

The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia

J. Abbot & J. Marohasy

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

Abstract

The Bowen Basin contains the largest coal reserves in Australia. Prolonged heavy rainfall during the 2010-2011 wet-season severely affected industry operations with an estimated economic loss of A\$5.7 billion (£3.8 billion). There was no explicit warning of the exceptionally wet conditions in the seasonal forecast from the Australian Bureau of Meteorology, which simply suggested a 50-55% probability of above median rainfall for the Bowen Basin.

In this study, the value of using neural networks, a form of artificial intelligence, to forecast monthly rainfall for the town of Nebo in the Bowen Basin is explored. Neural networks facilitate the input of multiple climate indices and the exploration of their non-linear relationships. Through genetic optimisations of input variables related to temperatures, including atmospheric temperatures and sea surface temperatures expressed through the Inter-decadal Pacific Oscillation and Niño 3.4, it is possible to develop monthly rainfall forecasts for Nebo superior to the best seasonal forecasts from the Bureau of Meteorology. As neural networks employ far superior technology for exploring the patterns and relationships within historical data including climate indices they are to be preferred.

Keywords: rainfall, neural network, forecast, Southern Oscillation index Interdecadal Pacific Oscillation, coal, mining.

1 Introduction

The very wet summer of 2010/2011 severely affected mining operations in Queensland, Australia. It is estimated that 85% of Queensland coalmines had to



either restrict production, or close entirely. By May 2011, Queensland’s coal mining sector had recovered to only 75% of its pre-flood output. These events lead to a loss of A\$5.7 billion (£3.8 billion) equivalent to 2.2% of Queensland’s gross state product for the financial year ending June 2011. A report prepared for Australia’s National Climate Change Adaptation Research Facility examined the impacts of the extreme weather event on Queensland’s mining industry and concluded that currently available climate forecasts are not useful enough to the industry, lacking localised information, and other micro details, to enable focused pro-active planning and risk management [1]. The report suggested that the mining industry would benefit from obtaining and using more precise seasonal and shorter-term rainfall forecasts.

The Australian Bureau of Meteorology, BoM, currently uses both statistical and dynamic forecasting models to generate seasonal rainfall predictions up to a year in advance. Statistical forecasting methods rely on the availability of climate indices, for example the El Niño Southern Oscillation (ENSO), and historical data. Dynamic forecasting systems, such as general circulation models, use computer simulations that attempt to directly model physical processes relevant to climate and climate change. The dynamic models are more expensive to operate, and despite substantial research efforts and technological advances, are still unable to consistently outperform the more simple statistical prediction systems [2].

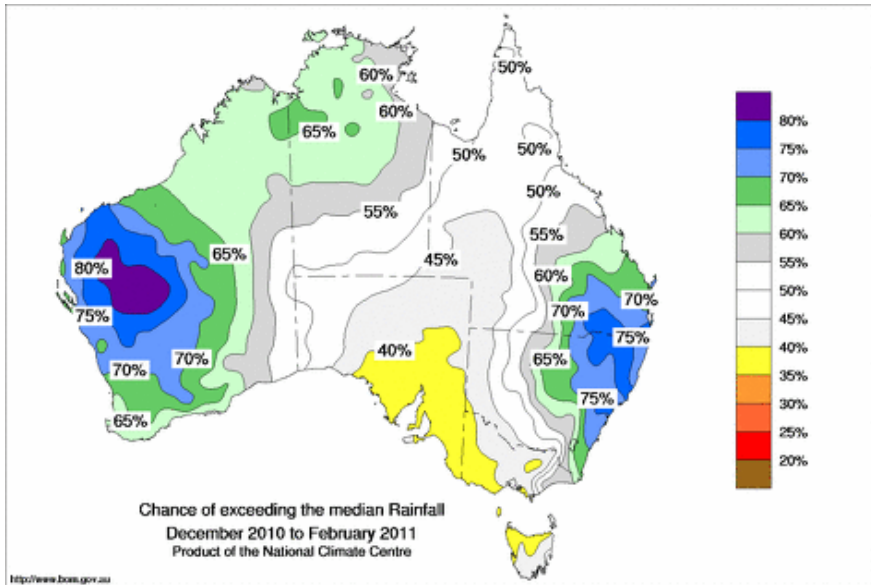


Figure 1: Seasonal rainfall forecast issued in November 2010 by the Australian Bureau of Meteorology.

