The effectiveness of ARIMAX model for prediction of summer rainfall in northwest Western Australia

F Islam¹ and M A Imteaz¹,²

¹Department of Civil & Construction Engineering, Swinburne University of Technology, Hawthorn, VIC 3122, Australia.
²Corresponding author email: mimteaz@swin.edu.au

Abstract. Inherent climate mechanism plays a large role in Australian rainfall events, while most interestingly, such influence of climate drivers on rainfall generation depends largely on a place’s geographical location and the season it has been passing through. For Northwest Western Australia (NWWA), summer has been the main rainfall season; therefore, it became necessary to develop a seasonal forecast model several months ahead of the event. This study involves the selection of potential climate drivers for NWWA and using them to forecast and model long-term seasonal summer rainfall (Dec to Feb) several months advance. Two rainfall stations (Bidyadanga and Gogo) were selected considering the availability of continuous monthly rainfall data (from 1915 to 2015) with minimal missing value. Simple multiple linear regression (MLR) and ARIMAX have been performed on acquired rainfall data and lagged climate indices. It has been observed that the Southern Oscillation Index (SOI) and the Western Indian Ocean Index (WIO) has a great influence on NWWA rainfall. As such, the WIO-SOI model (4 months advance) shows a significantly higher correlation in the ARIMAX model (0.68 and 0.65 for Bidyadanga and Gogo station, respectively) compared to the MLR model (0.35 and 0.36) and confirms its effectiveness to be used for rainfall prediction in the region. It is believed that the simple but effective nature of the developed models can produce optimum benefit on stakeholder’s decision making to tackle future socio-economic challenges.

1. Introduction

Rainfall forecasting plays an important role in the agroeconomic interest, socioeconomic interest, and disaster management initiatives. Having a reliable and trustworthy forecast several months ahead of the event can help the farmers and stakeholders to take timely decisions to mitigate any potential damage. Also, accurate prediction of rainfall events can help to set strategic plans to encounter extreme weather conditions. Rainfall forecasting can be performed using statistical and dynamic principles [1]. Statistical methods are mostly used to forecast rainfall as it is considered as less complex, time-efficient however, it may require a long period of uninterrupted data. In contrast, dynamic models are very complex, time-consuming, and costly [2, 3]. Ideally, statistical models are preferred over its counterparts, due to its simplicity and easy to use application.

At Present, time series analysis has become a very popular statistical method to develop forecast models and its use has got a global spread. The use of time series analysis has been evident to forecast rainfall events [4, 5], drought events [6], rainfall-runoff patterns [7, 8], and rainfall-temperature interaction [9]. In particular, for Australian rainfall, the use of the ARIMA model was also conveyed and reported [10, 11]. Current literature suggests that seasonal rainfall forecasting in Western Australia...
(WA) is mostly location-oriented. Researchers have divided the WA region into different parts for their convenience of study. Several studies were found that dealt with South West Western Australia (SWWA)[12-15], while, a limited number of studies considered North West Western Australia (NWWA) rainfall variability [16]. Moreover, most of these studies tried to establish a concurrent relationship between climate drivers and WA rainfall. However, no evidence was found that evaluated the lagged relationship between climate indices and seasonal rainfall in NWWA. A recent study conducted in Southwest Western Australia that predicts seasonal rainfall a few months in advance using the ARIMAX model also depicted promising results [17]. As summer rainfall is the main rainfall season for NWWA, therefore, this study evaluated the lagged relationship between climate indices and seasonal summer rainfall in the Kimberly division of WA. Both the MLR and ARIMAX methods were used to assess summer rainfall variability. The effect of potential climate drivers namely ENSO grouped indices and WIO were analyzed against WA summer rainfall as they have shown significant contribution to such variability [15, 16, 18, 19]. It is expected that this study will identify the dominant climate indices responsible for NWWA summer rainfall. Furthermore, having a reliable summer rainfall forecast model several months advance for NWWA is expected to help the stakeholders and policy makers in decision making for feasibility analysis, crop production plan, and disaster and risk management.

