

# The application of artificial intelligence for monthly rainfall forecasting in the Brisbane Catchment, Queensland, Australia

J. Abbot & J. Marohasy

*Centre for Plant and Water Sciences, Central Queensland University,  
Noosa, Queensland, Australia*

## Abstract

Concurrent relationships between climate indices and Australian spring rainfall have been used extensively to explain weather events. In order for climate indices to be useful for rainfall forecasting there must be relationships between their lagged values and rainfall. The methods currently used by the Australian Bureau of Meteorology for seasonal weather forecasting have limited capacity to exploit the often non-linear relationships that potentially exist between the lagged values for these indices and rainfall. This paper reports on the application of a method of forecasting based on the use of neural networks, a form of artificial intelligence. Neural networks facilitate the input of multiple variables simultaneously. The variables most useful for determining rainfall are elucidated by application of algorithms during the optimisation process.

Brisbane, the capital of Queensland, Australia, has flooded periodically and catastrophically. The neural network described in this study was used to forecast rainfall for three locations in the Brisbane River catchment one to three months in advance, including the 2011 flood event. Results are compared on the basis of root mean square error with output from the Australian Bureau of Meteorology's general circulation model, POAMA. The Neural Network model shows considerable more skill. The Neural Network incorporates lagged values for key climatic indices and also rainfall and atmospheric temperatures.

*Keywords: rainfall, forecast, lagged climate indices, SOI, IPO, Niña 3.4, Brisbane, flood, neural network.*



## 1 Introduction

Brisbane, the capital of Queensland, Australia, flooded in 1841, 1893 and 1974. The Wivenhoe dam was built for flood mitigation and completed in 1985 but this dam did not prevent catastrophic flooding again in 2011. In fact, the dam was in part responsible for the extensive flooding in 2011 because when its structural integrity was challenged, large volumes of water were abruptly released [1]. The Brisbane City Council reported a clean-up bill of A\$440 million (335.3 million Euro equivalent) [2] and insurance claims totalled approximately A\$2.4 billion (1.83 billion Euro equivalent) [3]. In addition, there is a pending class action lawsuit, potentially the largest ever in Australia, with claims for damages estimated to exceed A\$1 billion (0.76 billion Euro) [4].

The Queensland Floods Commission of Inquiry [5, 6] was most concerned with forecasts from the Australian Bureau of Meteorology (BoM) in the week preceding the flooding. The focus of the present study, however, is on longer-term rainfall forecasting from 1 month to 3 months in advance, potentially valuable information in the context of managing water infrastructure, such as Wivenhoe dam.

Rainfall variability in the Brisbane catchment is high [7], with alternating wet and dry decades [8]. The Interim Report Queensland Floods Commission of Inquiry [6] referred briefly to the seasonal forecast issued by the BoM in October 2010, predicting a 75% chance of above median rainfall in south-east Queensland for the period November 2010 to January 2011, and an active cyclone season. Those briefings included statements that: there was a well-established and quite strong La Niña pattern, that was more than ‘run-of-the-mill’, which was expected to persist until at least March; there was a historical correlation between La Niña events and tropical cyclones in the Coral Sea; and above normal rainfall would continue over much of Queensland.

La Niña events can be measured through the climate indices Niña 3.4 and Southern Oscillation Index (SOI). In this study, we also use the Inter-decadal Pacific Oscillation (IPO), which is known to modulate phases of La Niña and rainfall [9–11]. Numerous studies attempt to relate individual indices to concurrent values of rainfall [12–18]. This can be valuable in terms of understanding causation of weather phenomena. For forecasting purposes, however, it is only lagged values that can legitimately be used as input [19–24], unless an independent model for forecasting a particular climate index is incorporated into a rainfall forecast.

The BoM relies on both statistical and dynamic systems to generate rainfall forecasts. Dynamic systems, such as general circulation models, use computer simulations that attempt to model physical processes in the climate system. This method can give seasonal rainfall predictions from 1 month up to a year in advance. The dynamic forecasting approach is more expensive to implement and operate than statistical climate prediction systems. Despite substantial research efforts and technological advances, sophisticated dynamic systems are still unable to consistently outperform simple statistical prediction systems [24].



