

Forecasting of monthly rainfall in the Murray Darling Basin, Australia: Miles as a case study

J. Abbot & J. Marohasy
*School of Medical and Applied Sciences,
 Central Queensland University, Australia*

Abstract

The Murray Darling Basin accounts for nearly 40% of the value of agricultural production in Australia, and 65% of the irrigated land. We use an artificial neural network (ANN), a form of machine learning, to show the potential for more reliable monthly rainfall forecasts with a lead time of 3 months, and the potential skill of the same model for 6, 9, 12 and 18 month lead-times for the township of Miles, in the northern Basin. The skill of these forecasts is contrasted with the skill of the Predictive Ocean Atmosphere Model for Australia (POAMA), a general circulation model used for operational forecasts by the Australian Bureau of Meteorology. Forecasts from the ANN are significantly more skilful for all lead times. The ANN's capacity to integrate information from the climate indices Niño 1.2, 3, 3.4 and 4, the Dipole Mode Index (DMI), and also a composite local maximum and minimum atmospheric temperature series, contributes on average approximately 60% of the skill of the ANN forecast.

Keywords: rainfall, forecasting, artificial neural network, Murray Darling Basin.

1 Introduction

In Australia, water infrastructure and irrigated agriculture are most developed in the Murray Darling Basin, which covers 14% of Australia's land area [1]. The Water Act 2007 requires government to make decisions on water allocations between agriculture and the environment [2–4]. However, this planning, and on-farm productivity more generally, is constrained by the reliability and lead-time of rainfall forecasts [5].

Historically, statistical models based on climate indices including Pacific and Indian Ocean sea surface temperatures and Southern Oscillation Index (SOI) were



used by the Bureau of Meteorology to generate seasonal (3-month) forecasts from one to three months in advance [6]. The skill of these forecasts was always limited by the high variability in the strength of the associations of climate indices with rainfall, both temporally and spatially [7]. For example, Murphy and Ribbe [8] showed that the SOI and Pacific Ocean temperature indices Niño 3 and Niño 4 are influential, but quantitative relationships with seasonal rainfall are highly variable by time and geographic location, with linear correlations typically in the range -0.2 to 0.2. Similar results were reported by McBride and Nichols [9] and Cai *et al.* [10] for relationships between rainfall and SOI. Schepen *et al.* [11] also examined the relationships between 13 lagged monthly climate indices, including SOI, Niño 3, Niño 3.4, Niño 4 and DMI with seasonal rainfall over the Australian continent, reporting high variability in the strength of association, both temporally and spatially.

While traditional statistical models are limited in the number of input variables that can be effectively combined, advances in machine learning have significantly expanded this potential [12]. Machine learning is an interdisciplinary field that has close relationships with artificial intelligence, pattern recognition and data mining. While data mining focus on the discovery of previously unknown properties in data [13], machine learning focuses on prediction based on known properties learned from exposure to data sets during a process known as “training”. During training, a model is constructed from algorithms. A core objective of the learning process is to be able to generalize from experience [14]. Machine learning has become important in the medical diagnostic field [15–18] where information needs to be combined from different tests in which each test may carry some relevant, but limited, diagnostic information. There may be no consistently useful method of combining the information, with a traditional reliance on the skill and experience of the medical practitioner. There are obvious parallels here with rainfall forecasting, particularly in an Australian context, where individual climate indices convey useful predictive information about rainfall, but the associations are typically fragmented.

The mainstream approach to rainfall forecasting, is to attempt to simulate the actual physical processes through general circulation modelling. The Predictive Ocean Atmosphere Model for Australia (POAMA) is a general circulation model (GCM) developed by the Australian Bureau of Meteorology [19–21]. It has been run operationally since 2002, and used for all official seasonal forecasts since May 2013.

This study assesses the reliability and skill of an ANN versus POAMA to forecast rainfall for the location of Miles, in the northern Murray Darling Basin, building on earlier research using ANNs to forecast monthly rainfall for locations in south-eastern, central, western and northern Queensland [22–26].

2 Data and method

We used Neurosolutions Infinity software (NeuroDimensions Inc., Florida) to build a neural network model based on the attributes considered potentially the most important predictors of rainfall in the northern Murray Darling. The



reliability of this model was tested by running 7 trials for a three-month lead for the township of Miles on the western Darling Downs using the same initial inputs. This model was then used to predict monthly rainfall for Miles for 1, 3, 6, 9, 12 and 18 months in advance, for the 10 year period April 2004 to March 2014.

