

Application of Lagged Climate Indices for Forecasting Autumn Rainfall in South Coast region of Western Australia using ARIMA Model

F. Islam, M.A. Imteaz and H.M. Rasel

*Department of Civil & Construction Engineering, Swinburne University of Technology, Melbourne, Australia
Email: farhanaislam@swin.edu.au*

Abstract: Climate indices are very effective predictors for forecasting seasonal rainfall. For any rainfall forecasting approach, it is necessary to understand the behavior of potential climate indices with rainfall variability. At present, rainfall forecasting using climate indices is one of the most reliable method to predict rainfall variability in many parts of the world. But, most of the studies have concentration on concurrent relationship between climate indices and seasonal rainfall. While a very few studies were conducted on concentration of lagged relationship between climate indices and seasonal rainfall. This study explores the significant correlation among lagged climate indices with autumn rainfall for south coast of Western Australia. As several climate indices such as Indian Ocean Dipole (IOD), El Nino Southern Oscillation (ENSO), Southern Oscillation Index (SOI), Southern Annular Mode (SAM), Blocking highs, Enso Modoki Index (EMI) are responsible for rainfall variability in Australia; therefore, this study evaluates the major climate indices and their interaction in generating rainfall variability in south coast of Western Australia. In the south coast part of Western Australia, two stations (Albany and Mount Barker) were considered for this study. From the single statistical correlation analysis, it was found that DMI (IOD indicator), SOI, Nino3.4, Nino3 and Nino4 (ENSO indicators) have significant correlation with autumn rainfall in Albany and Mount Barker. However, for these two stations EMI did not show any significant correlation with autumn rainfall. A time series analysis approach (Auto Regressive Integrated Moving Average-ARIMA) was conducted using climate indices, which showed significant correlation with autumn rainfall. In ARIMA model, lagged DMI-Nino3; lagged DMI-Nino4; Lagged DMI-Nino3.4 and Lagged DMI-SOI were chosen as independent variables (predictors) as it has showed significant correlation. From the ARIMA model analysis, it was evident that lagged DMI-Nino3 models showed highest predictability that is 56% and 33% for Albany and Mount Barker respectively. Finally, among those models, statistically significant models that showed high performance in predictability were suggested to forecast long-term rainfall for this region.

Keywords: *Climate Indices, rainfall, correlation, ARIMA*

1. INTRODUCTION

Rainfall is the most important hydro-climate phenomenon that plays a vital role for agricultural production, land management, risk management and water management systems of any region due to its long lasting social, economic and environmental impacts. A significant change in rainfall pattern could arise certain problems such as flood, bushfire, drought, stream-flow discharge reduction, reduced agricultural production, increased demand for water supply. Therefore, having a reliable rainfall forecasting several months ahead is important as it can help in decision making and set prior strategies to overcome the aforementioned problems during crisis period. In addition, a reliable forecasting system is also essential for farmers, which may provide necessary information to help in setting strategies for future crop management as well as for water management systems.

Many researches were conducted to investigate Australian rainfall mechanism. To understand this mechanism, researchers evaluated the relationships among climate indices and Australian seasonal rainfall (Chiew *et al.*, 1998; Ramsey *et al.*, 2008; Tularam, 2010). From the detailed study, it was evident that Australian rainfall largely varies with the interaction among climate indices generated in Southern, Pacific and Indian oceans (Chowdhury and Beecham, 2013; Risbey *et al.*, 2009). However, their influences vary depending on the different seasons and locations. Rainfall in different location is controlled by different climate indices generated within the region and more than one climate indices can be responsible for rainfall variability. Therefore, knowledge of different climate indices and their influences on specific location is highly important.

For Australia, it was found that Pacific Ocean SST anomalies highly influence rainfall in tropical and eastern regions while Indian Ocean SST anomalies were influential for southern and western regions. Especially, IOD and SAM were found as influential for south-eastern Australian rainfall (Rasel *et al.*, 2016, Risbey *et al.*, 2009) and Western Australian rainfall (Islam *et al.*, 2016; Ummenhofer *et al.*, 2008). Meanwhile, Blocking highs is deemed as responsible for southern Australian rainfall (Tibaldi *et al.*, 1994). ENSO Modoki and Madden Julian Oscillation (MJO) affects north-western and northern Australian rainfall (Marshall and Hendon, 2014; Taschetto and England, 2009). Among all these indices, ENSO grouped indices were found as the major climate indices for all over Australia (Montazerolghaem *et al.*, 2016).

