Improving monthly rainfall forecasts using artificial neural networks and single-month optimisation: a case study of the Brisbane catchment, Queensland, Australia

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

Abstract

Official medium-term rainfall forecasts failed to warn of the impending heavy rainfall in the Brisbane catchment during the summer of 2010–2011 with resulting catastrophic flooding causing loss of life, extensive property damage and disruption of economic activity in south-eastern Queensland, Australia. Since the flooding, the Australia Bureau of Meteorology has changed its method of forecasting from an empirical statistical scheme to the use of a general circulation model, the Predictive Ocean and Atmospheric Model for Australia (POAMA). More skilful forecasts, however, can be achieved through the use of an altogether different technique involving artificial neural networks (ANN), a form of machine learning. Building on previous studies comparing the skill of the ANN forecasts for Gatton and Harrisville in the Brisbane catchment with output from POAMA, this study shows how monthly ANN rainfall forecasts can be further significantly improved through a technique referred to as "single-month optimisation". This technique enables the temporal variability in the influence of key climate indices to be better incorporated into the rainfall forecast. In particular, correlation coefficients exceeding 0.85 at a 12 month lead time can be achieved for both Harrisville and Gatton.
Keywords: rainfall, forecast, artificial neural network, flood, Queensland.

1 Introduction

Prolonged and extensive rainfall over large areas of Queensland, coupled with already saturated catchments, led to flooding of historic proportions in Queensland in December 2010, extending into January 2011 [1]. Thirty-three people died in the 2010/2011 floods, and more than 78 per cent of the state (an area bigger than France and Germany combined) was declared a disaster zone, with over 2.5 million people affected [1]. Some 29,000 homes and businesses suffered some form of inundation. The Queensland Reconstruction Authority estimated that the cost of flooding events to be in excess of AUS$5 billion. The scale of the disaster led to the establishment, in January 2011, of the Commission of Inquiry into the Queensland floods of 2010/2011 [1].

The Brisbane River is the longest river (309 km) in south-east Queensland, with its source located in the Brisbane Range, 120 km north-west of the city of Brisbane. From there it makes its way south before joining the Stanley River, downstream of Somerset Dam, to run into Lake Wivenhoe [2]. This lake provides the main water supply for Brisbane, the state capital of Queensland, and was created by construction of the Wivenhoe Dam [2]. The dam was completed in 1984, in response to severe flooding in 1974, with the principal aim of safeguarding the city of Brisbane from future floods. Nevertheless, in January 2011 Brisbane experienced its second highest flood in over 100 years. Major flooding occurred throughout most of the catchment and an estimated 18,000 properties were inundated [2].

The 2010/11 Brisbane flood has been designated as a "dam release flood" by hydrologists appointed by the Insurance Council of Australia [2]. This suggests that the sudden release of water from the Wivenhoe Dam was a principal cause of flooding along the mainstream and tributaries of the Brisbane River [2]. These events remain contentious, and are the subject of a major class action lawsuit against the Queensland government, scheduled to be heard in court during 2016.

Exactly where and how much rain falls within a catchment will critically determine a dam's effectiveness for flood mitigation [2], as well as how much water should be held before the onset of each summer wet season. The official rainfall forecasts issued by the Australian Bureau of Meteorology, BOM, do not provide adequately detailed information at the localised level, or with sufficient lead times. The BOM forecasts are only provided to the public in the form of probabilities relative to the median, and do not differentiate between anticipated rainfall slightly above the median and an extreme rainfall event as occurred in January 2011. Until May 2013 these forecasts were based on a statistical scheme using an El Niño Southern Oscillation (ENSO) index as a primary predictor in a relatively simple statistical model [3, 4]. The BOM switched to the use of the Predictive Atmospheric Model for Australia (POAMA) which is a general circulation model, in June 2013 [5–7]. General circulation models, however, do not generally perform well at forecasting rainfall, despite substantial efforts to enhance performance over many years [8–10]. Various statistical models continue to be developed for rainfall prediction in Australia, generally using climate indices as inputs [11–14].

Artificial neural networks, ANNs, a form of machine learning, provide a third potential technique for medium-term rainfall forecasting. While ANNs can be classified as a type of statistical model, they offer several important advantages over the more simple statistical models that were used by the BOM until May 2013. ANNs can accommodate non-linear relationships, and test multiple inputs, particularly important when the influence of climate indices may vary geographically and temporarily in poorly understood ways [13]. ANNs have been applied to produce seasonal and monthly rainfall forecasts in many parts of the world [15–17], including Australia [18–21].

