Analysing lagged ENSO and IOD as potential predictors for long-term rainfall forecasting using multiple regression modelling

F. Mekanik[a], M. A. Imteaz[a]

[a]Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Melbourne, VIC, Australia,
Email: fmekanik@swin.edu.au

Abstract: Several climate indices around Australia were found to have strong correlations with south-east Australian seasonal rainfalls. Any such correlation with lagged climate indices and seasonal rainfall afterwards can be used for forecasting long-term seasonal rainfall. In this study, long-term forecasting of Victorian spring rainfall has been investigated using lagged El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) indices using multiple regression analysis. Three stations of Buchan (east VIC), Malmsbury (Central VIC) and Kaniva (West VIC) were chosen as case study. Rainfall was classified according to El Nino, La Nina and Neutral years of ENSO, and also positive and negative years of IOD. It was discovered that categorizing rainfalls based on the years of ENSO and IOD do not have significant effect on its relationship with these climate modes. It was also found that the Pearson correlation coefficient between ENSO and Buchan in east Victoria is very weak; for Malmsbury and Kaniva ENSO indicators are showing higher correlations compared to Buchan. DMI effect is stronger in these two regions as well.

Using the non-classified rainfalls, correlation coefficient between spring rainfall at year n and Dec_{n-1}-Aug_n monthly values of ENSO and IOD indicators (Nino3.4, SOI and DMI) were calculated ("n" being the year for which spring rainfall is being predicted); It was discovered that only the three months of June, July and August of Nino3.4, SOI and DMI have significant correlation with spring rainfall. Several multiple regression models were investigated using lagged ENSO and IOD as potential predictors of spring rainfall; the models that satisfied the limits of statistical significance and multicollinearity were used to forecast spring rainfall three consecutive years in advance. Multiple regression analysis showed poor results in regards to forecasting ability in east Victoria, however it was able to forecast spring rainfall three consecutive years in advance for central and west Victoria with a correlation of 0.48 and 0.67 respectively.

Keywords: *Rainfall, El Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), multiple regression model*
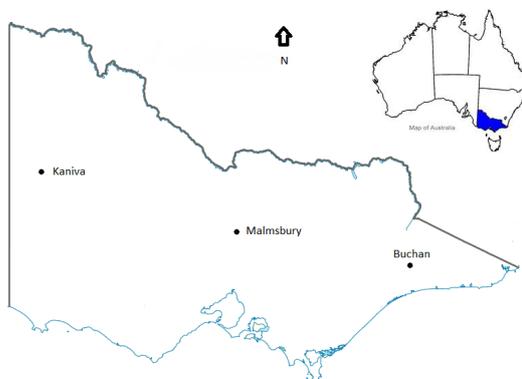
1.   INTRODUCTION

Rainfall is a product of complex global atmospheric phenomena and long-term prediction of rainfall remains a challenge for many years. Forecasting rainfall several months or seasons in advance can be beneficial for the management of water resources. Australian rainfall is highly variable both in space and time. It is believed that Australian rainfall is affected by several major climate patterns. The major drivers bringing rainfall over Australia which have been investigated by many researchers are, El Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD) and Southern Annular Mode (SAM) (Kirono et al., 2010). Many researchers have conducted different studies in different parts of Australia trying to establish the relationship between these climate modes and Australian rainfall. Some of these studies cover the whole Australia (Kirono et al., 2010; Meneghini et al., 2007; Risbey et al., 2009) while the others are more focused on a specific region like South West Western Australia (SWWA) (England et al., 2006; Ummenhofer et al., 2008), South Australia (Evans et al., 2009; Nicholls, 2010), South East Australia (SEA) and East Australia (Murphy and Timbal, 2008; Verdon et al., 2004). Other than the work of Keim and Verdon-Kidd (2009) which analysed the combined impact of ENSO and SAM on Victorian rainfall, other studies focused only on finding the relationship between rainfall and a single driver. According to Verdon et al. (2004) compared to other parts of Australia ENSO signals are relatively weak for Victoria. This study will further investigate this matter. On the other hand, the majority of studies on ENSO-rainfall relationships did not consider the effect of lagged climate modes on future rainfall predictions. According to Schepen et al. (2012) a strong relationship between simultaneous climate modes and rainfall does not essentially mean that there is a lagged relationship as well. Of the few studies focusing on the lagged climate–rainfall relationship one can mention Mekanik and Imteaz (2013), Abbot and Marohasy (2012), Drosdowsky and Chambers (2001), Kirono et al. (2010), Mekanik and Imteaz (2012), Schepen et al. (2012) and Mekanik et al. (2013). Thus, the objective of this study is to investigate the relationship of combined ENSO and IOD lags on Victoria's spring rainfall, as a case study. To achieve this objective multiple regression analysis (MR) is used. Three rainfall stations in Victoria, each representing a region is chosen as the case study. Model outputs were aimed to be deterministic forecast as opposed to probabilistic forecast.