Many researchers have tried to evaluate the relationship among climate indices and rainfall in Australia (Kirono *et al.*, 2010; Montazerolghaem *et al.*, 2016). Most of these studies were mainly focused on concurrent relationship between rainfall and climate indices. A very limited number of studies were conducted on evaluating lagged relationships among seasonal rainfall and climate indices (Schepen *et al.*, 2012). However, it is necessary to evaluate the lagged relationship because a concurrent relationship is the relationship between daily/monthly climate indices to daily/monthly rainfall. It is not suitable for predicting future seasonal rainfall. On the other hand, a lagged relationship is the relationship where climate indices has the delayed response to the seasonal rainfall after a period. As a result, it can predict future seasonal rainfall with a lagged period of several month advance which offer flexibility and beneficial for disaster management, emergency and evacuation planning.

From the previous literature, it was apparent that ENSO is the most influential climate index for Australia (Montazerolghaem *et al.*, 2016; Risbey *et al.*, 2009) and IOD mostly influences Western Australian rainfall (Fierro and Leslie, 2013; Ummenhofer *et al.*, 2008). So, these climate indices can be used as potential predictors for autumn rainfall. In this study, Bivariate correlation analysis and ARIMA method were conducted to evaluate the correlation and rainfall predictability for two rainfall stations from south coast of Western Australia.

2. DATA COLLECTION AND STUDY AREA

Western Australia (WA) is the largest state of Australia with 2529875 square kilometres land area. It has around 2.6 million inhabitants, which is almost 11% of the total population of Australia. The southwest division of WA has been identified as the most populous region as nearly 90% of WA population lives there.

In this study, two rainfall stations from south coast of Western Australia have been selected. Rainfall stations were selected considering 100years continuous rainfall data with minimal missing values. The collected data period was from January 1916 to December 2015. Autumn rainfall (March-April-May) data was used for this study. The detailed geographical location of the rainfall stations and study area are illustrated in Figure. 1. Red star marks in Figure. 1 showed the geographical location of selected rainfall stations.



Figure 1. Geographical location of rainfall stations and study area

From the available literature, it can be clearly seen that ENSO grouped indices namely SOI, Nino3.4, Nino3, Nino4, EMI and DMI have great influence on Western Australia rainfall (Fierro & Leslie, 2013; Risbey *et al.*, 2009; Ummenhofer *et al.*, 2008). So, these climate indices have been chosen as potential predictors to forecast WA seasonal rainfall. The collected data period for each climate index was from January 1915 to December 2015. A description of rainfall stations and climate indices can be found in Table 1(a) and 1(b)

Table 1(a) Description of the geographical location and recorded data of eight rainfall stations

Station Number	Station	Station District Name	Latitude	Longitude	Elevation	Data Source
9500	Albany	South Coast	35.03° S	117.88° E	3m	www.bom.gov.au/climate/data/
9581	Mount Barker	South Coast	34.63° S	117.64° E	300m	www.bom.gov.au/climate/data/

Table 1(b) Description of the Climate Indices

Climate Indices	Region	Data Set	Data Source
SOI	Pacific Ocean	Jones	http://climexp.knmi.nl/
Nino3.4, Nino3 and Nino4	Pacific Ocean	HadISST	http://climexp.knmi.nl/
EMI	Pacific Ocean	HadISST	http://www.jamstec.go.jp/frcgc/research/d1/iod/modoki
DMI	Indian Ocean	HadISST	http://climexp.knmi.nl/

3. METHODOLOGY

3.1. Auto-Regressive Integrated Moving Average (ARIMA) Modelling

Several different approaches are used in time series modelling. Statistical modelling namely, moving average, exponential smoothing and ARIMA are used to predict future values using their past values. In this study, ARIMA modelling was used as a linear combination of its past values and also independent variables can be included if they have significant relationship with dependent variables.

ARIMA is the dual combination of auto-regressive (AR) and moving average (MA) models. In the identification stage, data set was checked whether data are stationary or not? If the data found non-stationary than first order differencing conducted to make the data stationary from non-stationary. The general notation of ARIMA is given below:

$$\text{ARIMA}(p, d, q)*(P, D, Q)$$

Where, (p, d, q) is the non-seasonal part and (P, D, Q) is the seasonal part; p= non-seasonal auto-regressive order, d= non-seasonal differencing, q= non-seasonal moving average; P= seasonal auto-regressive, D= seasonal differencing, Q = seasonal moving average. In this study, only non-seasonal part was considered.