Our first attempt at using ANN models to forecast rainfall at two key locations in the Brisbane catchment was reported at the Seventh River Basin Management Conference [21]. In the present study, the ANN approach has been extended to investigate the application of a novel method of optimising the ANN models, in particular instead of optimising a given set of inputs for all months simultaneously, the program is provided with data for only a single month, i.e. all Januaries, then all Februaries, etcetera. In effect, this results in the optimisation of predicted rainfall for each month individually and thus the creation of a unique model for each month for all years. These models are combined to arrive at a final forecast and corresponding skill score for the test periods of July 2004 to August 2011, and July 2004 to June 2014. This single-month optimisation technique is more time consuming, requiring the program to be run 12 times in order to have a result for every month of the year. The final monthly rainfall forecast is, however, significantly more skilfully measured in terms of root mean square error, mean absolute error and Pearson correlations coefficients.

2   Data and methods

The skill of a rainfall forecast using any statistical model, including ANNs, will depend on the quality and relevance of the data provided as input to the model, with longer time series generally giving a superior forecast. Gatton and Harrisville are both within the Brisbane catchment, and in close proximity to a grid area where comparative monthly forecasts have been published using the general circulation model POAMA [9]. Both stations have relatively long rainfall records. Gatton (station number 40082, Latitude 27.54° S, Longitude 152.34° E, elevation 89 m) opened in 1897, while Harrisville post office (station number 40094, Latitude 27.81° S, Longitude 152.67° E, elevation 61 m) opened a year earlier in 1896. Both locations are still operating as weather stations today.

Variations in rainfall in many parts of the world, including the Brisbane catchment, are associated with large-scale climate phenomena which can be described by climate indices typically measuring changes in temperatures and pressures across oceans [22–24]. ENSO, a Pacific Ocean phenomena can be measured by both the Southern Oscillation Index (SOI) and a combination of four different Niño values (Niño 4, Niño 3.4, Niño 3, Niño 1.2). The Inter-decadal Pacific Oscillation (IPO) also measures pressure and temperature changes in the Pacific Ocean. The Indian Ocean Dipole measured by the Dipole Mode Index (DMI), is a measure of pressure and temperature changes in the Indian Ocean.

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 minimum and maximum temperatures used in the development of the local temperature composite were obtained from the Australian Bureau of Meteorology.

All of the above climate indices, local rainfall and a composite of temperature data from Brisbane Regional Office (station 40210) and Brisbane (Aero station 40842), were provided as inputs to *Neurosolutions Infinity* software (NeuroDimensions Inc., Florida) and used to build probabilistic ANN models. Each attribute was lagged up to 12 months. The data was divided into training (75%), evaluation (15%) and test sets (10%).

Two approaches were used for ANN optimization. With the first approach, designated as "all-month optimization", data for all 12 months of the year was included as input and optimised together, as in our previous studies [18–21]. With the second approach, designated as "single month optimisation", forecasts corresponding to each calendar month were performed individually, so that 12 optimisations were carried out to produce monthly rainfall forecasts for the entire year.

The two different approaches were compared for lead-times of 12 months for the test periods July 2004 to June 2014, and also July 2004 to August 2011. Comparisons with POAMA could only be made to August 2011, as the BOM has been unwilling to provide the output from POAMA used in operational forecasts as deterministic forecasts for point locations beyond this period (National Climate Centre, BOM, pers. comm. July 2014). It should be noted that while POAMA has only been used for operational forecasts since June 2013, it was being used to produce forecasts since 1997. The data used in this study are from POAMA Version 1.5, which was used by the BOM until August 2011.

Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and Pearson correlation coefficient (r) were used to compare the skill of rainfall forecasts generated by the two different optimisation techniques relative to observed rainfall, and to compare these forecasts with output from POAMA and also climatology (the long-term average).

3   Results and discussion

In previous studies using ANN models to forecast monthly rainfall for Gatton and Harrisville an all-month method of optimisation was used to forecast for one, two or three months in advance [19, 21]. To be specific, all months for all years were provided as input to the program, and the skill of the forecast calculated as the difference between the forecast and observed rainfall for each month averaged over the test period. In the case of Gatton and Harrisville for 1 to 3 month leads, an average RMSE of 54 mms and 46 mms respectively was achieved for the 13 year period of July 1997 to December 2010 [19]. This was better than the forecast

for POAMA for the same period, where a mean RMSE of 70 mms for both Gatton and Harrisville was reported [19].

