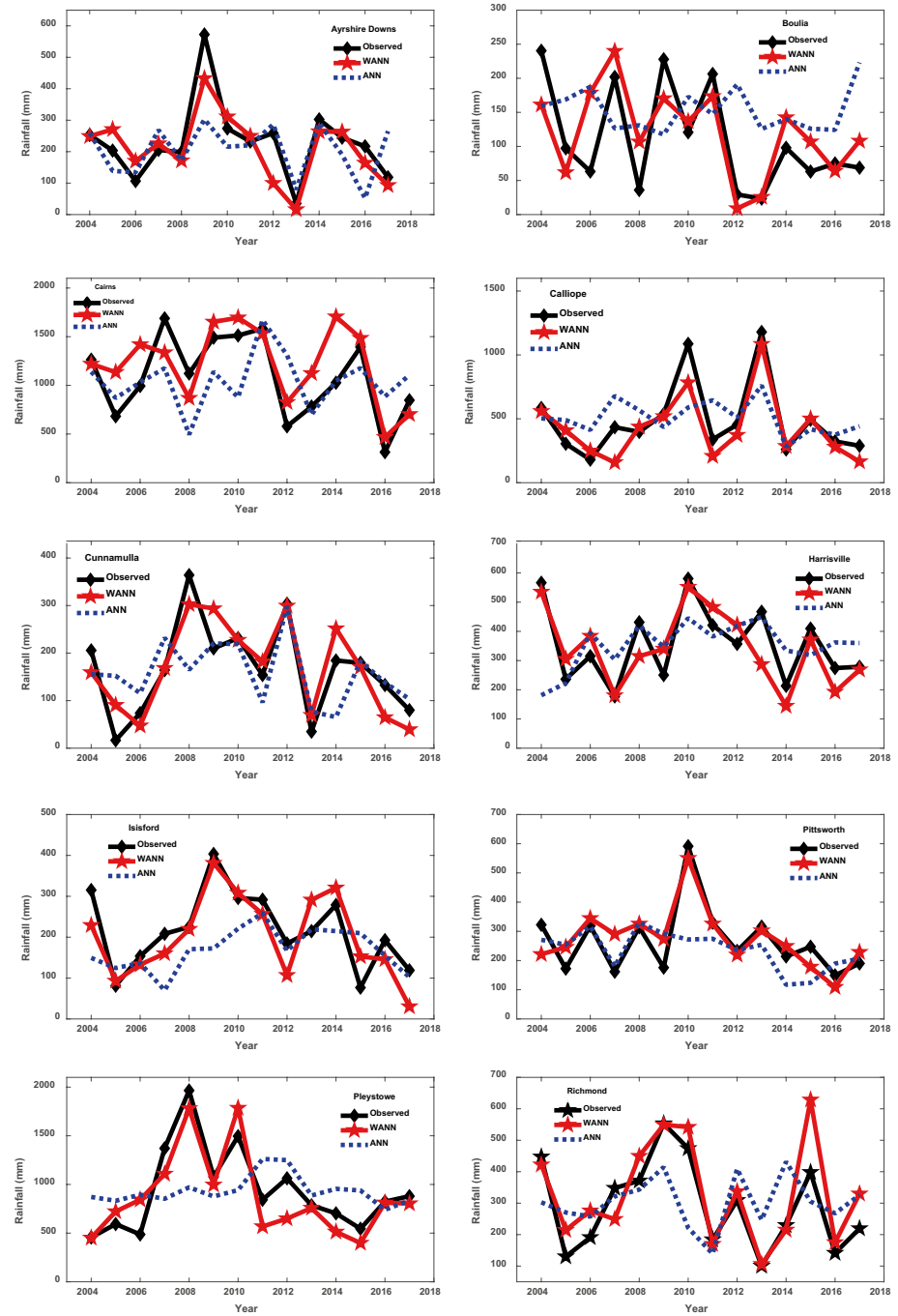


Fig. 4 Observed and modeled rainfall using ANN and WNN during the verification period at one-year lead time

Table 6 Comparison of predictions of WANN with ACCESS-S, and climatology during the testing period