Inclusion of any independent variables in ARIMA model is termed as ARMAX model (Box and Pierce, 1970; Dutta *et al.*, 2012). The notation for ARMAX model is ARMAX (p, q, b). Where, p= autoregressive, q= moving average and b= independent variables. The mathematical expression for ARMAX model is given below:

$$X_t = \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-1} + \sum_{i=1}^b \eta_i d_{t-i} \tag{1}$$

Here, $\varphi_1, \dots, \varphi_p$ and $\theta_1, \dots, \theta_q$ are the AR and MA parameters; $\varepsilon_t, \varepsilon_{t-1}$ are white noise error and η_1, \dots, η_b are the parameters of the independent variables input d_t and t is the time period.

3.2. Statistical Analysis for Model Performances

Statistical parameters namely Pearson Correlation Coefficient (R), Co-efficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Willmott Index of Agreement (d) have been evaluated for the fitness of the developed model. Willmott (1981) introduced index of agreement (d) as a method of computing degree of error and the fitness of the developed model. The value of d ranges from 0 to 1. A value of near 0 indicates not a perfect model while value near to 1 indicates a perfect model. Equation for index of agreement (d) is given below:

$$d = 1 - \left(\frac{\sum(\text{Predicted} - \text{Observed})^2}{\sum(|\text{Predicted} - \text{MeanObserved}| + |\text{Observed} - \text{MeanObserved}|)^2} \right) \tag{2}$$

Initially, Bivariate correlation analysis was performed. Only correlations significant at 1% and 5% level, were used for further analysis. In this study, 100 years (rainfall and climate indices) data were divided into two stages; calibration stage (1916-1985) and validation stage (1986-2015). For dividing the data set for calibration and validation stages, a ratio of 2/3 and 1/3 was followed which also satisfied the World Meteorological Organization (WMO) standard verification system requirement of minimum 20 years' data (Ferranti, 2012; Vaze *et al.*, 2012). All these statistical analyses were performed using *IBM Statistics 24* software.

4. RESULT AND DISCUSSION

For this study, lagged monthly values (June_{n-1} to February_n) of selected climate indices were considered to evaluate their correlation with autumn rainfall. Where n is the year of predicted rainfall year and $(n-1)$ is the immediate previous year. Bivariate correlation analysis was performed to evaluate the correlation between autumn rainfall and climate indices for selected rainfall stations of WA. Results are presented at Table 2.

Table 2 Pearson Correlation (R) of Lagged Climate Indices with Autumn Rainfall

Significant Climate Indices	Rainfall Stations	
	Albany (Pearson Correlation, R)	Mount Barker (Pearson Correlation, R)
DMI _{Oct}	-0.309**	-0.242*
DMI _{Nov}	-0.375**	----
DMI _{Dec}	-0.266*	----
DMI _{Jan}	----	----
DMI _{Feb}	----	----
SOI _{Oct}	----	----
SOI _{Nov}	0.254*	----
SOI _{Dec}	0.401**	----
SOI _{Jan}	0.269*	----
SOI _{Feb}	0.346**	0.299*
Nino3.4 _{Oct}	----	----
Nino3.4 _{Nov}	-0.267*	-0.254*
Nino3.4 _{Dec}	-0.251*	----
Nino3.4 _{Jan}	-0.352**	-0.302*
Nino3.4 _{Feb}	-0.382**	-0.366**
Nino3 _{Oct}	-0.253*	-0.243*
Nino3 _{Nov}	-0.329**	-0.317**
Nino3 _{Dec}	-0.301*	-0.264*
Nino3 _{Jan}	-0.348**	-0.291*
Nino3 _{Feb}	-0.365**	-0.360**
Nino4 _{Oct}	----	----
Nino4 _{Nov}	----	----
Nino4 _{Dec}	----	----
Nino4 _{Jan}	-0.332**	-0.282*
Nino4 _{Feb}	-0.327**	-0.299*
EMI _{Oct}	----	----
EMI _{Nov}	----	----
EMI _{Dec}	----	----
EMI _{Jan}	----	----
EMI _{Feb}	----	----