Statistical methods require the existence of sufficiently strong empirical relationships between rainfall and other input variables particularly climate indices. This requires the availability of long data records and stationary relationships over time. Simple statistical models rely exclusively on the existence of linear relationships between rainfall and the input variables such as lagged climate indices. In practice, linear relationships for rainfall are highly discontinuous and fragmented when considered on a spatial and a temporal basis [24]. Using only simple linear models, it is essentially impossible to create a unified statistical model that can be applied over an extended geographical area such as south-east Queensland, and also incorporate statistically significant relationships between input and output for more than a limited period beyond a few months of the year. The result is essentially a set of rules corresponding to a highly complex classification system for rainfall prediction, with limited practical utility. The inevitable conclusion is that continuous simple linear relationships are not fundamental properties of the climate system, and it is preferable to utilize statistical modelling techniques that are designed to accommodate non-linear relationships.

Neural networks (NNs) have been used to forecast rainfall in many parts of the world [25]. This is a statistical approach that enables non-linear relationships to be considered as well as facilitating the input of multiple variables. For example, Shulka *et al.* [26] found with a NN model, inputting Nino indices produced superior forecasts compared to linear models for forecasting Indian monsoon rainfall. However, the NN approach has rarely been used to forecast rainfall in Australia. NNs have been used to forecasting Victorian spring rainfall using lagged SOI, Nino 4 and Indian Ocean Dipole (IOD) as input parameters, achieving superior performance to simple linear correlation models [27]. Abbot and Marohasy [28] used a neural network model with lagged climate indices to generate monthly rainfall forecasts for Queensland, that demonstrated more skill than POAMA.

## 2 Method

We forecast rainfall for three sites in the Brisbane catchment that are close to the Wivenhoe reservoir, and for which there is historical rainfall data extending back at least one hundred years. Monthly rainfall data for the sites was obtained from the BoM. Harrisville (station 040094) has an annual average rainfall of 820 mm and is approximately 30 km south-west of Wivenhoe. Gatton (station 040083) has an annual average rainfall of 778 mm and is located approximately 50 km south of Wivenhoe. Lowood (station 040120) has an annual average rainfall of 812 mm and is approximately 10 km south of Wivenhoe.

There are no correspondingly long temperature data series for these sites, or any other locality in south-east Queensland. Composite maximum and minimum temperature records were constructed from January 1887 to present using data from the BoM for three sites in Brisbane: the Brisbane Regional Office (station 40214), Brisbane Aero (station 40223) and Brisbane (station 40913).



Values for the climate indices SOI and Niña 3.4 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.

We used the NN software NeuroSolutions 6 for Excel (NeuroDimensions, Florida, USA) with an Elman neural network [28] to generate a model through which up to six possible input data sets were investigated. These sets were constructed to correspond to monthly values of rainfall (Rain), maximum and minimum atmospheric temperatures (MaxT, MinT), Southern Oscillation Index (SOI), Interdecadal Pacific Oscillation (IPO), and Niña 3.4 (Nino).

A unary data set was defined as the current monthly value of one of these input parameters, plus the corresponding lagged values for the previous twelve months, comprising a total of 13 input columns to the neural network. A binary data set was defined as a combination of two unary data sets. For example the combination of 26 input data columns for SOI and maximum temperature (SOI/MaxT) comprises a binary set. Similarly, ternary combinations of the unary sets consist of 39 input data columns, as for example the combination of SOI, maximum temperature and rainfall (SOI/MaxT/Rain).

In our NN model, outputs were assigned as the monthly rainfall with a lead-time of 1 month, 2 months or 3 months. For each input data set, the NN was optimised for 3000 epochs using a genetic optimization algorithm for 10 generations. Training sets comprised approximately 85% of the total data available for each location, with the remaining approximately 15% used for testing. To compare the skill of this model among sites with differing annual rainfall totals, root mean square error (RMSE) values were normalized by dividing the RMSE by the corresponding monthly average rainfall to give a weighted non-dimensional index (WNDI) [28].