Miles was chosen as a case study because it has an exceptionally long rainfall record. The post office (station number 42023, Latitude -26.66S, Longitude 150.18E, elevation 302 m) began recording rainfall in 1885, and is still operating as a weather station today.

The other reason for choosing Miles, is that it is one of two sites in the Murray Darling Basin for which we were provided with output from POAMA, allowing direct comparison of output from the ANN with the general circulation model used for operational forecasts by the Australian Bureau of Meteorology. We compared output from the ANN with POAMA for 1, 3 and 9 months lead times for the period April 2004 to March 2011.

The attributes included monthly values of an atmospheric temperature composite and the following climate indices: SOI, DMI, Niño 4, Niño 3.4, Niño 3, Niño 1.2 and the Inter-decadal Pacific Oscillation (IPO). The forecasts were run with the full set of attributes, and also local rainfall, each lagged up to 12 months. The data was divided into training (70%), evaluation (15%) and test sets (15%). The test set was not used in training or evaluation.

Values for DMI and the four Niños were sourced from the Royal Netherlands Meteorological Institute Climate Explorer – a web application that is part of the World Meteorological Organisation and European Climate Assessment and Dataset project. Values for IPO were provided by Chris Folland from the UK Met Office. Values of SOI and also local minimum and maximum atmospheric temperatures used in the development of the temperature composite were obtained from the Australian Bureau of Meteorology.

We use Pearson correlation coefficient (r), the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to compare the skill of the rainfall forecast generated by the ANN, POAMA and also climatology (the long-term average).

3 Results

The neural network model was run seven times for a 3-month lead with very little variability of output. The mean Pearson correlation coefficient was 0.75 (SE Mean 0.009, StDev. 0.025), mean RMSE was 33.8 (SE Mean 0.622, StDev. 1.645) and MAE was 25.8 (SE Mean 0.390, StDev. 1.031). Even the least skilful of the seven forecasts, measured in terms of lowest Pearson correlation ($r = 0.69$), and highest RMSE (35.6) and MAE (26.9), successfully forecast the exceptionally wet summer of 2010–2011 in south east Queensland [27], Figure 1.

No single attribute was responsible for the skill of the forecasts, and there was considerable variability in the relative contribution of the different inputs attributes, Table 1.

For the period from 2002 through until 2011, seasonal rainfall predictions were made by the Australian BOM using POAMA 1.5. In July 2011, the Bureau

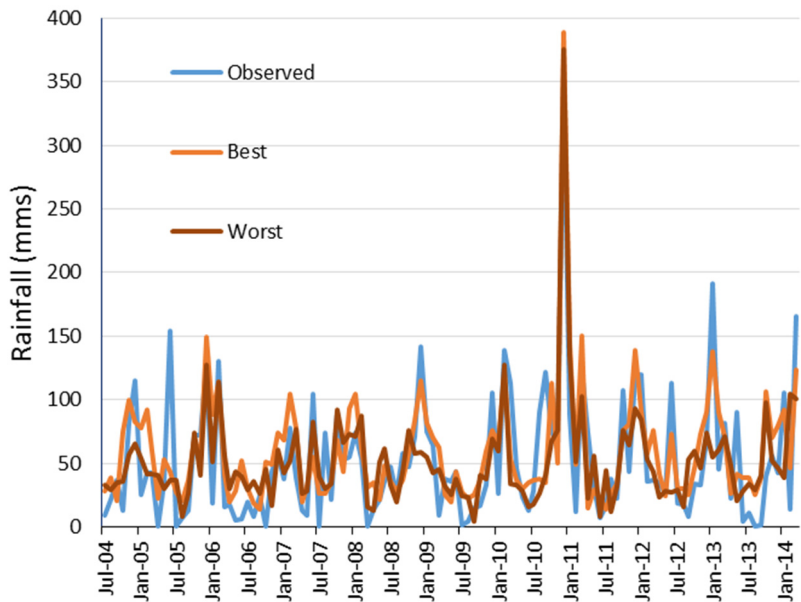


Figure 1: The best (orange) and worst (brown) forecast monthly rainfall for a 3-month lead, April 2004 to March 2014, and also observed rainfall (blue).

Table 1: The relative percentage contribution of the different input attributes (complex refers to a combination of attributes) to the three month forecast for the 7 different model runs.