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

From the analyses it can be clearly seen that except EMI all the remaining climate indices showed low to high significant correlation with autumn rainfall with five months (October-February) lagged period. SOI (Sea level Pressure based ENSO) showed significant positive correlation with autumn rainfall (values from +0.25 to +0.40); Nino3.4, Nino3, Nino4 (Sea Surface Temperature based ENSO) have showed significant negative correlations (values from -0.25 to -0.38); Similarly, DMI showed negative correlations (value from -0.27 to -0.38) with autumn rainfall. All these results are consistent with the previous findings of Fierro and Lesile, 2013; Ashok *et al.*, 2007, and Risbey *et al.*, 2009. In addition, EMI did not show any correlation with rainfall, which also aligned with the findings of Taschetto and England, 2009 who found that only northwestern Australian autumn rainfall is correlated with El Nino Modoki (EMI).

From the outcomes of single correlation analyses, several model sets were developed for these two rainfall stations. These model sets were the combination of lagged ENSO-DMI and lagged SOI-DMI considering lagged months October to February. After model set development, ARIMA test were conducted for the time periods (1916-1985) for the selected rainfall stations. In ARIMA model, lagged ENSO-DMI and SOI-DMI were chosen as input variables (predictors) as it showed significant correlation in bivariate correlation analyses. Furthermore, in ARIMA model input variable can be included as independent variable if they show significant correlation with dependent variable ("IBM SPSS Forecasting 22," 1989, 2013). Moreover, as rainfall is non-stationary, therefore first order differencing has been conducted to make the data stationary (first order differencing is adequate for linear trend). Hence, climate indices data are stationary so no need to do differencing for these data. In this study, only non-seasonal part is used for rainfall predictability. So, ARMAX (1,1,1) (1,1,0)(1,1,0) is the selected model for the two rainfall stations.

From the selected ARIMA model analysis, it was evident that in calibration period lagged DMI-Nino3 model outperformed lagged DMI-SOI, lagged DMI- Nino4 and lagged DMI-NINO3.4 models. Lagged DMI-Nino3 can predict rainfall up to 56% and 33% for Albany and Mount Barker respectively. For this two rainfall stations, MAPE (mean absolute percentage error), Normalized BIC (Bayesian Information Criterion) and DF (Degree of freedom) were calculated in ARIMA Model. A summary of best ARIMA model is illustrated in Table 3.

Table 3. Summary of the ARMAX model Statistics

Region	Station	Model Type	Model Input			Number of Predictors	R	R ²	MAPE	Normalized BIC	DF
			Rainfall	Predictor 1	Predictor 2						
South Coast	Albany	ARMAX	(1,1,1)	(1,1,0)	(1,1,0)	2 (DMI _{Nov} , Nino3 _{Feb})	0.75	56%	17.93	6.28	16
	Mount Barker	ARMAX	(1,1,1)	(1,1,0)	(1,1,0)	2 (DMI _{Oct} , Nino3 _{Feb})	0.51	33%	22.59	6.19	16

After the completion of model development for calibration period using ARIMA method, a validation test (time period 1986-2015) was conducted. A validation test was conducted to evaluate the accuracy of the developed models for rainfall predictability. In validation tests, statistical parameters namely: R, R², RMSE, MAE and Willmott index of agreement (d) were calculated. A detail summary of the method is illustrated in Table 4 for both the calibration and validation stages.

Table 4. Summary of the ARMAX Model Considering both Calibration and Validation Period

Method	Station	Predictor	Calibration Period					Validation Period				
			R	R ²	RMSE	MAE	d	R	R ²	RMSE	MAE	d
ARMAX (1,1,1)(1,1,0)(1,1,0)	Albany	DMI _{Nov} &Nino3 _{Feb}	0.75	56%	16.98	12.99	0.71	0.56	32%	17.51	10.89	0.64
	Mount Barker	DMI _{Oct} &Nino3 _{Feb}	0.51	33%	17.27	12.53	0.66	0.64	42%	20.02	12.60	0.75

From the outcome of ARIMA/ ARMAX model, it is apparent that in validation stages rainfall predictability is 32% and 42% for Albany and Mount Barker respectively. Moreover, in both periods, RMSE and MAE are found lowest and index of agreement (d) was above 0.6, which also indicates that the developed ARIMA model has better forecasting ability. The comparisons between observed and predicted rainfall using ARIMA Models output are illustrated in Figure 2.