2. Data and Study area
In this study, two rainfall stations from the Kimberly division of WA have been selected. Figure 1 and table 1 show the location and description of selected rainfall stations, respectively. 101 years of rainfall data (1915-2015) were collected considering minimum missing value. These data were divided into two parts using the 70:30 split-calibration period (1915-1985) and validation period (1986-2015). Australian standard summer monthly rainfall data for December, January, and February were used.

![Figure 1. The geographical location of Bidyadanga and Gogo Station.](image)
SOI and WIO were found as the most influential climate indices for NWWA rainfall. SOI is the measure of sea level pressure (SLP) difference between Darwin and Tahiti, and WIO is the average sea-surface temperature (SST) anomaly index which is the indicator of the surface temperatures in the western tropical Indian ocean. WIO calculated by SSTs in the box 50°E - 70°E, 10°S - 10°N, and is one half of the dipole mode index (DMI). DMI is defined as the average SST anomalies between the tropical western Indian Ocean (WIO) bounded by 10°N to 10°S, from 50°E to 70°E, and the tropical south-eastern Indian Ocean (SEIO) bounded by 110°S to the equator, from 90°E to 110°E.

Several data sets were available for both SOI and WIO climate indices, however, Jones for SOI and HadISST1 for WIO were used for this study. This is due to the popularity and acceptability of these data set among the researchers, and the availability of longer period data. Monthly climate index data for 100 years were collected and prepared as lagged climate indices for WIO (Marchn-1 - Novn) and SOI (Marchn-1 - Novn). Here, ‘n’ being the year of which seasonal summer rainfall is predicted. The data description has been presented in table 2.

### Table 1. Description of selected rainfall stations

| Station No. | Station Name | Latitude | Longitude | Elevation (m) | Data Period   |
|-------------|--------------|----------|-----------|---------------|---------------|
| 3030        | Bidyadanga   | 18.68° S | 121.78° E | 11            | 1915-2015     |
| 3014        | Gogo Station | 18.29° S | 125.59° E | 150           | 1915-2015     |

*Note: Data Source: www.bom.gov.au/climate/data/

### Table 2. Data description of selected climate indices

| Climate Indices | Region       | Data Period   | Data Set   |
|-----------------|--------------|---------------|------------|
| SOI             | Pacific Ocean| 1915-2015     | Jones      |
| WIO             | Indian Ocean | 1915-2015     | HadISST1   |

3. Methodology

A simple linear relationship has been employed to determine the significant correlation between lagged climate indices and rainfall. Multiple linear regressions technique was used to determine the predictability of rainfall using two or more climate indices. From the results of bivariate correlation analysis, significant climate indices were used as predictors in the ARIMAX Method.

#### 3.1. Multiple linear regression (MLR)

Multiple Linear Regression (MLR) describes the linear relationship between a dependent and two or more independent variables. The dependent variable is also called the outcome variable and independent variables are named as Predictors. Three assumption check is necessary before conducting linear regression analysis: linearity, equal of variance, and normality. In a linearity check, the linear relationship between the dependent and independent variables is verified. In equal of variance, the spread of the residuals is checked and in normality check, data distribution is sought whether it is normally distributed or not. If selected variables satisfy all these assumption checks, linear regression analysis can proceed further. The equation for the MLR model is given below:

\[ Y = a + b_1 X_1 + b_2 X_2 + e \]  

where, \( Y \) is the dependent variable (i.e. Summer rainfall in this study); \( X_1 \) is the first independent variable (i.e. lagged WIO) and \( X_2 \) is the second independent variable (i.e. lagged SOI); \( b_1 \) and \( b_2 \) are the model coefficient, \( a \) is constant and \( e \) is an error. A factor that can compromise the effectiveness of multiple linear regression analysis is the evidence of multi-collinearity among the predictors. It can be identified using Tolerance (T) and Variance Inflation Factor (VIF), as defined as follows:

\[ \text{Tolerance} = 1 - R^2 \]  