Station	WANN			ACCESS-S			Climatology		
	RMSE	MAE	NSE	RMSE	MAE	NSE	RMSE	MAE	NSE
Cairns	307.9	255.4	0.45	564.5	443.7	-0.78	932.3	804.4	-2.75
Calliope	130.0	96.0	0.79	368.1	280.5	-2.24	277.6	201.1	-0.99
Pleystowe	215.1	174.9	0.73	685.7	520.5	-1.14	653.6	469.6	-0.99
Ayrshire Downs	67.7	51.6	0.67	215.0	171.0	-1.71	149.8	105.5	-0.29
Richmond	85.8	62.9	0.60	171.4	133.3	-0.63	202.6	154.7	-0.93
Boulia	52.7	44.6	0.48	110.8	93.1	-0.33	110.0	80.5	-0.33
Isisford	54.6	46.4	0.63	150.4	122.6	-1.75	101.9	80.0	-0.18
Harrisville	85.4	72.3	0.51	276.0	222.6	-8.13	141.0	105.7	-0.07
Cunnamulla	55.3	46.5	0.65	118.8	93.1	-0.63	121.0	89.2	-0.68
Pittsworth	67.8	54.6	0.62	156.8	116.5	-0.80	107.6	82.4	-0.01
Average	112.2	90.5	0.61	281.8	219.7	-1.8	279.7	217.3	-0.72

for WANN (RMSE=307.9mm), whereas the highest values of 564.5mm and 932.3mm are acquired for the same station through ACCESS-S and climatology models, respectively. It can be seen that the WANN outperforms the ACCESS-S in forecasting the seasonal rainfall by 60% over the ten selected stations in terms of RMSE values.

The NSE values are negative through ACCESS-S and climatology for all the stations. For the WANN model, the NSE values are positive (from 0.45 to 0.79), which denotes the perfect fit of predicted and observed values. In other words, the NSE values of WANN indicate that WANN is an effective method to predict rainfall with lower errors. Figure 5 illustrates the time series values of WANN and ACCESS-S rainfall predictions along with the observed rainfalls. It is evident that the WANN prediction results are remarkably accurate than the ACCESS-S forecast results. Also, the WANN performs better in predicting the extreme seasonal rainfall, while ACCESS-S is incapable of generating exact peak values. For example, the ACCESS-S underestimates the extreme rainfall over the testing period by 73% in the Ayrshire Downs station. Also, its performances for different stations can be summarised as 62% overestimation for Calliope, 77% underestimation for Cunnamulla, 80% underestimation for Isisford, and 66% underestimation for Richmond in regards to the extreme seasonal rainfall.

Table 7 shows the prediction accuracies of the ARIMA and MLR in terms of RMSE, MAE, d_r , and R are presented. It is evident that the values of RMSE through WANN modeling are much lower than the RMSEs' by ARIMA and MLR models. The percentages of RMSE improvements by WANN are remarkably high; for Ayrshire Downs, 68% and 47% compared to MLR and ARIMA, respectively. The lowest improvements through WANN are 24% and 27% compared to MLR and ARIMA for Cairns, respectively. For all the stations, the WANN provides better results compared to the MLR and ARIMA.

The skill score values for the ACCESS-S, WANN, ANN, MLR, and ARIMA predictions are shown in Table 8. The skills of all three models for Cairns and Richmond have positive values, which means that those models' RMSE values are lower than the climatology forecast's values. However, for all the stations (except Cairns and Richmond