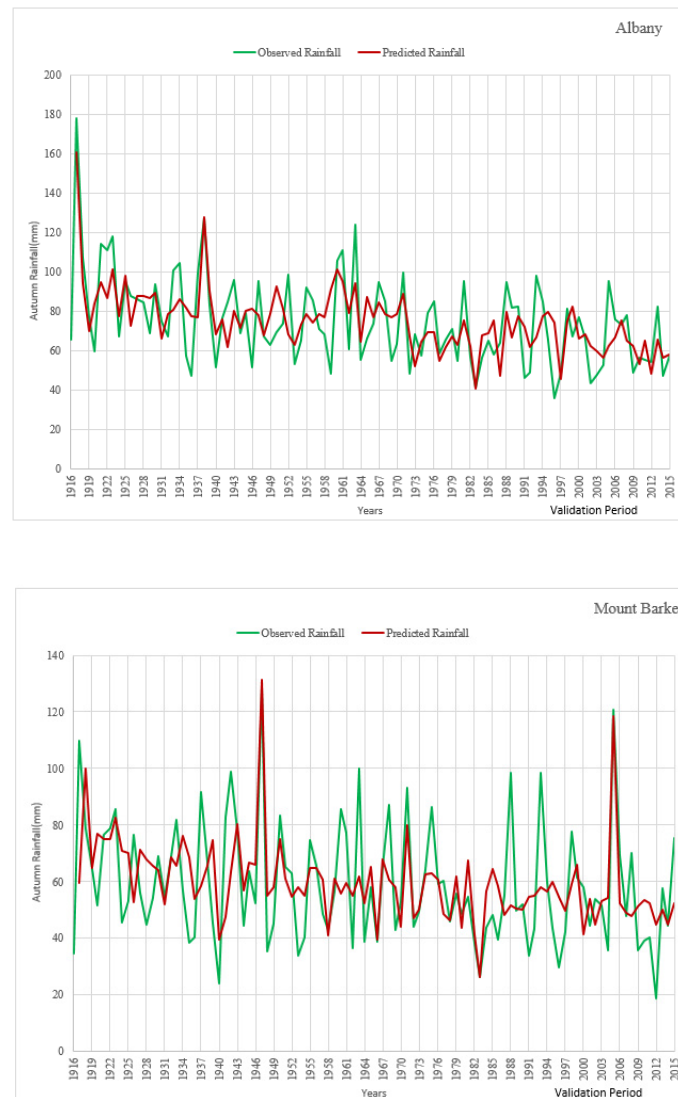


Figure. 2. Comparison between observed and predicted rainfall using ARIMA Model

From Figure. 2 it is can be clearly seen that, these developed models followed nearly same rainfall pattern as observed rainfall. However, developed models was not successful to pick all the extreme rainfall events but successful to predict some extreme rainfall cases (Year 1917, 1938 for Albany and Year 1947, 1971, 2005 for Mount Barker). Finally, it can be concluded that, ARIMA model using lagged DMI-Nino3 as input variable showed better performance to predict autumn rainfall in south coast of WA with better accuracy.

5. CONCLUSIONS

This study was concentrated on forecasting autumn rainfall in WA using climate indices with a lagged period of 1 to 9 months. Australian rainfall is mostly influenced by Pacific Ocean and Indian Ocean climate indices. However, their influence changes over different locations and seasons. Therefore, it is highly important to be familiar with different climate indices and their influences over specific locations and seasons. In this study, two rainfall stations from south coast region of WA were chosen. From the previous available literature, it was apparent that climate index namely: SOI, Nino3.4, Nino3, Nino4, EMI and DMI are the influential indices that are responsible for WA rainfall. So, these climate indices can be used as potential predictors of autumn rainfall. In this study, both Bivariate correlation and ARIMA techniques used as linear techniques to evaluate the correlation and rainfall predictability. Bivariate correlation analysis was conducted between rainfall and selected climate indices. After that, ARIMA method was conducted using different model sets and model that showed highest predictability along with minimum error chosen as best models for predicting WA rainfall. Lagged DMI-Nino3 model chosen as a best model for south coast region as it showed highest rainfall predictability with minimal errors.