Variable	Mean (%)	St. Dev.	Min.	Median	Max.
SOI	7.13	6.11	0	8.5	15.9
IPO	9.01	6.73	0	8.3	20.3
DMI	10.33	3.32	5.3	9.3	15.5
Niño	19.46	7.79	9.2	19.3	29.9
Temps.	28.16	5.17	22.1	27.5	37.3
Rainfall	7.10	3.42	3.5	8.5	11.7
Complex	18.86	11.77	1.5	22.7	31.4

provided us with monthly forecasts for the period to March 2011 as simple bilinear interpolations of surrounding grid points which were calculated from the ensemble mean which in turn had been calculated from many runs of this general circulation model, POAMA 1.5. Despite repeated requests, we have been unable to obtain forecasts for the subsequent period, nor forecasts from the newer version of POAMA, Version 2.4. Our analysis is thus limited to the period to March 2011, Figure 2.



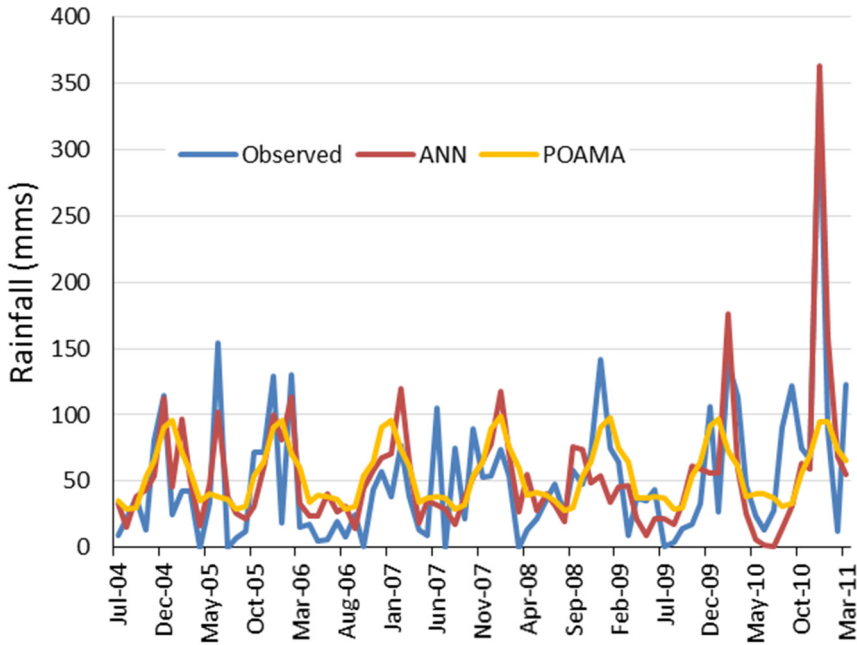


Figure 2: Forecasts from POAMA 1.5 (yellow line), and the ANN (red line) benchmarked against actual observed rainfall (blue line) for a 3-month lead for the period July 2004 to March 2011.

The inability of POAMA 1.5 to forecast the exceptionally wet summer of 2010–2011 contrasts dramatically with the ANN models' consistent forecast of this important feature, Figure 2.

Consistent with reported skill scores for POAMA in the literature [28–30], the forecast from POAMA downscaled from an ensemble mean for the locality of Miles was relatively poor at all lead times, Table 2.

4 Discussion

Seasonal rainfall predictions for Australia from POAMA 1.5 were run operationally from 2002 to 2011. The Bureau then changed to POAMA versions 2 and 2.4. In May 2013, the operational forecasts from POAMA 2.4 became the official forecasts for 250 x 250 kilometre grid areas (total area 62,500 km²) for the entire Australian continent, including the Murray Darling Basin. POAMA versions 2 and 2.4 are considered somewhat more reliable than version 1.5 [28], however, the improved accuracy is generally reported simply in terms of an improved capacity to forecast above or below median rainfall for each grid area. This is consistent with the format in which the official forecast is announced publicly, that is official forecasts are simply presented as the probability of being less, or more than, the long-term median for each grid location.



Table 2: Comparison of rainfall forecast skill for different forecasting methods at leads of 1, 3, 6 and 9 months for the period July 2004 to March 2011.