It was found that lower overall RMSE values can, in some instances, be generated by combining outputs from different input sets. This was done by calculating a weighed linear averaging of output rainfall values corresponding to each time period for a specific site. The resultant output time-series can produce a near equivalent, or lower RMSE values, compared to the minimum value without post-processing, and also achieve better representation of extreme rainfall events when charted. This improvement can possibly be explained by suggesting that the observed rainfall time-series signals can be viewed as the resultant of simpler sets of overlapping component signals, each having specific dominant drivers. Signal decomposition and analysis based on this premise will be the subject of future studies. Another post-processing procedure that we tested was the application of an expansion function to the output to decompress values on the rainfall axis. The raw output from the neural network tends to reflect optimal solutions that are concentrated around the mean and may underemphasize very low and very high values. It is possible to expand the output with respect to the rainfall axis without an increase in overall RMSE for the model, and in this way the high and low rainfall values are better represented.

### 3 Results and discussion

Linear correlations between many individual climate indices and rainfall have been examined for different areas across the Australian continent. Typically, concurrent correlations fall in the range of 0 to 0.5 and show high variability both spatially and temporally [14]. There is also high variability in quantitative associations between rainfall and lagged climate indices [24]. Despite the poor correlations, single input variables, typically SOI, have been used extensively for seasonal rainfall forecasting in Australia using simple statistical models.

In contrast, the NN approach enables input of a large number of variables simultaneously, with no assumptions as to which may be most significant in the prediction of rainfall for a specific lead period. If particular columns of inputs within a set, or an entire unary set, are not useful for prediction, there is a high probability these will be preferentially culled by the genetic algorithm, thus progressively refining the model by retaining only more significant inputs.

Feeding unary, binary and ternary combinations of input data sets into the NN, RMSE values were generated for Harrisville, Gatton and Lowood using all possible combinations of unary sets for 1 month lead time. There was no simple, unique, input combination of sets that always stood-out as providing a best forecast. This was also true with regard to the optimal duration of the lag period for inputs.

This result is consistent with the recent study by Schepen *et al.* [24] where 13 oceanic and atmospheric climate indices lagged at periods of 1, 2 and 3 months were evaluated as predictors of Australian seasonal rainfall. The overall picture that emerged in Schepen *et al.* [24], when considered spatially and temporally, is a complex mosaic, or patchwork, for assignment of dominant lagged indicators of rainfall.

RMSE values for the present investigation for the site of Harrisville, considering a 1 month lead forecast, ranged from a high of 59.9 mm for the ternary set (Rain/IPO/Nino) to a low of 44.9 mm for the ternary input set (SOI/Nino/MinT). Values for Gatton ranged from a high of 60.7 mm for the ternary set (Rain/MinT/MaxT) to a low of 48.9 mm for the unary input set (Rain). Output for Lowood ranged from a high of 71.3 mm for the ternary set (SOI/Nino/MinT) to a low of 56.7 mm for the input set (Rain/SOI). Climate indices, atmospheric temperatures and rainfall were all important components of the optimal data input sets.

#### 3.1 Comparing NN output with BoM statistical model forecast

The statistical forecast issued in November 2010 by the BoM for the 3-month period December 2010, January 2011 and February 2011, was in the form of a seasonal forecast map, fig. 1, indicating a 75% probability of above median rainfall for the Brisbane River catchment. This forecast was consistent with total precipitation observed over the three-month period. The flooding, however, resulted from falls of high intensity over a period of much shorter duration and was not forecast by the BoM.



