

Analysing the Effect of Lagged Climate Indices on Rainfall Predictability for Western Australia's North Coast Region

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: In last few decades, Australia experiences several extreme floods, bushfires and droughts. The consequences of these natural calamities have great social, environmental and economic impacts. Therefore, having a reliable seasonal forecasting of rainfall is highly important to withstand these situations as rainfall is mostly correlated with extreme flood, bushfire and drought incidents. Over decades, scientists and researchers are looking for reliable forecasting systems, which could provide definite predictions of rainfall that is essential for water management, agricultural production and flood management systems and mitigating such destructive situations by proper planning and management. Rainfall forecasting using climate indices has become one of the best ways to predict rainfall variability in many parts of the world. This strategy has also been proved as an effective tool to determine such variability. As Australia is affected by many different weather systems, several studies have been conducted to identify potential climate indices like Indian Ocean Dipole (IOD), El Nino Southern Oscillation (ENSO), Southern Annular Mode (SAM), Blocking highs, MODOKI and their interactions with rainfalls in different parts of Australia. Different regions of Australia have been exposed to different climate indices, so it is often recommended to study different regions separately to understand the effects of climate indices and their effectiveness to predict long-term rainfall. This study explores the seasonal variability of rainfall in Western Australia (WA) considering the effects of dominant climatic indices responsible for this region and develop a rainfall model for predicting Western Australia seasonal rainfall using potential climate indices. Very few studies have ever been found which had a primary concentration in determining rainfall variability in WA. Moreover, most of these studies has evaluated concurrent relationship between rainfall and climate indices. 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 north coast region of Western Australia (NCWA). To do so, two stations (Mingenew and Northampton) from NCWA were considered for this study. From the single statistical correlation analysis, it was found that DMI (IOD indicator), SOI, Nino3.4 and Nino4 (ENSO indicators) and ENSO Modoki Index (EMI) have shown significant correlation with autumn rainfall for these two stations. Several multiple regression analysis were conducted using lagged ENSO and DMI and lagged DMI and EMI as a combined predictor of autumn rainfall. Results showed that multiple regression has significantly increased the correlation between autumn rainfall and climate indices than single correlation analysis. Lagged DMI-Nino4 model showed highest correlation (0.38) for Mingeneew and (0.44) for Northampton. Developed models were tested with out of sample dataset and statistically significant models were suggested to forecast long-term rainfall for this region.

Keywords: *Climate indices, rainfall, multiple regression, correlation*

1. INTRODUCTION

Rainfall forecasting is an important issue, which can address valuable strategies to withstand climate challenges such as flood, drought and bushfire. Effective and reliable forecasting models can help in disaster management, water management, agricultural production, flood management systems and hence can play major role in decision making approach during crisis period. Rainfall forecasting can be done statistically and dynamically (Goddard *et al.*, 2001). Statistical methods are quite straightforward but need reliable sources for data extraction. Empirical relationships between observed variables are investigated through statistical tests and finally prediction equations can be developed if certain statistical criteria are met. For statistical methods, availability of uninterrupted data is required. On the other hand, dynamical forecast involves numerical simulations, which are comparatively complex to analyses and expensive to implement (Anderson *et al.*, 1999; Schepen *et al.*, 2012). Moreover, the effectiveness of prediction models developed through complex dynamic methods has often found as questionable. This is due to the failure of complex dynamic methods to outperform simple statistical prediction models (Halide and Ridd, 2008). Therefore, simple statistical climate prediction models often become the first choice to many researchers due to its simplicity, cost effective and easy to implement characteristics.

Over years, many studies were conducted in order to evaluate relationships between climate indices and Australian rainfall. Among many, few studies were conducted on entire Australia, while the rest were conducted on specific regions within Australia. Studies that considered the entire Australia includes: (Cai *et al.*, 2011; Schepen *et al.*, 2012). On the other hand, detailed studies were conducted on several parts of Australia, where an insight on local climate indices were used to explain the rainfall variability. Such studies were conducted on rainfall behaviour of Southwest Western Australia (Feng *et al.*, 2010; Ummenhofer *et al.*, 2008), South Australia (Cai *et al.*, 2011; Neville Nicholls, 2010; Tozer, 2015), Queensland (Abbot and Marohasy, 2012), Western Australia (Fierro and Leslie, 2013; Islam *et al.*, 2016), Southeast Australia and East Australia (Mekanik and Imteaz, 2014; Murphy and Timbal, 2008).