3
VIF = \frac{1}{\text{Tolerance}} \quad (3)

where, R^2 is the coefficient of determinations, which can be obtained from the ratio of regression sum of squares (SSR) and the total sum of squares (SST). To determine whether multicollinearity among the predictors exist or not, tolerance value close to zero and VIF greater than 10 is a clear indication of the multi-collinearity problem [20, 21]. The Durbin Watson (D-W) test statistics is another test that detects the presence of existing autocorrelations in the residuals (i.e. prediction errors from a regression analysis). According to Field [22], D-W values less than 1 or greater than 3 are delineated as a definite indication of autocorrelation among the residuals.

3.2. Multivariate auto-regressive integrated moving average with exogenous input (ARIMAX)

In multivariate ARIMA models, ARIMA order for the dependent variable (i.e. summer rainfall) and transfer function order for predictors (i.e. climate indices). These models are used to predict future values using their past values. ARIMA model is capable to predict future values with or without using predictors. When independent variables or predictors are included in the ARIMA model, it is termed as the ARIMAX model. However, predictors can only be included in the ARIMA model if it has a significant correlation with the dependent variable [23, 24]. General ARIMA model is expressed as ARIMA (p, d, q) *(P, D, Q). This expression has got two parts: one is non-seasonal and the other one is seasonal. Here, (p, d, q) is the non-seasonal part and (P, D, Q) is the seasonal part: ‘p’ is non-seasonal auto-regressive order (AR), ‘d’ is non-seasonal differencing, ‘q’ is non-seasonal moving average (MA); ‘P’ is seasonal auto-regressive, ‘D’ is seasonal differencing and ‘Q’ is the seasonal moving average. The mathematical expression for the ARIMAX model is given below:

\[
\Delta Y_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i \Delta Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{m=1}^{M} \beta_m X_{t-m} \quad (4)
\]

where, \(\varphi_1, \ldots, \varphi_p\) and \(\theta_1, \ldots, \theta_q\) are the parameters; \(\varepsilon_t, \varepsilon_{t-1}\) are white noise error and \(\beta_1, \ldots, \beta_m\) are the parameters of independent variables input \(X_t\) and \(t\) is the time.

Setting up the ARIMAX model consists of three steps: identification, parameter estimation and selection, and diagnostic checks. Identification involves with decision making on data stationarity, parameter estimation and selection involves with the order selection of AR and MA, while diagnostic check involves with verification of no existence of autocorrelation among the residuals. A detailed description of these steps is presented in Islam and Imteaz [17].

3.3. Statistical parameter for MLR and ARIMAX model

The predictability of a regression equation can be justified by evaluating statistical performance tests such as t-test, F-test, Pearson correlation coefficient (r), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). For both RMSE and MAE, a value of 0 indicates a perfect predictability performance. Thus, the lower the RMSE, the better the performance. The statistical parameters considered to evaluate the ARIMAX model are Pearson correlation(r), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Normalized BIC (Bayesian Information Criterion) values. The model with the highest r, lowest error, and Normalized BIC values are considered as a good model. Additionally, Ljung Box statistics is also checked whether the P-value is greater than 0.05 or not. A P-value greater than 0.05 indicates residuals are white noise which indicated there is no autocorrelation among the residuals and the developed model is adequate. However, all these tests have some shortcomings which can be overruled by the introduction of a refined index of agreement (\(d_r\)) developed by Willmott [25].
4. Results and discussion

Two rainfall stations from NWWA namely: Bidyadanga and Gogo were selected to conduct this study. Lagged monthly climate indices values (March_{n-1} to Nov_{n}) were used to investigate the correlation with summer. Here, n is the predicted year, and n-1 is the previous year. Bivariate correlation analysis showed that NWWA summer rainfall has a significant correlation with WIO and SOI climate indices. However, it did not show any correlation with DMI and ENSO based climate indices. Table 3 and 4 are showing the bivariate correlation result for the selected rainfall stations.