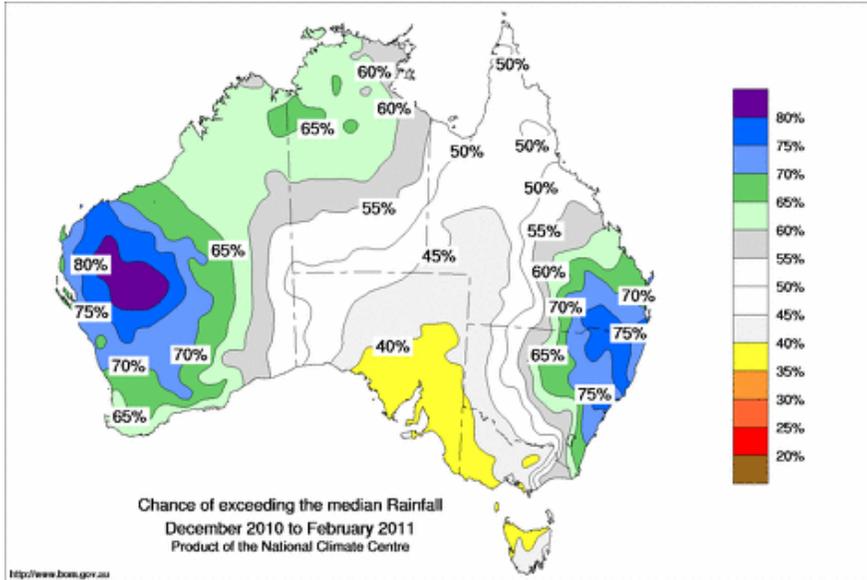


Figure 1: Seasonal rainfall forecast by the BoM issued in November 2010.

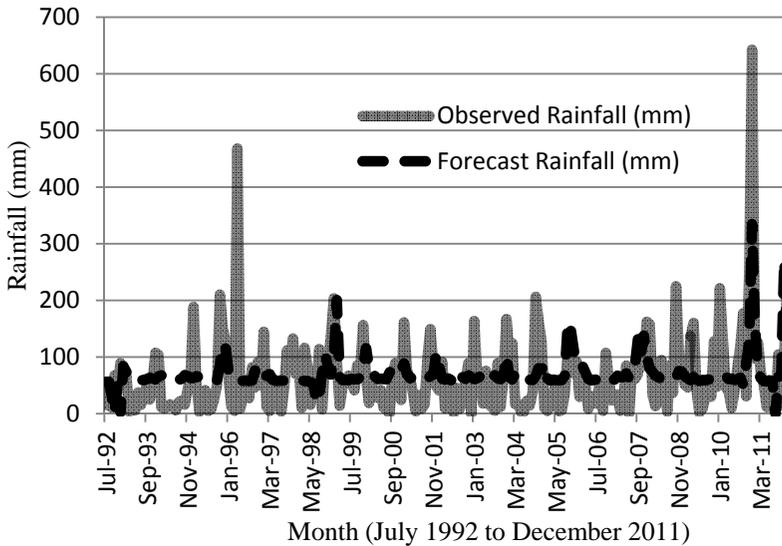


Figure 2: Neural network monthly rainfall forecast for Lowood (May 1992 to December 2011).

The general 3-month BoM forecast, fig. 1, is contrasted with the more specific time-series forecast from our NN, fig. 2. The NN forecast is for monthly rainfall for Lowood with a one-month lead-time for the 235-month period

(approximately 20 years) starting in May 1992 and ending in December 2011. The observed peak in December 2010 of 642.6 mm is represented in the forecast by a prominent forecast peak of 454.4 mm. This signal is clearly distinguishable from the prolonged period of drought in the prior 10 years when the maximum rainfall in any month never exceeded 220 mm, fig. 2.

This forecast, fig. 2, was produced by combining two primary NN output time-series forecasts, both generated with ternary input data sets. The first model used the ternary input set (Rain/IPO/Nino) to give an RMSE of 61.2 mm. The second model used the ternary input set (Rain/MaxT/IPO) to give output with an RMSE of 61.3 mm. The former model provided a better differentiation of the high rainfall period during the summer of 2010-2011. The latter model gave a better representation of rainfall during dryer periods. The combined output had an RMSE value of 57.2 mm.

### 3.2 Comparing NN output with BoM general circulation model

Output from the NN is directly comparable with output from the BoM's Predictive Oceanic Atmospheric Model for Australia, POAMA. The comparison is limited to the 13.5-year period from July 1997 to December 2010 for Harrisville and Gatton as this was the only output available from the BoM for POAMA. Comparisons of RSME and WNDI values show that the NN gave a more skilled forecast than POAMA for both sites and for each of the lead times, Table 1. Interestingly, the forecasts for both POAMA and the NN do not show clear trends in skill improvement moving from a three- to one-month lead time, Table 1.