Most of these studies were mainly focused on concurrent relationships between rainfall and climate indices. But, evaluating the lagged relationship among the climate indices and rainfall also requires special attention. This is because concurrent relationship does not offer enough flexibility and sufficient time for precautionary measures and efficient disaster management strategies. On the other hand, lagged relationship can predict future seasonal rainfall several months' advance which could be useful for sustainable use of land and water resources, disaster management, emergency and evacuation planning.

From current literatures, it was evident that very limited number of studies were conducted on lagged relationships between seasonal rainfall and climate indices (e.g. Drosdowsky and Chambers, 2001; Hasan and Dunn, 2012; Schepen *et al.*, 2012). Hasan and Dunn (2012) investigated one-month lag relationship between climate indices and Australian rainfall while Drosdowsky and Chambers (2001) investigated the predictability of Australian rainfall using 1 to 3 months lagged predictors for different seasons and locations. These findings were further strengthened by Schepen *et al.* (2012), who showed the strength of lagged climate indices to predict seasonal rainfall in Australia. Most of these studies were limited to maximum 3 months lagged period and they did not show any combined application of climate indices for forecasting seasonal rainfall. Risbey *et al.* (2009) stated that only one climate indices cannot predict future rainfall precisely because other climate indices can also be responsible for rainfall generation. In this regard, very few have investigated combined influence of climate indices. However, their concentration of study was strictly limited to concurrent relationships only.

Rainfall in Western Australia varies according to the location investigated, as different indices plays their role at different locations. To date, most of the studies were focused on concurrent relationship between climate indices and rainfall however; no study has evaluated lagged relationship to predict future seasonal rainfall in NCWA. This study will evaluate the lagged relationship among climate indices and seasonal rainfall in NCWA. Moreover, rainfall prediction models will be developed by using linear techniques. To do so, two rainfall stations from NCWA has been considered. Effect of potential climate indices namely ENSO grouped indices and DMI will be analysed against WA rainfall as they showed significant contribution to the variability (Fierro and Leslie, 2013; Ummenhofer *et al.*, 2008).

2. DATA

2.1 Rainfall

Two rainfall station namely: Mingenew and Northampton from NCWA have been selected as a case study. These rainfall stations were considered based on their continuous rainfall data availability. 100 years

continuous monthly rainfall data was derived from the website of Bureau of Meteorology. Only Autumn rainfall (March to May) data from the year 1916 to year 2015 were considered in this study.

2.2 Climate Indices

Based on the available literature, it was apparent that El Nino Southern Oscillation (ENSO), ENSO Modoki Index (EMI) and Indian Ocean Dipole (IOD) are most interactive climate indices for Western Australia. Among these indices, ENSO and IOD exist in Pacific and Indian Ocean respectively. ENSO grouped indices are illustrated by two indicators: SLP and SST anomalies. ENSO grouped indices includes Southern Oscillation Index (SOI), Nino 3.4, Nino 3, Nino 4 and Nino 1.2. SOI is the measure of sea level pressure (SLP) difference between Darwin and Tahiti in the tropical western Pacific. Nino 3.4 is the average sea surface temperature (SST) anomalies in western Pacific bounded by 5°N to 5°S, from 170°W to 120°W. Nino3 is the average sea surface temperature (SST) anomalies in eastern Pacific bounded by 5°S to 5°N, from 90°W to 150°W; while Nino4 is the average SST anomalies in central Pacific bounded by 5°N to 5°S, from 150°W to 160°E. The difference between El-Nino events and traditional events, which have maximum warming in central Pacific and east Pacific, is denoted as ENSO Modoki Index (EMI). Indian Ocean Dipole (IOD) is also measured by Dipole Mode Index (DMI). The variation of average Sea Surface Temperature generated between tropical western Indian Ocean bounded by 10°N to 10°S, from 50°E to 70°E and the tropical south-eastern Indian Ocean bounded by 10°S to equator, from 90°E to 110°E is known as DMI (Saji *et al.*, 1999). Apart from EMI, all other climate indices data was derived from Climate Explorer website (<http://climexp.knmi.nl/>). EMI data was obtained from the website: (<http://www.jamstec.go.jp/frcgc/research/d1/iod/modoki>).