In this study we choose a longer lead forecast time of one year (12 months) and make the comparison with the 8 month-lead for POAMA, the longest lead forecast available (Table 1). The test period for the comparison with POAMA is 7 years from July 2004 to August 2011.

Table 1:   Skill parameters for monthly rainfall forecasting for Harrisville and Gatton on an annual basis for the test period July 2004 to August 2011.

Harrisville			
	RMSE (mm)	MAE (mm)	r
Climatology	49.8	34.6	0.57
POAMA	49.3	34.2	0.59
All-month optimisation	39.4	29.8	0.78
Single-month optimisation	28.4	19.1	0.88
Gatton			
	RMSE (mm)	MAE (mm)	r
Climatology	59.7	40.2	0.50
POAMA	59.2	40.0	0.51
All-month optimisation	49.9	34.7	0.70
Single-month optimisation	32.4	22.2	0.91

The skill of the POAMA forecasts are almost the same as climatology (the long-term average) (Table 1). This is in part because of the way the POAMA values are calculated, which is that the BOM converts the output from POAMA into anomalies that are then added to climatology.

In contrast, the all-month and single-month optimisation are generated independently and give lower RMSE and MAE, and higher r-values for both Gatton and Harrisville (Table 1). The single-month optimisation gives the better result for all skill measures and for both sites (Table 1). The nature of the improvement in forecast skills can be illustrated through charting (Figures 1 and 2).

As expected, climatology reflects the repeating seasonal periodic signal of rainfall, with higher rainfall during the summer months at Gatton (Figure 1) and Harrisville (Figure 2). These climatology signals are very similar to the forecast generated using POAMA for both Gatton and Harrisville, suggesting that POAMA adds very little to the overall skill of the forecast (Figures 1 and 2).