Table 1: Comparing the skill of the rainfall forecasts for Harrisville and Gatton.

Lead	Neural Network			POAMA	
	Inputs sets	RMSE	WNDI	RMSE	WNDI
Harrisville					
1 month	MinT/SOI/Nino	44.9	0.66	74.9	1.10
2 months	MinT/IPO	45.7	0.67	69.1	1.01
3 months	Nino/MaxT	46.4	0.68	64.7	0.95
Gatton					
1 month	Rain	48.9	0.76	74.9	1.16
2 months	IPO/MinT	57.6	0.89	70.2	1.08
3 months	IPO/Rain	56.3	0.87	64.4	0.99

When the output for the NN is compared with the output for POAMA by way of charting, fig. 3 versus fig. 4, it is evident that the NN forecast generally shows less deviance from the observed rainfall values, as would be expected from the lower RMSE values, Table 1. Both methods show the summer peaks in rainfall for Harrisville, but there is no clear signal in the POAMA forecast, that differentiates the forecast rainfall for December 2010 from monthly rainfall



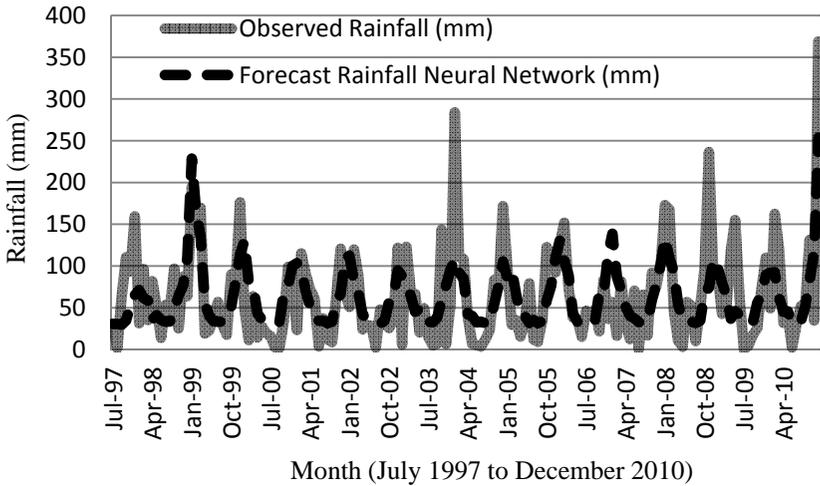


Figure 3: Observed rainfall versus Neural Network forecast for Harrisville.

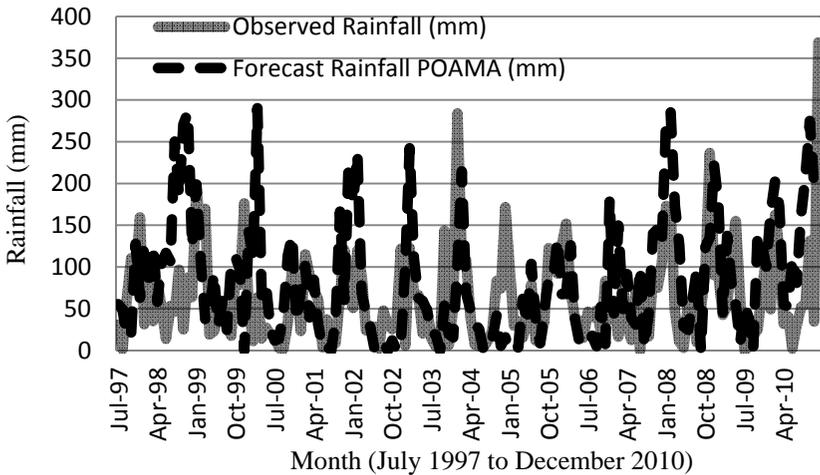


Figure 4: Observed rainfall versus POAMA forecast for Harrisville.

during the previous decade, fig 4. The POAMA forecast for the 2010-2011 summer peaks too early reaching a maximum in November 2010, before heavy rains were actually observed, fig 4.