2.   METHOD AND DATA

2.1   Data

Historical monthly rainfall data was obtained from the Australian Bureau of Meteorology website (BoM) (www.bom.gov.au/climate/data/). Three rainfall stations in Victoria, each representing a region is chosen as the case study (Figure.1). The stations were chosen based on their recorded length of data and having fewer missing values. Spring (September - November) rainfalls in millimeters were obtained from monthly rainfall data from January 1900 to December 2009.

El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) were chosen as rainfall drivers based on the previous studies (Kirono et al., 2010); ENSO is represented by two different types of indicators: the Southern Oscillation Index (SOI) which is a measure of Sea Level Pressure (SLP) anomalies between Darwin and Tahiti; and the Sea Surface Temperature (sst) anomalies in equatorial Pacific Ocean noted as Nino3 (5^oS -5^oN, 150°-90^oW), Nino3.4 (5° S -5^oN, 170°-120^oW) and Nino4 (5^oS -5^oN, 160°-150^oW) (Risbey et al., 2009). Nino3.4 and SOI which are the common indices in identifying El Nino/ La Nina years are used as ENSO indicators in this study. IOD is also a coupled ocean-atmosphere phenomenon in the equatorial Indian Ocean (Saji et al., 1999). A measure of IOD is the Dipole Mode Index (DMI) which is the difference in average SST anomalies between the tropical Western Indian Ocean (10^oS - 10^oN, 5^oO - 70^oE) and the tropical Eastern Indian ocean (10^oS - Equator, 90° - 110^oE) (Kirono et al., 2010). The climate indices data were obtained from Climate Explorer website (http://climexp.knmi.nl/). The data were divided in to two sets, from 1900-2006 for calibration of the models and the three years 2007-2009 were selected as the out-of-sample set to evaluate the generalization ability of the developed models. The data were normalized between the range of 1 and 0 using Eq. (1).

Figure 1. Map of the study area

Mekanik and Imteaz, Analysing lagged ENSO and IOD as potential predictors for rainfall forecasting

$$\bar{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

The models were evaluated using Mean Square Error (MSE), Mean Absolute Error (MAE), and Pearson correlation coefficient (r) which are widely used for prediction purposes; the models were further assessed using Willmott index of agreement *(d)* (Eq. 2)

$$d = 1 - \left(\frac{[\Sigma|\hat{y}_i - x_i|^2]}{[\Sigma(|\hat{y}_i - \bar{x}_1| + |x_i - \bar{x}_1|)^2]}\right) \tag{2}$$

where, $\hat{y}_i$ is the predicted value of the ith observation and x_i is the ith observation. The closer the *(d)* is to one the better the model has fitted the observations (Willmott, 1982).

2.2   Preliminary analysis

El Nino Southern Oscillation (ENSO) consists of two phases of warm SST anomalies (El Nino) and cold SST anomalies (La Nina) in the Equatorial Pacific Ocean; it is believed that El Nino brings below-average rainfall to Australia while La Nina is associated with above-average rainfall. IOD also has two phases of positive and negative IOD associating with below and above average rainfall respectively.

Meyers et al. (2007) classified the years of 1878-1998 into three ENSO categories (El Nino/ La Nina/ Neutral) and autonomously three IOD categories (IOD positive/IOD negative/Neutral); Later Umenhofer et al. (2009) extended the categorization to recent years. In this study, rainfall is categorized according to the years of ENSO and IOD based on the work of Umenhofer et al. (2009). The relationship between concurrent spring rainfall and the climate modes (ENSO/IOD) as well as the lag relationship between these phenomena are investigated; Pearson correlation was used to examine the strength and significance of the relationships. Table 1 shows the correlation coefficient of concurrent non-categorized spring rainfall with ENSO and IOD for three stations. It can be seen from this table that the correlation between ENSO and Buchan in east Victoria is very weak (r=-0.25). For Malmsbury and Kaniva ENSO indicators are showing higher correlations compared to Buchan. Also, DMI effect is stronger in these two regions as well.