The all-month ANN forecasts produces a much better alignment between forecast and observed rainfall signals for Gatton (Figure 1) but in particular for Harrisville (Figure 2). Peaks and troughs corresponding to periods of heavier and reduced rainfall in different years are becoming differentiated (Figures 1 and 2). This becomes even more pronounced examining the ANN forecasts using single-month optimisations (Figures 1 and 2). Of particular significance is the emergence of the peak corresponding to December 2010 for Harrisville, clearly indicating a forecast of higher monthly rainfall in comparison to the remainder of the 7-year test period.
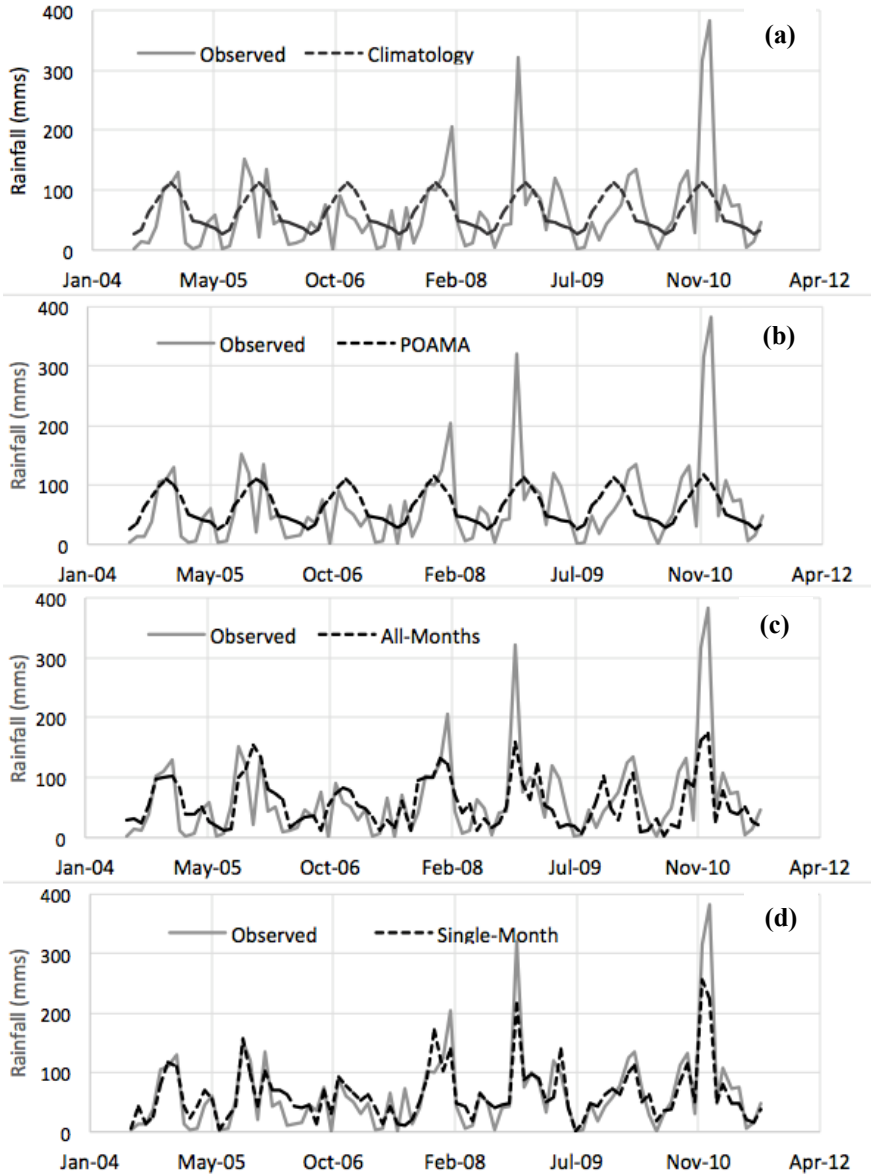
Figure 1:   Forecast and observed monthly rainfall for Gatton with 12 months lead time: (a) climatology; (b) POAMA; (c) all-month optimisation; (d) single-month optimisation.
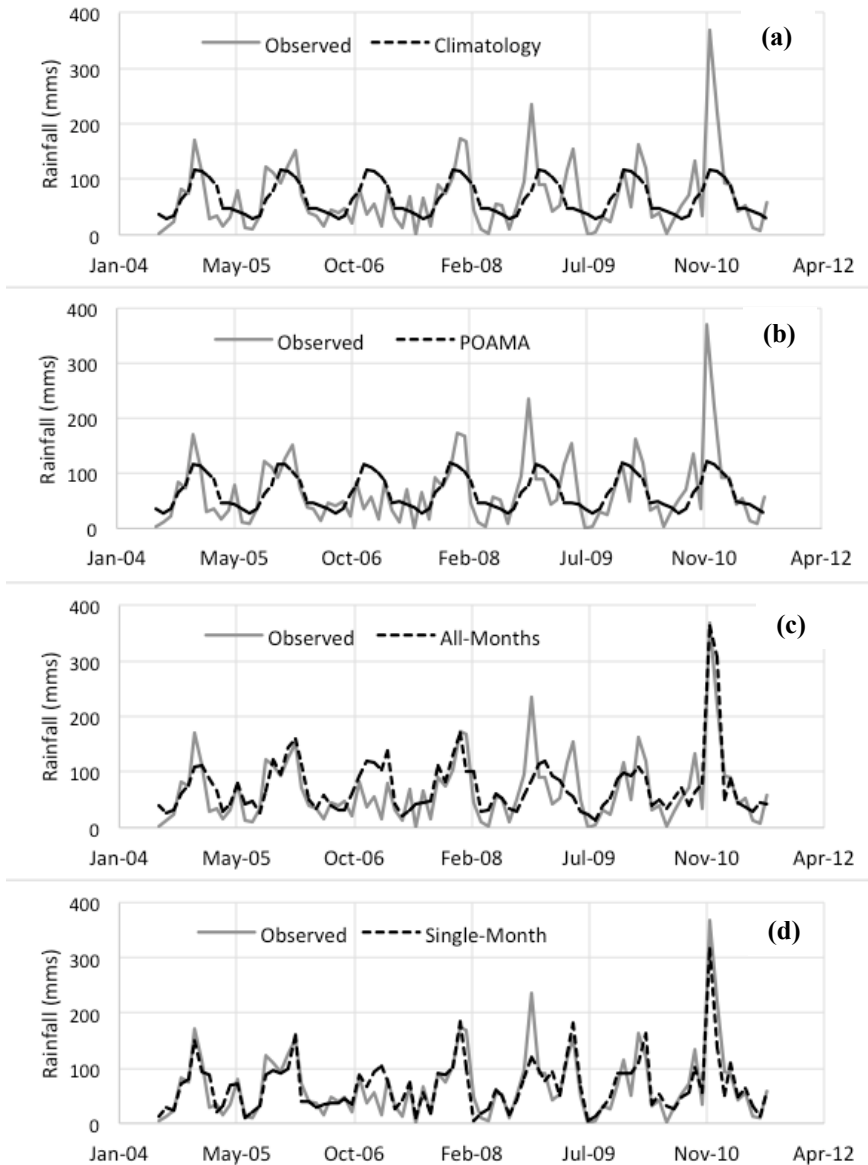
Figure 2:   Forecast and observed monthly rainfall for Harrisville with 12 months lead time: (a) climatology; (b) POAMA; (c) all-month optimisation; (d) single-month optimisation.