Model	R	MAE (mm)	RMSE (mm)
Lead 1 month			
ANN	0.82	22.8	29.4
Climatology	0.46	32.1	44.6
POAMA	0.43	33.1	45.5
Lead 3 months			
ANN	0.74	25.2	32.6
Climatology	0.44	32.8	45.2
POAMA	0.43	33.0	45.5
Lead 6 months			
ANN	0.74	26.2	33.8
Climatology	0.33	33.6	47.6
POAMA	0.33	34.7	48.0
Lead 9 months			
ANN	0.74	25.6	33.2
Climatology	0.34	33.9	47.3
POAMA	0.33	34.2	47.9

The few reported studies that have quantified the skill of the forecasts from POAMA 1.5, 2 and 2.4, beyond simply registering above or below medium rainfall, suggest that at a 1-month lead time, the correlations between monthly observed rainfall and output from POAMA can approach a correlation coefficient of about 0.4 [29]. This forecast skill quickly falls away with time, and by a lead-time of three months the forecast from POAMA is generally consistent with climatology, that is the long term average for the region [29]. Hawthorne *et al.* [30] reported that the skill of monthly rainfall forecasts derived from POAMA 2.4 at lead times up to 9 months is generally low.

This assessment of POAMA forecast skill is consistent with the results we have obtained, Table 2. At a lead of one month POAMA gives a slightly more skilful forecast than climatology, but at 3, 6 and 9 months the forecast from POAMA is essentially only equivalent to climatology, Table 2. It can be concluded that POAMA is not introducing significant additional rainfall forecast skill additional to climatology beyond 1 month lead time. Considered another way, POAMA generates anomalies which, when added to climatology, are effectively introducing low level noise. In contrast, the ANN gives a consistently higher correlation coefficient, r , and lower MAE and RMSE than climatology at 1, 3, 6 and 9 months, Table 2.

The combined four Niño values (1.2, 3, 3.4 and 4) provide on average about 20% of the forecast skill for the ANN models, Table 1. Another measure of ENSO, SOI, provides approximately 7% of the skill, Table 1. But the actual relative percentage contribution of the different input attributes was highly variable, Table 1. This is consistent with the extensive literature on the variable and often small, but nevertheless important, relative contribution of the different indices to rainfall in Australia, including in the Murray Darling Basin [11].

The relatively high skill scores for Miles for all lead times, Table 2, combined with the consistency of the forecast as demonstrated at the 3-month lead ($r = 0.75$, StDev 0.0246, $n = 7$) indicates that ANNs have significant application for rainfall forecasting in the Murray Darling Basin where there is a need for more reliable rainfall forecasts with longer lead-times [31]. That the seven different runs of the ANN at three months each forecast the wet summer of 2010–2011, also suggests that the ANNs can create a model that can effectively accommodate the extreme variability that characterises rainfall in the Murray Darling Basin, Figure 1.

Our findings are consistent with the work of Mekanik *et al.* [32], who also used an ANN to forecast rainfall in the Murray Darling Basin. In particular a comparison of multiple regression and ANNs for monthly forecasting in the state of Victoria, indicated that ANNs provided more skilful forecasts and that correlation coefficients of 0.58–0.97 could be achieved for western Victoria that is part of the south-eastern Murray Darling Basin [32].

5 Conclusion

The forecasting of rainfall in eastern Australia, including the Murray Darling Basin, has traditionally been approached through the application of simple statistical models or GCMs. We suggest a third way using machine learning, in particular ANNs.

A major limitation of the first approach, using simple statistical models that rely on empirical relationships between observed rainfall and one or two input parameters, is that there are no simple relationships extending temporally or geographically where any specific climate index dominates. At most, any individual climate index may explain only about 20% of the rainfall variability at a particular location [11]. It is also apparent that there is no simple method to ascertain what lag period for the particular indices selected should be included to forecast a specific lead time, for any one location [11]. Assessment of the many studies undertaken over recent decades [6–11, 33–35], indicates that the problem of pre-selecting optimal inputs is intrinsically difficult, and that forecast solutions are not amenable to the application of simple formulas incorporating sparse information.

The second approach, using GCMs, assumes that ocean and atmospheric processes are well enough understood that they can be represented mathematically in physical models that can provide useful forecast information. However, GCMs do not perform well at forecasting rainfall, and are very expensive to run. Furthermore, the GCMs produce output corresponding to very large grid areas,

and this requires downscaling before useful localised rainfall forecast information could potentially be generated.

The third approach, our approach, is based on the assumption that there are patterns embedded in historical data that are useful predictors of rainfall. No assumptions are made about the actual physical processes that generate rainfall at a particular location. Much of the skill is in the construction of the data sets and choice of the ANN configuration.