**Table 3.** Pearson Correlation \( (r) \) between Summer rainfall of Bidyadanga and climate indices.

| Climate Index | WIOMar | WIOMay | WIOJune | WIOJuly | WIOAug | WIOSep | WIOOct | WIONov |
|---------------|--------|--------|---------|---------|--------|--------|--------|--------|
| Pearson Correlation \( (r) \) | 0.15   | 0.25*  | 0.23    | 0.17    | 0.24*  | 0.24*  | 0.12   | 0.11   | 0.06   |

Note: **Correlation is significant at the 0.01 level (2-tailed), *Correlation is significant at the 0.05 level (2-tailed).

**Table 4.** Pearson Correlation \( (r) \) between Summer rainfall of Gogo station and climate indices.

| Climate Index | WIOMar | WIOMay | WIOJune | WIOJuly | WIOAug | WIOSep | WIOOct | WIONov |
|---------------|--------|--------|---------|---------|--------|--------|--------|--------|
| Pearson Correlation \( (r) \) | 0.10   | 0.14   | 0.27*   | 0.27*   | 0.32** | 0.32** | 0.22   | 0.06   |

Note: **Correlation is significant at the 0.01 level (2-tailed), *Correlation is significant at the 0.05 level (2-tailed).

From the above outcomes, several multiple regression model sets were developed with the combination of lagged WIO and SOI index. Table 5 presents the developed MLR model sets for Bidyadanga and Gogo station.

**Table 5.** Developed MLR model sets.

| Bidyadanga | Gogo Station |
|------------|--------------|
| Combination: Lagged WIO-SOI |

\[ d_r = \begin{cases} \frac{1 - \sum_{i=1}^{n}(P_i - \bar{O})}{c \sum_{i=1}^{n}(O_i - \bar{O})}, & \text{when} \\ \frac{\sum_{i=1}^{n}(P_i - O_i)}{c \sum_{i=1}^{n}(O_i - \bar{O})} - 1, & \text{when} \end{cases} \]
For each MLR model, statistical parameters such as RMSE, MAE, D-W test, tolerance, and VIF were calculated. Among all these models, the model showed the highest Pearson correlation ($r$), lowest error, and no evidence of multicollinearity among the residual was selected as the best model. The description of the selected best model for these two rainfall stations is presented in Table 6. From the MLR analysis, it was observed that for both the stations, the Pearson correlation ($r$) value got increased if compared to bivariate or single correlation analysis. Among all these models, sets Bidyadanga (WIO$_{Aug}$-SOI$_{May}$) and Gogo station (WIO$_{Aug}$-SOI$_{Mar}$) showed the highest correlation, lowest error, and no autocorrelation among the residuals. Besides, these two developed models can predict NWWA summer (Dec to Feb) rainfall four (4) months advances.

Table 6. Description of the best model set in MLR.

| Station Name | Model Input | Constant | Coefficient | Pearson Correlation ($r$) | RMSE | MAE | Durbin Watson (D/W) | Tolerance (T) | VIF |
|--------------|-------------|----------|-------------|-----------------|------|-----|---------------------|---------------|-----|
| Bidyadanga   | WIO$_{Aug}$ | 104.33   | 72.96       | 0.35            | 64.72| 49.68 | 1.95                | 0.99          | 1.02|
|              | SOI$_{May}$ |          | 19.39       |                 |      |      |                     |               |     |
| Gogo Station | WIO$_{Aug}$ | 113.07   | 50.27       | 0.36            | 43.95| 35.88| 1.803               | 0.83          | 1.20|
|              | SOI$_{Mar}$ |          | -7.12       |                 |      |      |                     |               |     |