After model development, a validation test was performed using ARIMA method. As a final outcome, in calibration periods, this model can predict autumn rainfall 56% and 33% and in validation period, rainfall predictability is 32% and 42% for Albany and Mount Barker respectively. Moreover, ARMAX(1,1,1)(1,1,0)(1,1,0) model was successful to predict some extreme rainfall cases and almost followed the same rainfall pattern as observed rainfall.

The ARIMA model performance was good but not quite satisfactory in predicting overall autumn rainfall for these two stations. So, for future study non-linear techniques can be used to increase the predictability. It is anticipated that application of non-linear technique for such forecasting analysis would provide better accuracy.

REFERENCES

- Ashok, K., Guan, Z., & Yamagata, T. (2003). Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophysical Research Letters*, 30(15).
- Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332), 1509-1526.
- Chiew, F. H., Piechota, T. C., Dracup, J. A., & McMahon, T. A. (1998). El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, 204(1), 138-149.
- Chowdhury, R. K., & Beecham, S. (2013). Influence of SOI, DMI and Niño3.4 on South Australian rainfall. *Stochastic environmental research and risk assessment*, 27(8), 1909-1920.
- Dutta, D., Welsh, W. D., Vaze, J., Kim, S. S., & Nicholls, D. (2012). A comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in eWater source. *Water resources management*, 26(15), 4397-4415.
- Ferranti, L. (2012). *Calibration and validation of seasonal forecasts*. ECMWF Seminar on Seasonal Prediction, Fierro, A. O., & Leslie, L. M. (2013). Links between central west Western Australian rainfall variability and large-scale climate drivers. *Journal of Climate*, 26(7), 2222-2246.
- IBM SPSS Forecasting 22 (Version 22). (1989, 2013).
- Islam, F., Imteaz, M. A., Boulomytis, V. G., & Rasel, H. (2016). Combined regression modelling of autumn rainfall in Western Australia using potential climate indices. Paper presented at the 37th Hydrology & Water Resources Symposium 2016: Water, Infrastructure and the Environment.
- Kirono, D. G., Chiew, F. H., & Kent, D. M. (2010). Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological processes*, 24(10), 1237-1247.
- Marshall, A., & Hendon, H. (2014). Impacts of the MJO in the Indian Ocean and on the Western Australian coast. *Climate dynamics*, 42(3-4), 579-595.
- Montazerolghaem, M., Vervoort, W., Minasny, B., & McBratney, A. (2016). Long-term variability of the leading seasonal modes of rainfall in south-eastern Australia. *Weather and Climate Extremes*.
- Ramsay, H. A., Leslie, L. M., Lamb, P. J., Richman, M. B., & Leplastrier, M. (2008). Interannual variability of tropical cyclones in the Australian region: Role of large-scale environment. *Journal of Climate*, 21(5), 1083-1103.
- Rasel, H., Imteaz, M., & Mekanik, F. (2016). Investigating the influence of Remote Climate Drivers as the Predictors in Forecasting South Australian spring rainfall. *International Journal of Environmental Research*, 10(1), 1-12.
- Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C., & Hendon, H. H. (2009). On the remote drivers of rainfall variability in Australia. *Monthly Weather Review*, 137(10), 3233-3253.
- Schepen, A., Wang, Q., & Robertson, D. (2012). Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *Journal of Climate*, 25(4), 1230-1246.
- Taschetto, A. S., & England, M. H. (2009). El Niño Modoki impacts on Australian rainfall. *Journal of Climate*, 22(11), 3167-3174.
- Tibaldi, S., Tosi, E., Navarra, A., & Pedulli, L. (1994). Northern and Southern Hemisphere seasonal variability of blocking frequency and predictability. *Monthly Weather Review*, 122(9), 1971-2003.
- Tularam, G. A. (2010). Relationship between El Niño southern oscillation index and rainfall (Queensland, Australia). *International Journal of Sustainable Development and Planning*, 5(4), 378-391.
- Ummerhofer, C. C., Sen Gupta, A., Pook, M. J., & England, M. H. (2008). Anomalous rainfall over southwest Western Australia forced by Indian Ocean sea surface temperatures. *Journal of Climate*, 21(19), 5113-5134.
- Vaze, J., Jordan, P., Beecham, R., Frost, A., & Summerell, G. (2012). Guidelines for rainfall-runoff modelling. Australian Government Department of Innovation, Industry, science and Research.
- Willmott, C. J. (1981). On the validation of models. *Physical geography*, 2(2), 184-194.