Using Miles as a case study, we have shown that it is possible to forecast monthly rainfall in the Murray Darling Basin with a high level of skill for lead times between 1 and 18 months using an ANNs. The skill of the forecasts remains consistent across repeated runs with the same initial input data and does not deteriorate significantly as lead time is increased. Furthermore, the skill is significantly higher than can be achieved using POAMA. The output from the ANN provides localised rainfall forecasts without the need for complex downscaling procedures. Another advantage of the ANN is that it quantified the amount of rainfall expected, rather than just providing a categorisation of above or below median.

Acknowledgements

This work was funded by the B. Macfie Family Foundation and CQ University, Australia.

References

- [1] Australian Bureau of Statistics, Water and the Murray-Darling Basin – a statistical profile, 2000–01 to 2005–06, Catalogue. No. 4610.0.55.007, Canberra, <http://www.abs.gov.au/ausstats/abs@.nsf/mf/4610.0.55.007>
- [2] Qureshi, M.E, Grafton, R.Q, Kirby, M. & Hanjra, M.A. Understanding irrigation water use efficiency at different scales for better policy reform: a case study of the Murray–Darling Basin, Australia, *Water Policy* 13, pp. 1–17, 2011.
- [3] Qureshi, M.E, Whitten, S.M., Mainuddin, M., Marvanek, S. & Elmahdi, A., A biophysical and economic model of agriculture and water in the Murray-Darling Basin, Australia, *Env. Modelling and Software* 41, pp. 98–106, 2013.
- [4] Wang, J., Wang E. & Liu, D.L. Modelling the impacts of climate change on wheat yield and field water balance over the Murray-Darling Basin in Australia, *Theoretical and Applied Climatology*, 104, pp. 285–300, 2011.
- [5] Barros A.P. & Bowden, G.J., Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin, *Journal of Hydrology*. 357, pp. 349–367, 2008.
- [6] Fawcett R.J.B. & Stone, R.C., A comparison of two seasonal rainfall forecasting systems for Australia, *Australian Meteorological and Oceanographic Journal*, 60, pp. 15–24, 2010.



- [7] Drosowsky W. & Chambers L.E., Near-Global Sea Surface Temperature Anomalies as Predictors of Australian Seasonal Rainfall, *Journal of Climate* 14, pp. 1677–1687, 2001.
- [8] Murphy B.F. & Ribbe, J., Variability of southeastern Queensland rainfall and climate indices, *International Journal of Climatology*, 24, pp. 703–721 2004.
- [9] McBride, J.L. & Nichols N., Seasonal Relationships between Australian Rainfall and the Southern Oscillation, *Monthly Weather Review*, 111, pp. 1998–2004, 1983.
- [10] Cai, W., Whetton P.H. & Pittock A.B., Fluctuations of the relationship between ENSO and northeast Australian rainfall, *Climate Dynamics* 17, pp. 421–432, 2001.
- [11] Schepen, A., Wang Q.J. & Robertson D., Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall *Journal of Climate* 25, pp. 1230–1246, 2012.
- [12] Mitchell, T., *Machine Learning*, McGraw Hill: Maidenhead, UK, p. 2, 1997.
- [13] Simon P., *Too Big to Ignore: The Business Case for Big Data*. John Wiley & Sons: Somerset, New Jersey, p. 89, 2013.
- [14] Bishop C. M., *Pattern Recognition and Machine Learning*. Springer: New York, NY, 2006.
- [15] Trambaiolli, L.R., Lorena, A.C., Fraga, F.J. *et al.*, Improving Alzheimer's Disease Diagnosis with Machine Learning Techniques, *Clinical EEG And Neuroscience*, 42(3), pp. 160–165, 2011.
- [16] Martis, R.J., Krishnan, M., Muthu, R., Chakraborty, C. *et al.*, Automated Screening of Arrhythmia Using Wavelet Based Machine Learning Techniques, *Journal of Medical Systems*, 36(2), pp. 677–688, 2012.
- [17] Hariharan, M., Polat, K. & Sindhu, R., A new hybrid intelligent system for accurate detection of Parkinson's disease, *Computer Methods and Programs in Biomedicine*, 113(3), pp. 904–913, 2014.
- [18] Murukesan, L., Murugappan, M., Iqbal, M., *et al.*, Machine Learning Approach for Sudden Cardiac Arrest Prediction Based on Optimal Heart Rate Variability Features, *Journal of Medical Imaging and Health Informatics* 4(4), pp. 521–532, 2014.
- [19] Marshall A.G., Hudson D., Hendon H.H. *et al.*, Simulation and prediction of blocking in the Australian region and its influence on intra-seasonal rainfall in POAMA-2. *Climate Dynamics* 42, pp. 3271–3288, 2014.
- [20] Hudson, D., Alves, O., Hendon H.H. *et al.*, Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia, *Quarterly Journal of the Royal Meteorological Society* 137, pp. 673–689, 2011.
- [21] Schepen, A., Wang, Q.J., & Robertson, D.E. Seasonal Forecasts of Australian Rainfall through Calibration and Bridging of Coupled GCM Outputs *Monthly Weather Review*. 142(5), pp. 1758–1770, 2014.
- [22] Abbot, J. & Marohasy, J., Application of Artificial Neural Networks to Rainfall Forecasting in Queensland, Australia, *Advances in Atmospheric Science* 29(4), pp. 717–730, 2012.