Rainfall anomalies were calculated based on the categorization of Umenhofer et al. (2009) (Figure 2). The significance of median being different from zero is calculated using the Wilcoxon Signed Rank test (Table 2).

It can be seen from this table that for Buchan none of the categorization is showing significant difference on rainfall; meaning that different phases of ENSO/IOD does not

Table1. Pearson correlation (r) of spring climate indices and spring rainfall (1900-2006)

Indices	Buchan(East)	Malmsbury(Centre)	Kaniva(West)
Nino34(S-O-N)	-0.25[b]	-0.39[a]	-0.40[a]
SOI(S-O-N)	0.17	0.45[a]	0.43[a]
DMI(S-O-N)	-0.22[b]	-0.44[a]	-0.44[a]

a: correlation is significant at the 0.01% level
b: correlation is significant at the 0.05% level

have a significant effect on the amount of rainfall received in this region. For Malmsbury only the phases of El Nino-IOD positive, El Nino, IOD positive and La Nina phases have significant effect on rainfall; It can be seen that except La Nina the other phases are associated with lower than average rainfall for this region. For Kaniva, IOD positive and negative have a significance level of 0.050 which is acceptable but still high and El Nino has a significance level of 0.021. In General it can be concluded that the phases associated with drier than normal seasons affect rainfall more than the wet phases of ENSO and IOD; However, more stations need to be investigated in order to generalize this outcome for Victoria.

Table 2. Testing for significant difference of median from zero (Wilcoxon Signed Rank test)

Climate phase	EN,IOD+	EN	IOD+	Neutral	IOD-	LN	LN,IOD-
Buchan	0.063	0.213	0.638	0.851	0.327	0.715	0.499
Malmsbury	0.043*	0.013*	0.003*	0.674	0.401	0.046*	0.063
Kaniva	0.091	0.021*	0.050*	0.761	0.050*	0.192	0.063

*The significance level is 0.05
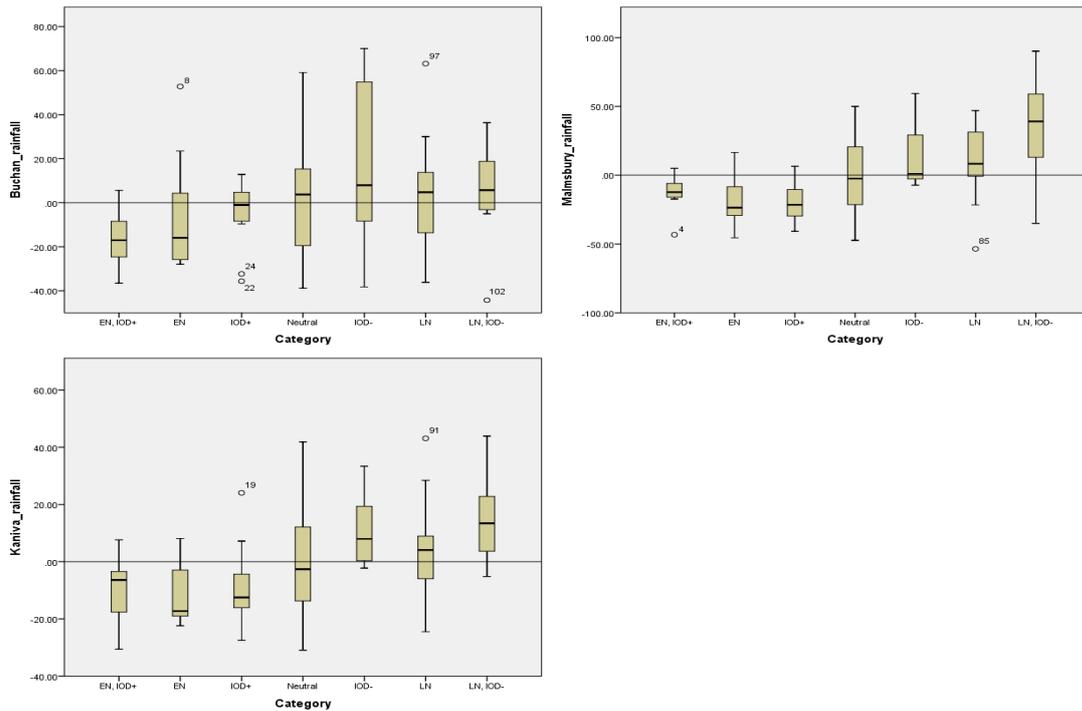
Figure 2. Rainfall anomalies for the different ENSO/IOD categories for east, centre and west VIC