For the ARIMAX model, variables that showed a significant correlation at 1% and 5% level were considered and their suitability was examined. A detailed investigation depicted that Western Australian summer rainfall and climate index data were found as non-stationary and non-seasonal. As the ARIMAX model data type has to be stationary, first-order differencing was applied. First-order differencing was chosen as it is deemed sufficient for any linear trend. Also, as the dataset has been found as non-seasonal, thus, the seasonal part has been disregarded. The next step was to select the appropriate AR and MA order for the model. To do so, the autocorrelation function (ACF) plot and partial autocorrelation plot (PACF) plot of rainfall were used. ACF plot is used for the selection of MA order and the PACF plot used for the selection of the AR order. Figure 2 shows ACF and PACF plots with the lag number for the selected rainfall stations. From these plots, after analyzing the spike and decay conditions, Bidyadanga: (0,1,1) and Gogo: (3,1,2) models were selected.
Figure 2. ACF and PACF plot of rainfall stations (a) Bidyadanga and (b) Gogo Station

Table 7 shows the selected model inputs for the AR and MA orders. Considering the results of the bivariate correlation analysis, several ARIMAX models were developed. A model summary of the selected ARIMAX model is given in Table 8. WIO$_{Aug}$–SOI$_{Mar}$ from Bidyadanga and WIO$_{Aug}$–SOI$_{Mar}$ from Gogo station showed the highest correlation ($r$) of 0.68 and 0.65, respectively with four months lagged period. Both these selected models showed the lowest error and the residual was found as white noise. It is previously mentioned that a p-value greater than 0.05 indicates the presence of no autocorrelation among the residuals. For all the selected rainfall stations, the p-value has found to be greater than 0.05 (refer to Table 8). Another alternative way of checking autocorrelation is a residual ACF and PACF plot. In these plots, if all the spikes remain between the confidence limit, the residuals are declared as white noise as illustrated in Figure 3.

Table 7. ARIMA and transfer function input.

| Station Name      | ARIMA order | Transfer Function order |
|-------------------|-------------|------------------------|
|                   | Autoregressive | Difference | Moving Average | Denominator | Difference |
|                   | (AR)       | (D)       | (MA)         | Numerator  | Denominator | Difference |
| Bidyadanga        | 0          | 1         | 1            | 1          | 1           | 1          |
| Gogo Station      | 3          | 1         | 2            | 1          | 1           | 1          |

Table 8. Summary of the ARIMAX model results.

| Rainfall Station | Model Type | Lag month | Predictor         | Model Fit statistics | Ljung-Box Q (18) |
|------------------|------------|-----------|-------------------|----------------------|------------------|
|                  |            |           |                   | $r$                  |                   |
| Bidyadanga       | ARIMAX     | 4         | WIO$_{Aug}$–SOI$_{May}$ | 0.68                 | 8.82 17 0.94     |
|                  | (0.1,1)    |           |                   | 57.90 41.97 8.94    |                   |
| Gogo Station     | ARIMAX     | 4         | WIO$_{Aug}$–SOI$_{Mar}$ | 0.65                 | 20.21 13 0.09    |
|                  | (3.1,2)    |           |                   | 40.01 27.49 8.12    |                   |
A validation test was also performed with the same model input sets for both MLR and ARIMAX analysis. Refined Willmott index of agreement ($d_r$) was also calculated for the developed models and finally, both the model outputs were compared. A detailed summary of both the methods is illustrated in table 9 for both the calibration and validation periods.

| Model | Station Name | Predictors | Calibration Period | Validation Period |
|-------|--------------|------------|-------------------|-------------------|
|       |              |            | Pearson correlation ($r$) | Refined Willmott Index of agreement ($d_r$) | Pearson correlation ($r$) | Refined Willmott Index of agreement ($d_r$) |
| MLR   | Bidyadanga   | WIO Aug-SOI May | 0.35 | 0.61 | 0.36 | 0.57 |
|       | Gogo Station | WIO Aug-SOI Mar | 0.36 | 0.54 | 0.35 | 0.52 |
| ARIMAX| Bidyadanga   | WIO Aug-SOI May | 0.68 | 0.63 | 0.77 | 0.70 |
|       | Gogo Station | WIO Aug-SOI Mar | 0.65 | 0.66 | 0.74 | 0.70 |