The output from POAMA for Harrisville is generally consistent with what is known of rainfall forecasts from general circulation models (GCMs) for the Australian east coast. For example, Vaze *et al.* [29] concluded that all 15 GCMs investigated in their study failed to reproduce the observed historical annual and seasonal mean rainfalls, across south-east Australia.

## 4 Conclusions

The catastrophic flooding in the Brisbane River catchment during the summer of 2010-2011 caused loss of life and property damage that could have been avoided if there had been more timely and accurate rainfall forecasts [1]. More recent episodes of intense rainfall and severe flooding in southern Queensland occurred during January 2013 [30]. The coastal city of Bundaberg 350km north of Brisbane was particularly affected. This was during a period where the BoM seasonal forecast for the region was for below median rainfall [31].

Both the statistical and dynamic forecasting models developed over many years by the BoM appear to be limited in their capacity to provide adequate warning of heavy rainfall. Alternative modelling techniques, such as NNs, provide more sophisticated methods for relating available input variables, such as lagged climate indices, to the desired output, rainfall, with specified lead times.

NN have not been extensively investigated for forecasting of Australian rainfall. However, studies from other regions of the world demonstrate their potential utility for forecasting. The present study has shown that it is possible to generate improved forecasts for the Brisbane River catchment by application of this approach using lagged climate indices, atmospheric temperatures and rainfall as inputs. There is also significant potential to further improve these forecast through further refinements of input parameters, for example by inclusion of rates of change of indices such as SOI, that appear to be of value in other statistical forecast models [32].

## Acknowledgement

This work was funded by the B. Macfie Family Foundation.

## References

- [1] van den Honert R.C. and McAnaney J., The 2011 Brisbane Floods: Causes, Impacts and Implications, *Water*, **3**, pp. 1149-1173, 2011.
- [2] Cratchley, D., Insurers prepared for mounting flood costs. *Herald Sun* Jan 13 2013 <http://www.heraldsun.com.au/news/breaking-news/insurers-blame-councils-for-flood-damage/story-e6frf7kf-1226564743427>.
- [3] Griffiths M., Brisbane Floods to cost city \$440 million. ABC News, February 8, 2011 <http://www.abc.net.au/pm/content/2011/s3133417.htm>.



- [4] Johnson, S., Queensland Floods Class Action could set record. Yahoo 7 News January 20, 2013. <http://au.news.yahoo.com/latest/a/latest/15887827/qld-floods-class-action-could-set-record/>.
- [5] Queensland Floods Commission of Inquiry, Final Report. 2012. [www.floodcommission.qld.gov.au](http://www.floodcommission.qld.gov.au).
- [6] Queensland Floods Commission of Inquiry, Interim Report., p. 45, 2012 [www.floodcommission.qld.gov.au](http://www.floodcommission.qld.gov.au).
- [7] Lough J.M., Rainfall variations in Queensland Australia: 1891–1986. *Int. J. Climatol.* **11**, pp. 745–768, 1991.
- [8] Murphy B.F. and Ribbe, J., Variability of Southeastern Queensland Rainfall and Climate Indices. *Int. J. Climatol.* **24**, pp. 703–721, 2004.
- [9] Salinger, M.J., Renwick, J.A. and Mullan, A.B., Interdecadal Pacific Oscillation and South Pacific Climate. *Int. J. Climatol.* **21**, 1705–1721, 2001.
- [10] Cai W. and van Rensch P., The 2011 southeast Queensland extreme summer rainfall: A confirmation of a negative Pacific Decadal Oscillation phase? *Geophys. Res. Lett.*, **39**, pp. L08702, 2012.
- [11] Verdon, D.C., Wyatt, A.M., Kiem, A.S. and Franks S.W., Multidecadal variability of rainfall and streamflow: Eastern Australia, *Water Resources Res.*, **40**, W10201, 2004.
- [12] Speer, M.S., Leslie, L.M. and Fierro, A.O., Australian east coast rainfall decline related to large scale climate drivers. *Clim. Dyn.* **36**, pp. 1419–1429, 2011.
- [13] Meneghini, B. Simmondsa, I. and Smith, I.N. Association between Australian rainfall and the Southern Annular Mode, *Int. J. Climatol.* **27**, pp. 109–121, 2007.
- [14] Risby, J.S., Pook, M.J. and Mcintosh, P.C., On the Remote Drivers of Rainfall Variability in Australia. *Monthly Weather Rev.* **137**, pp. 3233–3253, 2009.
- [15] Stone R.C. and Aulicemms, A., SOI phase relationships with rainfall in eastern Australia. *Int. J. Climatol.* **12**, pp. 625–636, 1992.
- [16] Vizard, A.L. and Anderson, G.A., The resolution and potential value of Australian seasonal rainfall forecasts based on the five phases of the Southern Oscillation Index. *Crop & Pasture Sci.*, **60**, pp. 230–239 2009.
- [17] Wang G. and Hendon, H., Sensitivity of Australian Rainfall to Inter–El Niño Variations, *J. Climate* **20**, pp. 4211–4226, 2007.
- [18] McBride J.L. and Nicholls N. Seasonal relationships between rainfall and the southern oscillation. *Monthly Weather Rev.* **111**, pp. 1998–2004, 1983.
- [19] Dutta, S.C., Ritchie, J.W. Freebairn D.M. and Abawi, G.Y., Rainfall streamflow response to El Niño Southern Oscillation: a case study in a semiarid catchment, Australia, *Hydrol. Sci. J.*, **51(6)**, pp. 1006–1020, 2006.
- [20] Kirono D.G.C., Chiew, F.H.S. and Kent, D.M., Identification of best predictors for forecasting seasonal rainfall and runoff in Australia, *Hydrol. Process.* **24**, pp. 1237–1247, 2010.