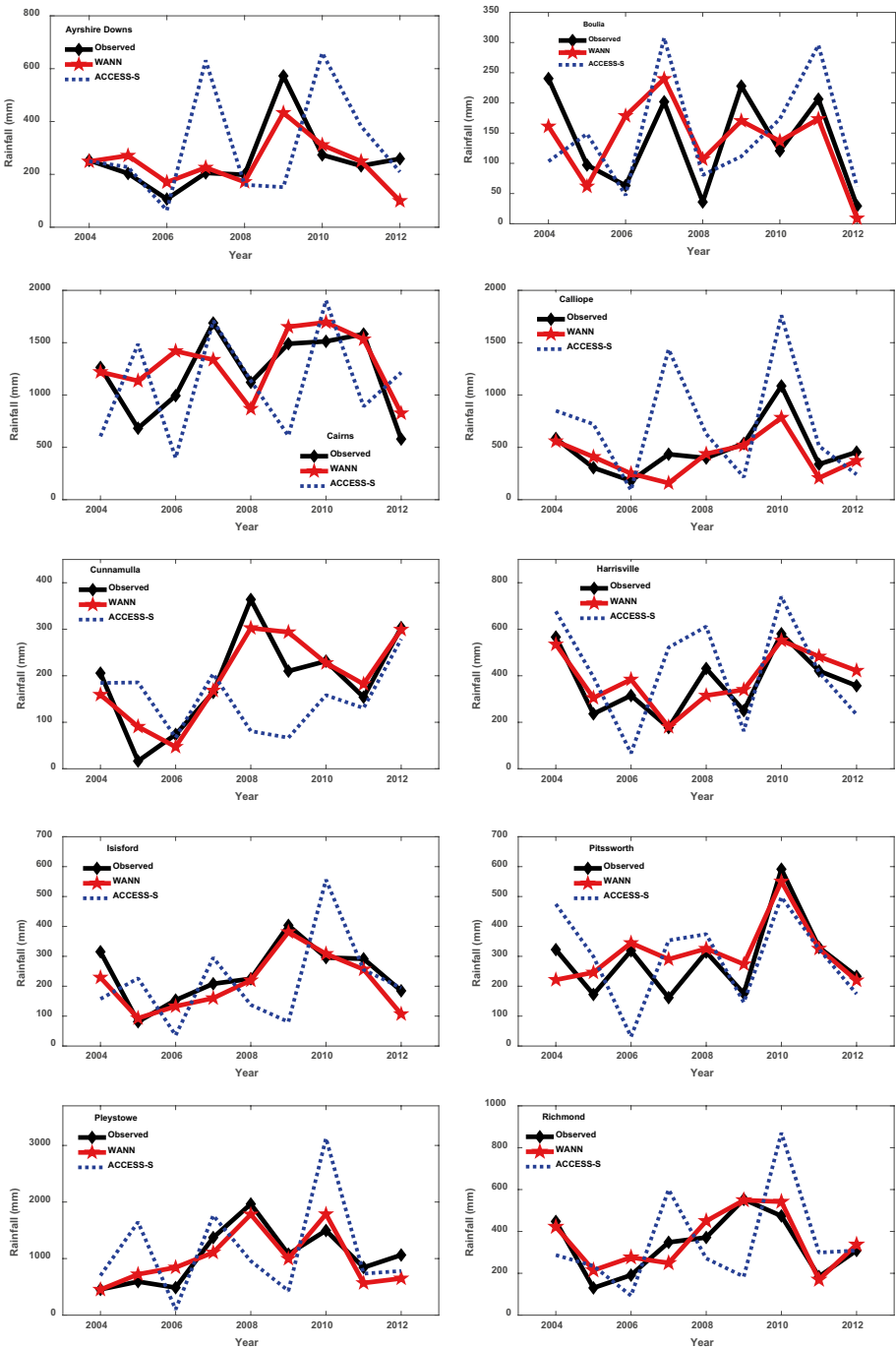


Fig. 5 Comparison of WANN and ACCESS-S predictions with observed rainfall at one-year lead time during the verification period

Table 7 Comparison of predictions of WANN with MLR, and ARIMA during the testing period

Station	WNN				MLR				ARIMA			
	RMSE	MAE	d_r	R	RMSE	MAE	d_r	R	RMSE	MAE	d_r	R
Cairns	307.9	255.4	0.63	0.75	405.7	353.0	0.49	0.11	423.5	361.2	0.48	0.01
Calliope	130.0	96.0	0.76	0.91	297.0	190.9	0.54	0.15	316.0	209.7	0.50	0.12
Pleystowe	215.1	174.9	0.73	0.87	404.3	292.7	0.46	0.54	370.4	282.4	0.56	0.45
Ayrshire Downs	67.7	51.6	0.66	0.84	212.7	199.7	-0.24	0.07	128.0	88.2	0.41	0.04
Richmond	85.8	62.9	0.74	0.87	133.6	117.8	0.50	0.10	141.4	122.8	0.48	0.10
Boulia	52.7	44.6	0.64	0.72	77.1	70.6	0.47	0.06	76.3	70.1	0.47	0.09
Isisford	54.6	46.4	0.67	0.84	90.1	70.3	0.50	0.12	90.6	70.5	0.49	0.07
Harrisville	85.4	72.3	0.66	0.76	127.5	110.1	0.50	0.39	129.0	11.0	0.49	0.41
Cunnamulla	55.3	46.5	0.70	0.80	96.9	76.8	0.49	0.02	96.2	77.1	0.49	0.05
Pittsworth	67.8	54.6	0.67	0.80	105.7	78.8	0.51	0.16	106.1	78.9	0.51	0.16
Average	112.2	90.5	0.68	0.81	195.1	156.1	0.4	0.2	187.7	137.2	0.5	0.2