3. METHODOLOGY

To establish relationship between climate indices and rainfall, linear technique has been used in this study. Simple linear relationship has been employed to determine significant correlation between lagged climate indices and rainfall. After that, multiple linear regressions technique was used to determine the predictability of rainfall using two or more climate indices. The methodology of this technique are explained in details in following sections.

3.1. Multiple Linear Regression

Multiple linear regression is the statistical measure that determine the best linear relationship between a dependent variable and several independent variables. For the evaluation of the best fit of the model, least square method is used. Dependent variable is also called outcome variable while independent variables are called Predictors.

The equation for this model is given below:

$$Y = a + b_1X_1 + b_2X_2 + e \quad (1)$$

Where, Y is the dependent variable (i.e. Autumn Rainfall in this study); X_1 is the first independent variable (i.e. lagged DMI) and X_2 is the second independent variable (i.e. lagged ENSO); b_1 and b_2 are the model coefficient. a is constant and e is error.

Three assumption check is necessary before conducting linear regression analysis; these are linearity, equal of variance and normality. In linearity check, linear relationship between dependent and independent variable is checked, in equal of variance the spread of the residuals are checked and in normality check data distribution is checked whether it is normally distributed or not. If selected variables satisfy all these assumption check, linear regression analysis can be proceeded.

3.2. Model Performances Analysis

For any developed regression model, it is recommended to check the goodness of the fit of the developed model. It can be checked by evaluating statistical tests namely: hypothesis testing (F-test), R squared and analyses of residuals. F-test is used to verify statistical significance and compare different models to evaluate and identify the model that best fitted the sample data. Another test called t-test is conducted to verify the parameters are statistically significant at 95% confidence level.

Another important evaluation for regression modelling is Multicollinearity and Durbin-Watson (DW) test. In regression modelling, when independent variable are highly correlated to each other is called Multicollinearity. Two indicators namely Variance Inflation Factor (VIF) and Tolerance (T) are used to identify Multicollinearity. Lin (2008) stated that, tolerance value less than 0.10-0.20 and VIF value greater than 5-10 indicates Multicollinearity. Durbin-Watson (DW) is another important parameter, which identify the autocorrelation

among the residuals. According to Field (2013) DW value lower than 1 or greater than 3 indicates autocorrelation. Several statistical parameters namely Pearson Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Willmott Index of Agreement (d) have been calculated to check the fitness of the developed regression models.

All these statistical analyses were conducted by using *IBM Statistics 24* software. Initially, single correlation analysis was performed. Results that showed significant correlation (1% and 5%) are considered for the multiple regression analysis. In this study, total data collection period for both rainfall and climate indices is 100 years. These data were divided into two stages namely: calibration stages and validation stages. Data dividation was done by following the ratio (2/3 for calibration and 1/3 for validation) Of World Meteorological Organization (WMO) standard verification system (Ferranti, 2012; Vaze et al., 2012).

4. RESULT AND DISCUSSION

At first three assumptions check (mentioned in section 3.1) were carried out to check the requirement of regression analysis. After that, a descriptive statistics for all the selected variables were created (Table 1).

Table 1. Descriptive Statistics for all variables

Variables	Data Size (N)	Minimum	Maximum	Mean	Standard Deviation	Variance
Mingenev	70	5.93	97.37	33.39	17.59	309.44
Northampton	70	9.50	116.60	37.66	21.59	466.12
DMI	70	-0.53	0.73	0.002	0.23	0.05
SOI	70	-2.66	2.27	-0.01	0.86	0.74
Nino3.4	70	-1.50	1.73	-0.04	0.68	0.47
Nino3	70	-1.54	2.00	-0.13	0.73	0.54
Nino4	70	-1.41	0.71	-0.07	0.46	0.21
EMI	70	-0.99	0.61	0.04	0.31	0.09

Single correlation analyses were conducted between climate index and selected rainfall stations. In this study, lagged monthly values (June_{n-1} to February_n, Here, n is the predicted rainfall year and n-1 is the previous year) of climate index and autumn rainfall (March-April-May) are considered to perform the analyses. From the analyses, it was evident that most of these climate indices showed significant correlation for maximum five months (Oct-Feb) lagged period. The outcome of single correlation analyses are presented in Table 2.