2.3 Multiple regression analysis

Multiple regression analysis (MR) is a linear statistical technique that allows for finding the best relationship between a variable (dependent, predicant) and several other variables (independent, predictor) through the least square method. Multiple regression models can be presented by the following equation:

$$Y = a + b_1 X_1 + b_2 X_2 + c \tag{3}$$

where, Y is the dependent variable (spring rainfall), X_1 and X_2 are first and second independent variable respectively (lagged ENSO and IOD indicators), b_1 and b_2 are model coefficients of first and second independent variable respectively, a is constant, and c is the error.

Verifying the multicollinearity is an important stage in MR modeling. Multicollinearity occurs when the predictors are highly correlated which will result in dramatic change in parameter estimates in response to small changes in the data or the model. The indicators used to identify multicollinearity among predictors are tolerance (T) and variance inflation factor (VIF):

$$\text{Tolerance} = 1 - R^2, \quad \text{VIF} = \frac{1}{\text{Tolerance}}, \quad R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \tag{4}$$

where, R^2 is the coefficient of multiple determination, SST is the total sum of squares, SSR is the regression sum of squares and SSE is the error sum of squares. According to Lin (2008) a tolerance of less than 0.20–0.10 or a VIF greater than 5–10 indicates a multicollinearity problem.

Analysing the pattern of residuals is another method of evaluating the goodness-of-fit of the models. If any autocorrelation exists among the residuals then the models have not captured all the relationship there is between the inputs and the output; The criterion that can evaluate this is the Durbin-Watson (DW) test, which tests for serial correlations between errors. The test statistics have a range of 0 to 4, according to Field (2009) values less than 1 or greater than 3 are definitely matter of concern.

3. RESULTS AND DISCUSSION

In this study correlation between spring rainfall at year n and $\text{Dec}_{n-1}\text{-Aug}_n$ monthly values of ENSO and IOD indicators (Nino3.4, SOI and DMI) were calculated ("n" being the year for which spring rainfall is being predicted); It was discovered that only the three months of June, July and August of Nino3.4, SOI and DMI have significant correlation with spring rainfall (Table 3); This result is in accordance to the findings of Cheiw et al. (1998) and Verdon et al. (2004), substantiating that not only the highest correlations between

rainfall and climate indicators are obtained up to three month lags i.e. there is no further significant relationship after lag 3;

Table 3. Correlation of spring rainfall and lagged climate indices

	Nino34(Jun)	Nino34(Jul)	Nino34(Aug)	SOI(Jun)	SOI(Jul)	SOI(Aug)	DMI(Jun)	DMI(Jul)	DMI(Aug)
Buchan	-0.22[b]	-0.26[a]	-0.24[b]	-0.20[b]	---	---	-0.30[a]	---	---
Malmsbury	-0.22[b]	-0.22[b]	-0.29[a]	---	0.32[a]	0.30[a]	---	-0.30[a]	-0.31[a]
Kaniva	-0.32[a]	-0.32[a]	-0.36[a]	0.23[b]	0.33[a]	0.31[a]	---	-0.30[a]	-0.31[a]

a: correlation is significant at the 0.01% level
b: correlation is significant at the 0.05% level

ENSO-IOD input sets were organized based on these months as potential predictors of spring rainfall for multiple regression analysis. F-test and t-test was conducted to evaluate the significance level of the models and the regression coefficients; among the constructed models the ones that did not violate the limits of statistical significance was selected, the models with lower error were chosen as the best model for each station. The regression coefficients, variance inflation factor (VIF), Durbin-Watson statistics (DW) and the Pearson correlation coefficient (r) of the best models are shown in Table 4 . It can be seen from this Table that VIFs for the selected models are near one, i.e. there is no multicolinearity among the predictors; also, the DW statistics is showing that the residuals of the models have no autocorrelation confirming the goodness-of-fit of the models. Nino3.4-DMI based models proved to be statistically significant and having better forecasting ability than SOI-DMI models for Victoria, with a Pearson r of 0.35 for Buchan, 0.36 for Malmsbury and 0.39 for Kaniva. Table 5 shows the MSE, MAE and Pearson correlation (r) of the best MR models for the three regions. It can be seen from Table 5 that the errors are relatively low for all the stations.