with 39% and 15%, respectively), the skill scores are negative through the ACCESS-S. This indicates that the RMSEs of the ACCESS-S are higher than the RMSE values of climatology prediction, and the prediction skill of the ACCESS-S is worse than the climatology. However, the prediction capability of the WANN is higher than the climatology forecast (lower RMSE values) for all the stations. The ANN, MLR, and ARIMA prediction skills are very close to the climatology skill for Pittsworth with 1%, 2%, and 1% of skill scores, respectively. In general, the WANN renders an excellent accuracy in terms of the average skill score for all the stations with a value of 53%, and ACCESS-S provides worse prediction compared to the climatology forecast with an average skill score of -22%. Overall, the achieved average skill scores through the ANN, ARIMA, and MLR are 24%, 20%, and 15%, respectively.

Table 8 Percentage of forecast skill score for ACCESS-S, WANN, ANN, MLR, and ARIMA during the testing period

Station	ACCESS-S	WANN	ANN	MLR	ARIMA
Cairns	39%	67%	57%	56%	55%
Calliope	-33%	53%	17%	-7%	-14%
Pleystowe	-5%	67%	36%	38%	43%
Ayrshire Downs	-43%	55%	33%	-42%	15%
Richmond	15%	58%	35%	34%	30%
Boulia	-1%	52%	13%	30%	31%
Isisford	-48%	46%	3%	12%	11%
Harrisville	-96%	39%	7%	10%	9%
Cunnamulla	-2%	54%	35%	20%	20%
Pittsworth	-46%	37%	1%	2%	1%
Average	-22%	53%	24%	15%	20%

5 Conclusion

In this study, a model called WANN was developed to forecast Queensland's seasonal rainfall one year in advance. The results were compared with the conventional ANN, MLR, ARIMA, and the current prediction system of the BOM (ACCESS-S) and climatology forecasts. It was found that the WANN improves the ANN and ACCESS-S prediction results in terms of RMSE over the ten selected stations by 37% and 60%, respectively. However, the WANN could improve the predictions by ARIMA and MLR by 40% and 42%, correspondingly. In terms of skill score, the WANN with 53% outperforms ANN, ACCESS-S, ARIMA, and MLR with 24%, -22%, 20%, and 15%, respectively. It was also found that ACCESS-S's performance is not satisfactory regarding the average NSE (-1.8), compared to the average NSE value of 0.61 achieved by WANN. While the skill score values are negative for ACCESS-S for most stations, they are positive (higher than 0.37) through the WANN.

Since the MLR and ARIMA are linear methods, they are incapable of properly capturing the non-linear characteristic of the rainfall time series, and they are unable to accurately replicate a great range of data. Nonetheless, the WANN exceptionally outperforms the MLR and ARIMA because it can appropriately capture the non-linearity of the subject time series. It should be noted that the ANN cannot improve the prediction skills of the ARIMA and MLR. Thereby, it is seen that combining the wavelet analysis with ANN should be taken into account for the prediction of the seasonal rainfall with a low number of available data. However, in WANN implementation, the mother wavelet must be carefully selected through the trial and error process as it highly impacts the model's performance. Also, selecting the appropriate decomposition level in wavelet analysis affects the performance of the model. Therefore, careful steps must be taken in determining the number of levels and type of wavelet function in wavelet analysis.

To the best of our knowledge, no model can predict the seasonal rainfall one year in advance with such a level of accuracy. Thus, this capability can be beneficial for better management of unusual events such as flooding and drought, especially for regions with extreme weather conditions. This study opened an opportunity to predict seasonal rainfall one year ahead with excellent precision using various effective climate indices. It should be noted that the outcomes of this study show that this technique can be applied to forecast rainfalls of other seasons, and application of this method in longer lead times could be attempted for those seasons.

Authors Contributions M.G. – data collection, model setup, analysis, draft paper writing; M.I. – project conceptualisation and supervision, final paper preparation.