- [23] Abbot J. & Marohasy J., The application of artificial intelligence for monthly rainfall forecasting in the Brisbane Catchment, Queensland, Australia. *WIT Transactions on Ecology and the Environment*, 172, pp. 125–135, 2013.
- [24] Abbot, J., & Marohasy, J., Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks, *Atmospheric Research*, 138, pp. 166–178, 2014.
- [25] Abbot, J. & Marohasy, J., Using artificial intelligence to forecast monthly rainfall under present and future climates for the Bowen Basin, Queensland, Australia, *International Journal of Sustainable Development and Planning*. In press. 2015.
- [26] Abbot, J. & Marohasy, J., Using lagged and forecast climate indices with artificial intelligence to predict monthly rainfall in the Brisbane Catchment, Queensland, Australia, *International Journal of Sustainable Development and Planning*. In press. 2015.
- [27] van den Honert, R.C. & McAneney J., The 2011 Brisbane Floods: Causes, Impacts and Implications, *Water* 3, pp. 1149–1173, 2011.
- [28] Langford, S. and Hendon H.H., Improving Reliability of Coupled Model Forecasts of Australian Seasonal Rainfall, *Monthly Weather Review*, 141, pp. 728–741, 2013.
- [29] Shao, Q. and Li, M., An improved statistical analogue downscaling procedure for seasonal precipitation forecast, *Stochastic Environmental Research and Risk Assessment* 27, pp. 819–830, 2013.
- [30] Hawthorne S., Wang, Q. J., Schepen, A & Robertson, D., Effective use of general circulation model outputs for forecasting monthly rainfalls to long lead times, *Water Resources Research*, 49, pp. 5427–5436, 2013.
- [31] Marohasy J., Murray Darling dams ‘out of balance’ as water flows to South Australia, October 23, 2013 *Australian Financial Review* http://www.afr.com/p/opinion/murray_darling_dams_out_south_balance_QIv1Tk3HbnyacTdJ1a4NMM
- [32] Mekanik F., Imteaz, M.A., Gato-Trinidad, S. & Elmahdi, A., Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes, *Journal of Hydrology*, 503, pp. 11–21, 2013.
- [33] Tularam, G.A., Relationship between El Niño Southern Oscillation Index and rainfall (Queensland, Australia), *International Journal of Sustainable Development and Planning*, 5(4), pp. 378–391, 2010.
- [34] Cobon, D.H. & Toombs, N.R., Forecasting rainfall based on the Southern Oscillation Index phases at longer lead-times in Australia, *The Rangeland Journal*, 35, pp. 373–383, 2013.
- [35] Hasan, M.M. & Dunn, P.K., Understanding the effect of climatology on monthly rainfall amounts in Australia using Tweedie GLMs, *International Journal of Climatology*, 32, pp. 1006–1017, 2012.
- [36] He, X., Guan, H., Zhang X. & Simmons, C.T., A wavelet-based multiple linear regression model for forecasting monthly rainfall, *International Journal of Climatology*, 34, pp. 1898–1912, 2014.



- [37] Chowdhury, R.K. & Beecham, S., Australian rainfall trends and their relation to the Southern Oscillation Index, *Hydrological Processes*, 24, pp. 504–514, 2010.
- [38] Meneghini, B., Simmonds, I., & Smith, I.N., Association between Australian rainfall and the Southern Annular Mode, *International Journal of Climatology*, 27, pp. 109–121, 2007.