Table4. Summary of the best regression models

Station	Models	Co-efficient							r	VIF	DW
		Const.	Ni34(Jun)	Ni34(Jul)	Ni34(Aug)	DMI(Jun)	DMI(Jul)	DMI(Aug)			
Buchan	Ni34(Jul)-DMI(Jun)	0.51	---	-0.17	---	-0.23	---	---	0.35	1.10	2.1
Malmsbury	Ni34(Aug)-DMI(Jul)	0.55	---	---	-0.20	---	-0.22	---	0.36	1.12	1.9
Kaniva	Ni34(Jun)-DMI(Jul)	0.67	-0.32	---	---	---	-0.27	---	0.39	1.10	2.0

After calibrating and validating the models, in order to evaluate the generalization ability of the developed MR out-of-sample tests were carried out on the years 2007-2009 (Table 6). It can be seen that MR model is showing very poor generalization ability for Buchan (r= -0.90), however the ability of MR models to forecast out-of-sample sets improves for Malmsbury in central Victoria and Kaniva in west Victoria (r=0.48 and 0.67 respectively). While Pearson correlation shows how well the models are following the trend of the actual observations, Willmott index of agreement "*d*" shows how well the models are fitting the observations; the closer the value of "*d*" is to one the better is the model accuracy (Table 6). Figures 3 to 5 show the models developed for the three stations. In general regression models are showing an underestimation of the actual observations.

Table 5. Performance of the regression models

Station	r	MSE	MAE
Buchan	0.35	0.026	0.171
Malmsbury	0.36	0.030	0.140
Kaniva	0.39	0.041	0.163

Table 6. Performance of multiple regression models for the out-of-sample test set

Station	R	MSE	MAE	d
Buchan	-0.90	0.023	0.180	0.30
Malmsbury	0.48	0.013	0.100	0.50
Kaniva	0.67	0.051	0.163	0.44

4. CONCLUSION

This study focused on investigating the use of combined lagged El Nino Southern Oscillation (ENSO) and Indian Ocean dipole (IOD) as potential predictors of spring rainfall. Three regions (east, centre and west) of Victoria were chosen as case study. Nino3.4 and Southern Oscillation Index (SOI) were used as ENSO indicators and Dipole Mode Index (DMI) was chosen as IOD indicator.
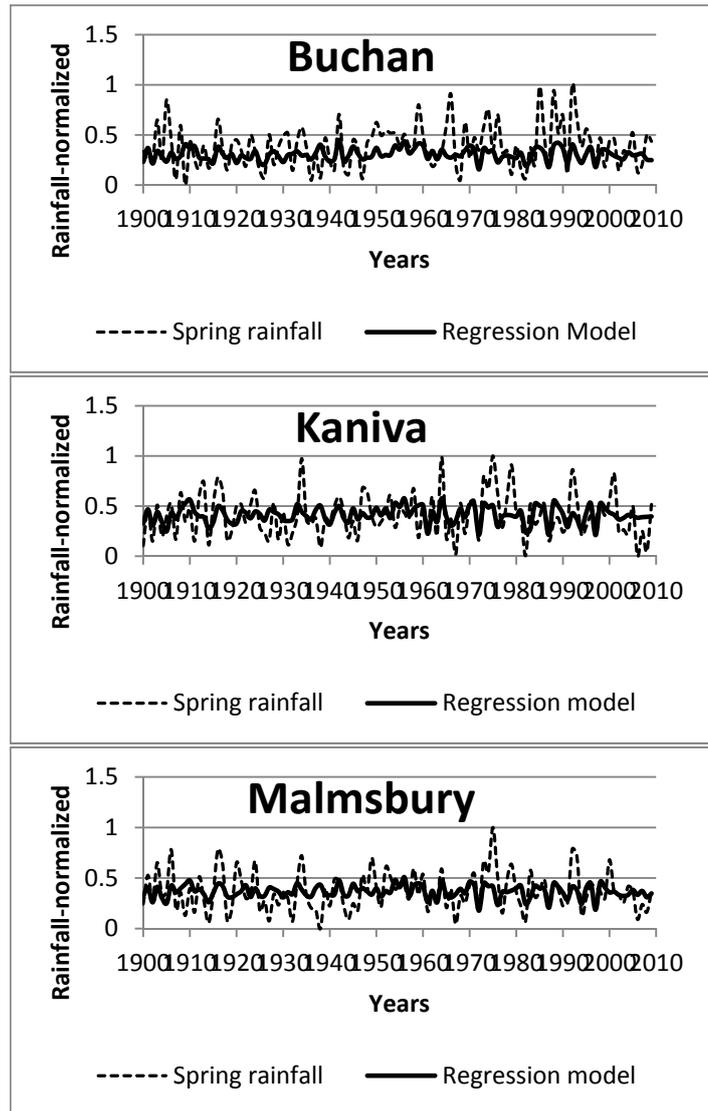
The effect of the phases of ENSO and IOD on spring rainfall was investigated by categorizing rainfall based on the phases of ENSO/IOD, Wilcoxon Signed Rank test was applied to test the significance of these effects. It was discovered that in the three stations considered El Nino and IOD positive affect rainfall more than La Nina and IOD negative.