From table 9, it was evident that ARIMAX models are superior to the multiple regression models in both calibration and validation periods. In ARIMAX modeling, both the correlation ($r$) and ‘$d_r$’ values were increased when compared with the results of regression models. To be more specific, in the ARIMAX model Pearson correlation($r$) has been increased from 0.35 to 0.68 and 0.36 to 0.65 in calibration and 0.36 to 0.77 and 0.35 to 0.74 in the validation period for Bidyadanga and Gogo station respectively. Besides, the Refined Willmott Index of agreement ($d_r$) value also increased in the ARIMAX model compared to the MLR model. Overall, the ARIMAX model overperformed compared to the MLR model in WA summer rainfall predictability. The comparisons between multiple regression analysis results and ARIMAX model results for the two rainfall stations are presented in figure 4. From these figures, it is observed that ARIMAX models' performances are relatively better than the multiple regression model's performances. Multiple regression models have shown some underestimations and overestimations from the actual rainfalls and it could not capture any extreme rainfall or drought events successfully. On the other hand, ARIMAX models were successful to predict some extreme rainfall magnitudes and few drought magnitudes and the model produced trend almost followed the same rainfall trend as of observed rainfall.
The statistical performance of rainfall prediction models developed for NWWA was evaluated and presented in table 10. From such an analysis, it is observed that the ARIMAX model’s performance is superior compared to the MLR model. Even though all these models showed the capability to predict four months in advance, however, the ARIMAX model showed a higher Pearson correlation ($r$) value of 0.77 while for MLR model, the correlation was found to be constrained within a value of 0.55.

Table 10. A comparison of MLR and ARIMAX model results in NWWA.

| Author | Region | Rainfall | Method      | Maximum Lagged Months | Pearson Correlation ($r$) |
|--------|--------|----------|-------------|-----------------------|--------------------------|

Figure 4. Comparison between observed and predicted rainfall for developed models.
The current study    NWWA    Summer    ARIMAX    4    0.65 to 0.68    0.74 to 0.77
The current study    NWWA    Summer    MLR    4    0.35 to 0.36    0.35 to 0.36
Hossain et al. [24]    NWWA    Spring    MLR    4    0.47 to 0.51    0.40 to 0.55

5. Conclusion
This study presents the comparison between MLR and ARIMAX model to predict summer rainfall using lagged climate indices for the Kimberly division of WA. Consideration of lagged indices resulted in the development of models that can predict the events several months advance. This is beneficial in terms of stakeholder’s perspective as it can offer enough time to anticipate any calamities and getting prepared for it. From MLR and ARIMAX analysis, WIO_{Aug} - SOI_{May} model was selected for Bidyadanga, and WIO_{Aug} - SOI_{June} was selected for Gogo Station. Both the models showed the highest correlation ($r$) with four months lagged period. In both calibration and validation periods, the ARIMAX model resulted in a higher Pearson correlation ($r$) values for both the stations (0.68 and 0.65 for Bidyadanga and Gogo station, respectively). Similarly, ARIMAX models’ $d_r$ values were found as close to 0.7 for both the stations, which depicts the fitness of the models as good. This was substantially higher than the reported $d_r$ value of 0.6 in MLR models. Furthermore, ARIMAX models were quite successful in predicting some of the extreme rainfall events and droughts, while MLR models failed to do so. This justified the combined effect of SOI and WIO in the ARIMAX model as capable to forecast long term summer rainfall for the studied region. However, none of the models were able to predict all the extreme cases, therefore, further studies are encouraged to assess the inherent non-linear relationship between the climate indices and rainfall for the investigated sites.