Funding No funding received for the study.

Availability of Data and Materials Data can be made available through contacting the first author

Declarations

Ethical Approval Not required as no animal/human was involved in the study.

Consent to Participate Authors agreeing to participate any survey or feedback tasks.

Consent to Publish Authors providing consent publication of the manuscript to the journal publisher.

Competing Interests There is no competing interest in regard to the study

References

- Abdel-Kader H, Abd-El Salam M, Mohamed M (2021) Hybrid Machine Learning Model for Rainfall Forecasting. *J Intell Syst* 1:5–12. <https://doi.org/10.5281/zenodo.3376685>
- Abtew W, Trimble P (2010) El Niño-Southern Oscillation link to South Florida hydrology and water management applications. *Water Resour Manag* 24:4255–71. <https://doi.org/10.1007/s11269-010-9656-2>
- Babel MS, Badgajar GB, Shinde VR (2015) Using the mutual information technique to select explanatory variables in artificial neural networks for rainfall forecasting. *Meteorol Appl* 22:610–6. <https://doi.org/10.1002/met.1495>
- Bagirov AM, Mahmood A (2018) A comparative assessment of models to predict monthly rainfall in Australia. *Water Resour Manag* 32:1777–94. <https://doi.org/10.1007/s11269-018-1903-y>
- Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. John Wiley & Sons, New Jersey
- Brown JD (2014) Linear models in matrix form Springer, Cham, Switzerland. <https://doi.org/10.1007/978-3-319-11734-8>
- Chong KL, Lai SH, Yao Y, Ahmed AN, Jaafar WZ, El-Shafie A (2020) Performance enhancement model for rainfall forecasting utilizing integrated wavelet-convolutional neural network. *Water Resour Manag* 34:2371–87. <https://doi.org/10.1007/s11269-020-02554-z>
- Diop L, Samadianfard S, Bodian A, Yaseen ZM, Ghorbani MA, Salimi H (2020) Annual rainfall forecasting using hybrid artificial intelligence model: integration of multilayer perceptron with whale optimization algorithm. *Water Resour Manag* 34:733–46. <https://doi.org/10.1007/s11269-019-02473-8>
- Farrokhi A, Farzin S, Mousavi SF (2020) A new framework for evaluation of rainfall temporal variability through principal component analysis, hybrid adaptive neuro-fuzzy inference system, and innovative trend analysis methodology. *Water Resour Manag* 34:3363–85. <https://doi.org/10.1007/s11269-020-02618-0>
- Galvao RK, Araujo MC, Fragoso WD, Silva EC, Jose GE, Soares SF, Paiva HM (2008) A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. *Chemometr Intell Lab Syst* 92:83–91. <https://doi.org/10.1016/j.chemolab.2007.12.004>
- Ghamariadyan M, Imteaz MA (2021a) A wavelet artificial neural network method for medium-term rainfall prediction in Queensland (Australia) and the comparisons with conventional methods. *Int J Climatol* 41:E1396–E1416. <https://doi.org/10.1002/joc.6775>
- Ghamariadyan M, Imteaz MA (2021b) Monthly rainfall forecasting using temperature and climate indices through a hybrid method in Queensland, Australia. *J Hydrometeorol* 22:1259–73. <https://doi.org/10.1175/JHM-D-20-0169.1>
- Gupta HV, Kling H (2011) On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. *Water Resour Res* 10. <https://doi.org/10.1029/2011WR010962>
- Hasan MM, Dunn PK (2012) Understanding the effect of climatology on monthly rainfall amounts in Australia using Tweedie GLMs. *Int J Climatol* 32:1006–17. <https://doi.org/10.1002/joc.2332>
- Hawthorne S, Wang QJ, Schepen A, Robertson D (2013) Effective use of general circulation model outputs for forecasting monthly rainfalls to long lead times. *Water Resour Res* 49:5427–36. <https://doi.org/10.1002/wrcr.20453>
- He X, Guan H, Qin J (2015) A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. *J Hydrol* 527:88–100. <https://doi.org/10.1016/j.jhydrol.2015.04.047>
- Islam F, Imteaz MA (2020) Use of teleconnections to predict Western Australian seasonal rainfall using ARIMAX model. *Hydrol* 7:52. <https://doi.org/10.3390/hydrology7030052>
- Kim TW, Valdés JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *J Hydrol Eng* 8:319–28. [https://doi.org/10.1061/\(asce\)1084-0699\(2003\)8:6\(319\)](https://doi.org/10.1061/(asce)1084-0699(2003)8:6(319))
- Knip DM, Heupel MR, Simpfendorfer CA, Tobin AJ, Moloney J (2011) Wet-season effects on the distribution of juvenile pigeye sharks, *Carcharhinus amboinensis*, in tropical nearshore waters. *Mar Freshw Res* 62:658–67. <https://doi.org/10.1071/MF10136>
- Mallat S (1999) A wavelet tour of signal processing. Elsevier, London. <https://doi.org/10.1016/b978-0-12-466606-1.x5000-4>
- Mekanic F, Imteaz MA, Talei A (2016) Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. *Clim Dyn* 46:3097–111. <https://doi.org/10.1007/s00382-015-2755-2>