The Pearson correlation coefficients of past values of the climate indices with spring rainfalls for the 3 stations were calculated; it was discovered that only the three months of June, July and August of Nino3.4, SOI and DMI have significant correlation with spring rainfall and these correlations are very weak. Nino3.4-DMI and SOI-DMI input sets were organized based on these months as potential predictors of spring rainfall for MR analysis. Among the several developed models the ones that did not violate the limits of statistical significance and multicollinearity and had lower model error were used for prediction purposes. The models were tested on out-of-sample sets. It was found that generalization ability of MR models for east Victoria is very poor compared to the other two regions. It was discovered that MR model is showing very poor generalization ability for Buchan ($r \sim -0.90$), however the ability of MR models to forecast out-of-sample sets improves for Malmsbury in central Victoria and Kaniva in west Victoria ($r=0.48$ and 0.67 respectively). By taking some care, multiple regression has the ability of forecasting rainfall based on lagged climate modes.

Figure 3. Multiple regression models for the three stations.

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.

Chiew, F., Piechota, T. C., Dracup, J., & McMahon, T. (1998). El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology, 204*(1), 138-149.

Drosdowsky, W., & Chmabers, L. E. (2001). Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *Journal of Climate, 14*(7), 1677-1687.

England, M. H., Ummenhofer, C. C. & Santoso, A. (2006). Interannual rainfall extremes over southwest Western Australia linked to Indian Ocean climate variability. *Journal of Climate, 19*(10), 1948-1969.

Evans, A. D., Bennett, J. M., & Ewenz, C. M. (2009). South Australian rainfall variability and climate extremes. *Climate Dynamics, 33*(4), 477-493.

Field, A. P. (2009). *Discovering statistics using SPSS*: SAGE publications Ltd.

Kirono, D. G. C., Chiew, F. H. S., & Kent, D. M. (2010). Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological Processes, 24*(10), 1237-1247.

Lin, F. J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality & Quantity, 42*(3), 417-426.

Mekanik, F. and Imteaz, M.A. (2013). Evaluating the effect of single and combined climate modes on rainfall predictability, in *"IAENG Transactions on Engineering Technologies"*, Chapter 40**,** Eds.: H.K. Kim, S. Ao, M.A. Amouzegar and B.B. Rieger, Springer Publisher, ISBN 978-94-007-6817-8.

Mekanik, F. & Imteaz, M. A. (2012). Forecasting Victorian spring rainfall using ENSO and IOD: A comparison of linear multiple regression and nonlinear ANN. International Conference on Uncertainty Reasoning and Knowledge Engineering, URKE 2012 , Jakarta, August.

Mekanik, F., Imteaz, M. A., Gato-Trinidad, S. & Elmahdi, A. (2013). Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology*, 503, 11-21.

Meneghini, B., Simmonds, I., & Smith, I. N. (2007). Association between Australian rainfall and the Southern Annular Mode. *International Journal of Climatology, 27*(1), 109-121.

Meyers, G., McIntosh, P., Pigot, L. & Pook, M. (2007). The years of El Niño, La Niña and interactions with the tropical Indian Ocean. *Journal of Climate, 20*(13), 2872-2880.

Murphy, B. F., & Timbal, B. (2008). A review of recent climate variability and climate change in southeastern Australia. *International Journal of Climatology, 28*(7), 859-879.

Nicholls, N. (2010). Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958–2007. *Climate Dynamics, 34*(6), 835-845.

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. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian ocean. *Nature, 401*(6751), 360-363.

Schepen, A., Wang, Q. J., & Robertson, D. (2012). Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *Journal of Climate, 25*(4), 1230-1246.

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.

Verdon, D. C., Wyatt, A. M., Kiem, A. S., & Franks, S. W. (2004). Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resources Research, 40*(10), W10201.

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society, 63*, 1309-1369.