Table 2. Outcome of Single Correlation Analyses

Pearson Correlation	Rainfall Station	Month	Significant Climate Indices					
			DMI	SOI	NINO3.4	NINO3	NINO4	EMI
R	Mingenev	OCT	----	----	-0.254*	-0.244*	-0.245*	----
		NOV	-0.240*	0.242*	-0.265*	-0.273*	-0.254*	----
		DEC	----	----	-0.290*	-0.255*	-0.313**	-0.269*
		JAN	----	0.256*	-0.262*	----	-0.344**	-0.330**
		FEB	----	----	-0.320**	-0.245*	-0.366**	-0.262*
R	Northampton	OCT	----	----	----	-0.256*	----	----
		NOV	----	----	-0.255*	-0.273*	----	----
		DEC	----	----	-0.281*	-0.255*	-0.280*	----
		JAN	0.251*	0.252*	----	----	-0.287*	-0.236*
		FEB	0.375**	----	-0.298*	----	-0.330**	----

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

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

From the result of correlation analyses, it is clearly visible that, selected climate indices showed low to high significant correlation with autumn rainfall. This outcome has aligned with the previous findings of Fierro and Leslie, 2013; Ashok et al., 2007; Taschetto and England, 2009. Moreover, it is also visible that SLP based ENSO Index (SOI) showed low lagged significant correlation compared to SST based ENSO indices. This

outcome is consistent with the previous finding of Drosowsky and Chambers (2001). According to them, SOI has less predictability skill for Western Australian autumn rainfall than SST based ENSO indices.

Several model sets (Presented in Table 3) were developed using the outcome of single correlation analyses. These model sets were the combination of lagged DMI-Nino3.4, lagged DMI- Nino3, lagged DMI- Nino4 and lagged DMI-EMI. Lagged DMI- SOI model set was not considered as it showed less significant correlation.

Table 3. Model Set for Multiple Regression for Two Rainfall Stations

Model Set	Mingenew	Northampton
DMI- Nino3.4	Nov-Oct, Nov-Nov, Nov-Dec, Nov-Jan, Nov-Feb,	Jan-Nov, Jan-Dec, Jan-Feb, Feb-Nov, Feb- Dec, Feb- Jan
DMI-Nino3	Nov-Oct, Nov-Nov, Nov-Dec, Nov-Feb,	Jan-Oct, Jan-Nov, Jan-Dec, Feb-Oct, Feb-Nov, Feb- Dec,
DMI-Nino4	Nov-Oct, Nov-Nov, Nov-Dec, Nov-Jan, Nov-Feb	Jan-Dec, Jan-Jan, Jan-Feb, Feb-Dec, Feb- Jan, Feb-Feb
DMI-EMI	Nov-Dec, Nov-Jan, Nov-Feb	Jan-Jan, Feb-Jan

Multiple regression analyses and statistical measure such as Durbin-Watson, tolerance were calculated using these aforementioned model sets. From the results of multiple regression analyses it was found that all these model set are free from multicollinearity and auto correlation among the predictors. In addition, among all these model sets, DMI-Nino4 model showed highest significant correlation 0.38 and 0.44 for Mingenev and Northampton respectively. It is also evident that multiple regression has significantly increased the correlation between autumn rainfall and climate indices than single correlation analysis. This outcome is compatible with the lagged map of Schepen *et al.*, 2012, where they depicted that climate index Nino4 has low to positive evidence of predicting autumn rainfall in north coast division of WA. A summary of the best regression models developed in this study is presented in Table 4.