6. References
[1] Goddard L, Mason S J, Zebiak S E, Ropelewski C F, Basher R and Cane M A 2001 Current approaches to seasonal to interannual climate predictions Int J Climatol 21(9) 1111-52
[2] Anderson J, Dool H V D, Barnston A and Chen W 1999 Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction Bull Am Meteorol Soc 80(7) 1349
[3] Schepen A, Wang Q and Robertson D 2012 Evidence for using lagged climate indices to forecast Australian seasonal rainfall J Clim 25(4) 1230-46
[4] Momeni P and Naill P 2009 Time series analysis model for rainfall data in Jordan: case study for using time series analysis Am J Environ Sci 5(5) 599
[5] Zakaria S, Al-Ansari N, Knutsson S and Al-Badrany T 2012 ARIMA models for weekly rainfall in the semi-arid Sinjar district at Iraq J of Earth Sci & Geotech Engg 2(3) 25-55
[6] Mishra A and Desai V 2005 Drought forecasting using stochastic models Stoch Environ Res Risk Assess 19(5) 326-39
[7] Chiwell F H, Piechota T C, Dracup J A and McMahon T A 1998 El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting J Hydrool 204(1) 138-49
[8] Lungu E 1993 Detection of changes in rainfall and runoff patterns J Hydrol 147(1-4) 153-67
[9] Tularam G 2010 Relationship between El Niño southern oscillation index and rainfall (Queensland, Australia) Int J Sustain Dev Plan 5(4) 378-91
[10] Kamruzzaman M, Beecham S and Metcalfe A 2013 Climatic influences on rainfall and runoff variability in the southeast region of the Murray-Darling basin Int J Climatol 33(2) 291-311
[11] Kamruzzaman M, Beecham S and Metcalfe A V 2017 Changing patterns in rainfall extremes in South Australia Theor Appl Climatol 127(3-4) 793-813.
[12] England M H, Ummenhofer C C and Santos A 2006 Interannual rainfall extremes over southwest Western Australia linked to Indian ocean climate variability J Clim 19(10) 1948-69
[13] Feng J, Li J and Li Y 2010 Is there a relationship between the SAM and southwest Western Australian winter rainfall? *J Clim* **23**(22) 6082-89

[14] Islam F and Imteaz M A 2019 Development of prediction model for forecasting rainfall in Western Australia using lagged climate indices *Int J Water* **13**(3) 248-68

[15] Ummenhofer C C, Sen Gupta A, Pook M J and England M H 2008 Anomalous rainfall over southwest Western Australia forced by Indian ocean sea surface temperatures *J Clim* **21**(19) 5113-34

[16] Fierro A O and Leslie L M 2013 Links between central west Western Australian rainfall variability and large-scale climate drivers *J Clim* **26**(7) 2222-46

[17] Islam F and Imteaz M A 2020 Use of teleconnections to predict Western Australian seasonal rainfall using ARIMAX model *Hydrol* **7**(3) 52

[18] Nicholls N 2010 Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958–2007 *Clim Dyn* **34**(6) 835-45

[19] Risbey J S, Pook M J, McIntosh P C, Wheeler M C and Hendon H H 2009 On the remote drivers of rainfall variability in Australia *Mon Weather Rev* **137**(10) 3233-53

[20] Belsley D A, Kuh E and Welsch R E 1980 *Regression Diagnostics* John Wiley & Sons New York

[21] Marquardt D W 1980 You should standardize the predictor variables in your regression models *J Am Stat Assoc* **75**(369) 87-91

[22] Field, A 2013 *Discovering statistics using SPSS* Sage publications London UK

[23] Box G E and Pierce D A 1970 Distribution of residual autocorrelations in autoregressive-integrated moving average time series models *J Am Stat Assoc* **65**(332) 1509-26

[24] Dutta D, Welsh W D, Vaze J, Kim S S and Nicholls D 2012 A comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in eWater source. *Water Resour Manag* **26**(15) 4397-4415

[25] Willmott C J, Robeson S M and Matsuura K 2012 A refined index of model performance *Int J Climatol* **32**(13) 2088-9

[26] Hossain I, Rasel H, Imteaz M A and Mekanik F 2018 Long-term seasonal rainfall forecasting: efficiency of linear modelling technique. *Environ Earth Sci* **77**(7) 280