Deconstructing the forecasts for each month for the two different methods (Tables 2 and 3), it becomes evident how a more skilful forecast is achieved using the single-month method of optimisation. Comparing r-values for Gatton, it is evident that while skills scores in excess of 0.90 are achieved for all months for the single-month optimisation, they are highly variable for the all-month optimisation (Table 2). Indeed, the ANN has no skill at forecasting the lower rainfall months of April through to September when using the all-month optimisation technique (Table 2).

Table 2:   Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and r values for monthly rainfall forecasts for Gatton at 12 month lead time using all-month and single month optimisation for the period July 2004 to August 2011.

Month	All-month optimisation			Single-month optimisation		
	RMSE (mm)	MAE (mm)	r	RMSE (mm)	MAE (mm)	r
January	83.6	57.6	0.78	63.9	40.1	0.95
February	50.3	44.6	0.70	29.0	23.4	0.97
March	35.0	29.6	0.66	20.4	17.6	0.97
April	47.7	43.6	-0.32	35.1	32.7	0.96
May	32.9	23.2	0.07	22.3	17.5	0.94
June	39.8	33.8	-0.50	15.8	13.6	0.88
July	18.8	17.5	0.22	8.3	4.4	0.93
August	28.6	22.5	-0.10	10.2	9.2	0.95
September	39.0	26.3	0.08	15.4	12.5	0.98
October	36.7	32.5	0.73	15.4	12.5	0.98
November	60.1	40.1	0.74	37.0	24.5	0.99
December	71.9	48.3	0.52	34.0	27.0	0.94

When undertaking an all-month optimisation the ANN must search for a best-fit across all climate indices for all months simultaneously. The results from this study confirm that the ANN is favouring the months of higher rainfall when undertaking all-month optimisations (Tables 2 and 3). In the case of Harrisville achieving high r-values for January and December has come at the expense of better forecasts for the lower rainfall months (Table 3).

Contrast this result with the single-month optimisations, where skill scores in excess of r = 0.80 where achieved for most months (Tables 2 and 3). Clearly the ANN is able to achieve a skilful forecast for the lower rainfall months of April to September when it is able to optimise these months individually, that is without having to consider the other calendar months (Tables 2 and 3).

Table 3:   RMSE, MAE and r for monthly rainfall forecast for Harrisville at 12 month lead time using (i) all-month and (ii) single month optimisation.

Month	All-month optimisation			Single-month optimisation		
	RMSE (mm)	MAE (mm)	r	RMSE (mm)	MAE (mm)	r
January	48.4	42.0	0.77	41.0	36.2	0.85
February	56.4	50.8	0.58	58.4	50.5	0.56
March	42.5	35.7	0.44	28.0	23.1	0.90
April	22.7	17.0	0.67	7.9	6.2	0.98
May	39.0	26.6	0.28	21.1	18.9	0.96
June	33.2	28.1	0.19	14.5	11.9	0.96
July	26.2	22.9	0.19	7.0	6.2	0.95
August	24.2	20.5	0.50	8.9	7.6	0.98
September	19.7	16.9	0.14	6.2	4.7	0.98
October	48.9	45.1	0.26	29.1	26.7	0.96
November	68.1	50.6	-0.20	39.2	22.4	0.83
December	39.1	22.7	0.91	23.4	17.3	0.98

4   Conclusion

In previous studies using ANNs to forecast monthly rainfall in Queensland [18–21], data for all 12 calendar months were optimised together for any given set of inputs. With the hindsight of this study, the all-month optimisation procedure would be expected to favour the influence of months with relatively high rainfall, compared to those with lower rainfall. This is evident in the results comparing all-month with single-month optimisations (Tables 2 and 3).