- [21] Hasan M.M. and Dunn P.K., Understanding the effect of climatology on monthly rainfall amounts in Australia using Tweedie GLMs, *Int. J. Climatol.* **32**, pp. 1006–1017, 2012.
- [22] Chiew, F.H.S., Piechota, T.C., Dracup, J.A. and McMahon, T.A., El Niño/southern oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. *J. Hydrol.* **204**, pp. 138-149, 1998.
- [23] Tularam, G.A., Relationship Between El Niño Southern Oscillation Index and Rainfall (Queensland, Australia), *Int. J. Sus. Dev. Plann.* **5(4)**, pp. 378–391, 2010.
- [24] Schepen, A., Wang, Q. J. and Robertson, D., Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall. *J. Climate* 1230-1246, 2012.
- [25] Dahamsheha, A. and Aksoyb, H., Artificial neural network models for forecasting intermittent monthly precipitation in arid regions. *Meteorol. Appl.* **16**, pp. 325–337, 2009.
- [26] Shukla R.P., Tripathi K.C., Pandey, A.C. and Das, I.M.L., Prediction Indian summer monsoon rainfall using Niño indices: A neural network approach. *Atmos. Res.* **102**, pp. 99–109, 2011.
- [27] Mekanik F. and Imteaz, M.A., 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. 2012
- [28] Abbot, J. and Marohasy J., Application of Artificial Neural Networks to Rainfall Forecasting in Queensland, Australia. *Adv. Atmos. Sci.* **29(4)**, pp. 717-730, 2012.
- [29] Vaze, J., Teng, J. and Chiew, F.H.S., Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia. *Hydrol. Process.* **25**, pp. 1486–1497, 2011.
- [30] Flood disaster unfolds as weather wreaks havoc, ABC News. January 29, 2013. <http://www.abc.net.au/news/2013-01-28/qld-flooding-alert-moves-south/4486666>.
- [31] Australian Bureau of Meteorology: Seasonal rainfall Forecast – December 2012 for the period January 2013 to March 2012: A drier summer favoured for parts of eastern Australia. <http://www.bom.gov.au/climate/ahead/archive/rainfall/20121219.shtml>.
- [32] Stone, R.C. and Auliciems A., SOI phase relationships with rainfall in eastern Australia. *Int. J. Climatol.* **12**, pp. 625–636, 1992.