- Mekanik F, Imteaz MA, Gato-Trinidad S, Elmahdi A (2013) Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *J Hydrol* 503:11–21. <https://doi.org/10.1016/j.jhydrol.2013.08.035>
- Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ Model Softw* 15:101–24. [https://doi.org/10.1016/s1364-8152\(99\)00007-9](https://doi.org/10.1016/s1364-8152(99)00007-9)
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. *J hydrol* 10:282–90. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nourani V, Komasi M, Alami MT (2012) Hybrid wavelet–genetic programming approach to optimize ANN modeling of rainfall–runoff process. *J Hydrol Eng* 17:724–41. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000506](https://doi.org/10.1061/(asce)he.1943-5584.0000506)
- Pakdaman M, Falamarzi Y, Babaieian I, Javanshiri Z (2020) Post-processing of the North American multi-model ensemble for monthly forecast of precipitation based on neural network models. *Theor Appl Climatol* 141:405–17. <https://doi.org/10.1007/s00704-020-03211-6>
- Poonia V, Tiwari HL (2020) Rainfall-runoff modeling for the Hoshangabad Basin of Narmada River using artificial neural network. *Arab J Geosci* 13:1. <https://doi.org/10.1007/s12517-020-05930-6>
- Power S, Casey T, Folland C, Colman A, Mehta V (1999) Inter-decadal modulation of the impact of ENSO on Australia. *Clim Dyn* 15:319–24. <https://doi.org/10.1007/s003820050284>
- Schepen A, Wang QJ, Robertson D (2012) Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *J Clim* 25:1230–46. <https://doi.org/10.1175/jcli-d-11-00156.1>
- Sharghi E, Nourani V, Najafi H, Molajou A (2018) Emotional ANN (EANN) and wavelet-ANN (WANN) approaches for Markovian and seasonal based modeling of rainfall-runoff process. *Water Resour Res* 32:3441–56. <https://doi.org/10.1007/s11269-018-2000-y>
- Sumner G, Bonell M (1988) Variation in the spatial organisation of daily rainfall during the North Queensland wet seasons, 1979–82. *Theor Appl Climatol* 39:59–72. <https://doi.org/10.1007/bf00866390>
- Tozer CR, Kiem AS, Verdon-Kidd DC (2017) Large-scale ocean-atmospheric processes and seasonal rainfall variability in South Australia: potential for improving seasonal hydroclimatic forecasts. *Int J Climatol* 37:861–77. <https://doi.org/10.1002/joc.5043>
- Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. *Int J Climatol* 32:2088–94. <https://doi.org/10.1002/joc.2419>
- Wilson L, Manton MJ, Siems ST (2013) Relationship between rainfall and weather regimes in south-eastern Queensland, Australia. *Int J Climatol* 33:979–91. <https://doi.org/10.1002/joc.3484>
- Yazdandoost F, Moradian S, Zakipour M, Izadi A, Bavandpour M (2020) Improving the precipitation forecasts of the North-American multi model ensemble (NMME) over Sistan basin. *J Hydrol* 590:125263. <https://doi.org/10.1016/j.jhydrol.2020.125263>
- Yadav P, Sagar A (2019) Rainfall prediction using artificial neural network (ANN) for Tarai region of Uttarakhand. *Curr J Appl Sci Technol* 33:1–7. <https://doi.org/10.9734/cjast/2019/v33i530096>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.