Table 4. Summary of the Selected Regression Model

Region	Station	Model	Constant	Coefficient	R	Durbin-Watson	Tolerance (T)	VIF
North Coast	Mingenew	DMI _{Nov} Nino4 _{Feb}	33.24	-7.19 -10.03	0.38	2.04	0.86	1.16
	Northampton	DMI _{Feb} Nino4 _{Feb}	36.42	+24.17 -9.40	0.44	2.07	0.92	1.08

After the model development in calibration period using multiple regression method, a validation test has been conducted. This test is required to verify the accuracy of the developed model for rainfall predictability. For both these calibration and validation period statistical parameters such as R, R², RMSE, MAE and Willmott index of agreement (d) were calculated. Summary of the result for both calibration and validation period is illustrated in Table 5.

Table 5. Summary of the Developed Model in both Stages (Calibration and Validation).

Region	Station	Model	Calibration Period				Validation Period			
			R	RMSE	MAE	d	R	RMSE	MAE	d
North Coast	Mingenew	DMI _{Nov} -Nino4 _{Feb}	0.38	16.49	12.49	0.48	0.44	17.31	11.87	0.56
	Northampton	DMI _{Feb} -Nino4 _{Feb}	0.44	19.64	14.45	0.55	0.48	15.09	10.90	0.58

From the outcome of these two stages, it was visible that in validation stages, correlation between rainfall and climate indices has been increased from 0.38 to 0.44 and 0.44 to 0.48 for Mingenev and Northampton respectively. Moreover, Willmott index of agreement (d) also increased in validation period which is above 0.5. On the other hand, RMSE and MAE value was lowest in both stages. A model with lowest error, highest 'R' and highest 'd' value is considered as a good model. From the analysis results, DMI-Nino4 model found as best model for predicting autumn rainfall for Western Australian north coast region. A comparison between observed and predicted rainfall are shown in Figure 1.

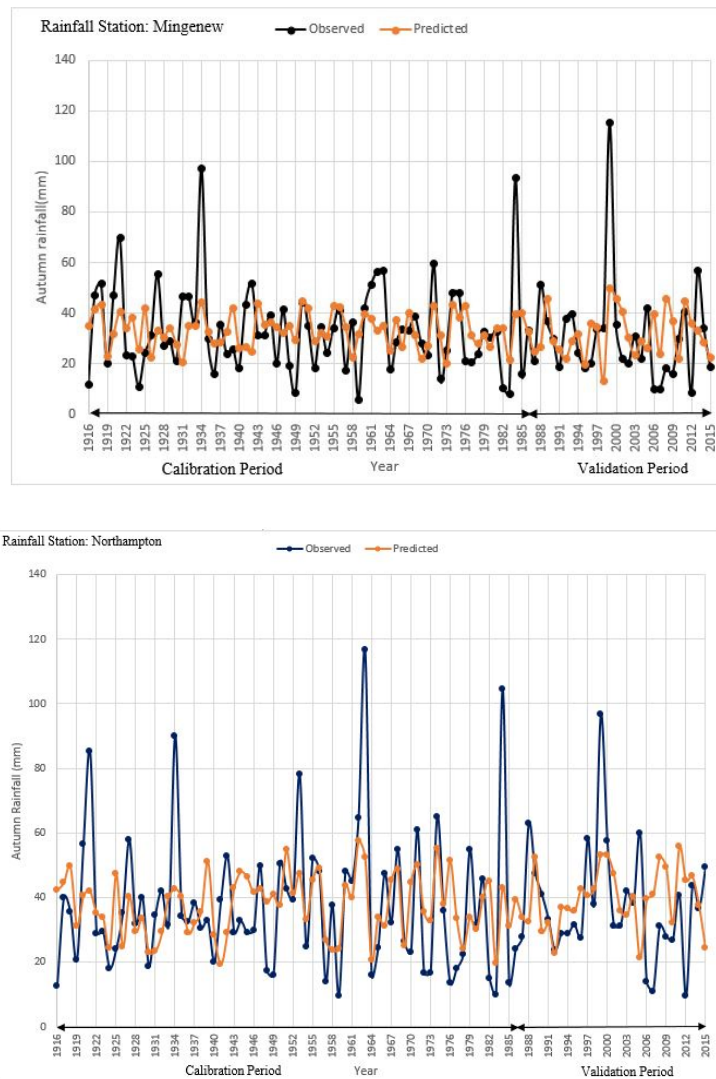


Figure 1. Comparison between Observed and Predicted Rainfall

Figure 1. depicts that, the developed model followed similar rainfall pattern as observed rainfall. However, unfortunately the developed model could not pick the extreme rainfall cases. Some overestimation of rainfall is also visible in the validation stages that might be due to Millennium Drought (Dijk *et al.*, 2013) occurred in Australia. Therefore, it can be said that the developed models can predict usual rainfall events, however it was not successful in predicting extreme rainfall cases.