Nevertheless the all-month optimisation approach consistently produced results with better skill than either climatology or POAMA (Table 1).

A complementary approach is demonstrated in this study where the ANN is optimised for each calendar month for all years individually, with a further significant improvement in skill (Figures 1 and 2, Tables 1–3). Of particular significance are the prominent peaks at December 2010 and January 2011 for the ANN forecasts using the single-month optimisation technique at Harrisville (Figure 2). This shows that the occurrence of very heavy rainfall can be forecast at long-lead times of at least 12 months.

The present study also shows that the monthly rainfall forecast skill using the all-month optimisation technique may be quite variable when individual months are considered (Tables 2 and 3). This can likely be attributed to variability in the influence of specified lagged climate indices at different times during the year [22–24]. Further exploration of the results, in particular an analysis of the relative contribution of the different climate indices for each month, and comparing all-month and single-month optimisation, may provide important insight into which climate indices are most influential at different times of the year for Harrisville

and Gatton. This is will be the focus of a follow-on study, with an expanded number of runs for the single-month optimisations for both Gatton and Harrisville.

Acknowledgements

References

[1]   Queensland Flood Commission of Enquiry, Final Report 2012. http://www.floodcommission.qld.gov.au/__data/assets/pdf_file/0007/1169 8/QFCI-Final-Report-March-2012.pdf

[2]   van den Honert, R.C. & McAneney, J. The 2011 Brisbane Floods: Causes, Impacts and Implications. *Water*, 3, pp. 1149-1173, 2011.

[3]   Drosdowsky, W. & Chambers L.E., Near-Global Sea Surface Temperature Anomalies as Predictors of Australian Seasonal Rainfall. *Journal of Climate,* 14, pp. 1677-1687, 2001.

[4]   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.

[5]   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, pp. 1758-1777, 2014.

[6]   Marshall, A.G., Hudson, D., Hendon, H.H., Pook M.J., Alves, O. & Wheeler, M.C., 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.

[7]   Langford, S. & Hendon, H.H., Improving Reliability of Coupled Model Forecasts of Australian Seasonal Rainfall. *Monthly Weather Review,* 141, pp. 728-741, 2013.

[8]   Shao, Q. & Li M., An improved statistical analogue downscaling procedure for seasonal precipitation forecast. *Stochastic Environmental Research Risk Assessment,* 27, pp. 819-830, 2013.

[9]   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.

[10] Vaze, J., Teng J. & Chiew, F.H.S., Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia. *Hydrological Processes,* 25, pp. 1486-1497, 2011.

[11] Clarke, A.J., Van Gorder, S. & Everingham, Y., Forecasting Long-Lead Rainfall Probability with Application to Australia's Northeastern Coast. *Journal of Applied Meteorology and Climatology,* 49, pp. 1443-1453, 2010.

[12] 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. 373-338, 2013.

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

[14] 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(7)**, pp. 1006-1017, 2012.

[15] Acharya, N., Chattopadhyay, S., Kulkarni M.A. & Mohanty, U.C., A neurocomputing approach to predict monsoon rainfall in monthly scale using SST anomaly as a predictor. *Acta Geophysica,* **60(1)**, pp. 260-279, 2012.

[16] Dahamsheha, A. & Aksoy H. Artificial neural network models for forecasting intermittent monthly precipitation in arid regions. *Applied Meteorology* 16, pp. 325-337, 2009.

[17] Shukla, R.P, Tripathi, K.C, Pandey, A.C., & Das, I.M.L, Prediction of Indian summer monsoon rainfall using Niño indices: A neural network approach. *Atmospheric Research,* 102, pp. 99-109, 2011.

[18] Abbot, J. & Marohasy, J., Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences*, 29(4), 717-730, 2012.

[19] 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.

[20] Abbot, J. & Marohasy, J., The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia. WIT Transactions on Ecology and the Environment, 171, pp. 287-297, 2013.

[21] 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.

[22] 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.

[23] Murphy, B.F. & Ribbe, J. Variability of Southeastern Queensland Rainfall and   Climate Indices. *International Journal of Climatology,* 24, pp. 703-721, 2004.

[24] Risbey, J.S., Pook M.J., Mcintosh P.C., Wheeler M.C. & Hendon H.H., On the Remote Drivers of Rainfall Variability in Australia. *Monthly Weather Review,* 137, pp. 3233-3253, 2009.