5. CONCLUSIONS

Western Australia is the largest state of Australia and its rainfall varied according to its location. This rainfall variation is the result of different climate indices that influence rainfall in different location at different time. To date, most of the studies were focused on concurrent relationship between climate indices and rainfall however; no study has evaluated lagged relationship to predict future seasonal rainfall in NCWA. This study will evaluate the lagged relationship among climate indices and seasonal rainfall in WA. To accomplish the objective, single and multiple regression technique has been considered. Moreover, this will fill up the existing research gap that prevails in north coast of Western Australian rainfall predictability. Two rainfall stations namely: Mingenew and Northampton from NCWA have been chosen as a case study. Effect of potential climate indices namely ENSO grouped indices and DMI were analysed against WA autumn rainfall as they showed significant contribution to the variability. From the outcome of correlation analyses, it was apparent that most of these climate indices showed significant correlation up to five months (Oct-Feb) lagged period. After that, several model sets have been developed for multiple regression analyses. Model which showed highest Pearson correlation (R), highest Wilmott of Index (d), lowest RMSE; MAE and satisfy all the statistical

requirement selected as a best model for NCWA region. DMI-Nino4 model has been chosen as a best model, which fulfilled all these requirement. The performances of this model is satisfactory in terms of predicting autumn rainfall in normal situation but not for extreme situations like extreme rainfall and drought. This developed model failed to predict extreme rainfall and drought cases. This is may be the result of some other factors namely: temperature, wind, humidity and soil moisture, which also influences rainfall but was not considered in this model. In addition, this study evaluate the linear relationship between climate indices and rainfall only, and no non-linear relationship was considered. The application of non-linear techniques will be considered in future studies.

REFERENCES

- Abbot, J., & Marohasy, J. (2012). Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences*, 29(4), 717-730.
- Anderson, J., Van Den Dool, H., Barnston, A., & Chen, W. (1999). Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bulletin of the American Meteorological Society*, 80(7), 1349.
- Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112(C11).
- Dijk, A. I., Beck, H. E., Crosbie, R. S., Jeu, R. A., Liu, Y. Y., Podger, G. M., . . . Viney, N. R. (2013). The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, 49(2), 1040-1057.
- Drosowsky, W., & Chambers, L. E. (2001). Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *Journal of Climate*, 14(7), 1677-1687.
- Feng, J., Li, J., & Li, Y. (2010). Is there a relationship between the SAM and southwest Western Australian winter rainfall? *Journal of Climate*, 23(22), 6082-6089.
- Ferranti, L. (2012). Calibration and validation of seasonal forecasts. ECMWF Seminar on Seasonal Prediction, Field, A. (2013). Discovering statistics using IBM SPSS statistics: Sage.
- 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.
- Halide, H., & Ridd, P. (2008). Complicated ENSO models do not significantly outperform very simple ENSO models. *International Journal of Climatology*, 28(2), 219-233.
- Hasan, M., & Dunn, P. K. (2012). Understanding the effect of climatology on monthly rainfall amounts in Australia using Tweedie GLMs. *International Journal of Climatology*, 32(7), 1006-1017.
- 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.
- Lin, F.-J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality & Quantity*, 42(3), 417-426.
- Nicholls, N. (2010). Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958–2007. *Climate dynamics*, 34(6), 835-845.
- Quan, X., Hoerling, M., Whitaker, J., Bates, G., & Xu, T. (2006). Diagnosing sources of US seasonal forecast skill. *Journal of climate*, 19(13), 3279-3293.
- 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.
- Saji, N., Goswami, B., Vinayachandran, P., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751), 360-363.
- 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.
- Tozer, C. R. (2015). Utilising insights into rainfall patterns and climate drivers to inform seasonal rainfall forecasting in South Australia| NOVA. *The University of Newcastle's Digital Repository*.
- Ummenhofer, 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.