The seasonal rainfall forecast for Australia issued by the BoM in November 2010 indicated a probability of 50-55% of above median rainfall for the Bowen

Basin for December 2011 to February 2011, fig. 1. This rainfall forecast was accurate when compared with observed rainfall for the entire three-month period. However, it did not include any indication of the magnitude and distribution of the rainfall within this three-month period.

In addition to the forecast from the BoM, the Queensland government issues two seasonal weather forecasts of comparable reliability also based on statistical methods using climate indices [3]. The Queensland government forecasts issued in November 2010 were for “a higher than normal probability of exceeding median rainfall across much of the state” for the period November to January and also for the period November to March [4]. There was no forecast specific to the flood periods and no mention of anticipated catastrophic flooding.

The statistical models produced by the BoM and Queensland government use a limited number of input variables, and often assume that there are linear relationships between those variables and rainfall. More sophisticated statistical modelling techniques, such as neural networks based on artificial intelligence, enable simultaneous inputs of many variables, and can also accommodate non-linear relationships common in historical rainfall data and climate indices.

Neural networks are not routinely used by government agencies in Australia to generate seasonal rainfall forecasts but have been used in other parts of the world, particularly in regions subject to highly variable monsoonal rainfall, reviewed in Abbot and Marohasy [6]. They are massive parallel-distributed, information-processing systems with characteristics resembling the biological neural networks of the human brain. In essence, they mine data for historical patterns that can then be applied to predict future events. The application of this approach requires sufficient relevant historical data, the presence of patterns that can be detected, and their continuance into the future period where forecasts are desired. From this perspective, the neural network approach does not differ from other statistical models, but is much more adaptable in recognising and utilizing complex patterns. Preliminary investigations comparing the skill of rainfall forecasts from the BoM’s general circulation model with output from a simple neural network for 17 sites in Queensland suggest they have can provide a more skilled seasonal forecast [6].

This study examines the application of neural networks for forecasting monthly rainfall one month in advance (one-month lead-time) for the site of Nebo, a town located close to a major open pit coal mine operated by Rio Tinto at Hail Creek in the Bowen Basin of Queensland. In particular, it investigates the potential to generate forecast signals corresponding to periods of very heavy rainfall, with a lead-time that would potentially alert mine operators of impending flood risks. The site of Nebo was chosen because of its proximity to a coal mine in the Bowen Basin and because it is a site with over 120 years of historical rainfall data.

Input variables used in this study included the ENSO climate indices used by the BoM and Queensland government, Niño 3.4 and SOI respectively, and also the Inter-decadal Pacific Oscillation (IPO) that is thought to modulate the influence of ENSO on rainfall along the east coast of Australia [7–9].



2 Methods

Neural network software, NeuroSolutions 6 for Excel (NeuroDimensions, Florida, USA), was used with an Elman neural network. The desired output, which is the observed rainfall, was assigned as the monthly rainfall with a lead-time of one month ahead of the current month. For each input data set, the neural network was optimised for 3000 epochs using a genetic optimisation algorithm for 10 or 20 generations. Training sets comprised approximately 85% of the total data with the remaining 15% used for testing. The test period, which is also the forecast period, was for 137 months from August 2000 to December 2011 that included the exceptionally wet summer of 2010-2011.

The monthly rainfall data for Nebo (station 033054) were obtained from the BoM. Atmospheric temperature data, both maximums and minimums, were obtained from the Te Kowai Experimental Station in Mackay (station 033047) available from the year 1908. Values for SOI were also sourced from the BoM. Values for the climate index 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.

Six unary input data sets were constructed corresponding to monthly values of rainfall (Rain), Southern Oscillation Index (SOI), Inter-decadal Pacific Oscillation (IPO), Niño 3.4 (Nino), maximum atmospheric temperature (MaxT), and minimum atmospheric temperature (MinT). Each unary data set comprised the current monthly value, plus twelve lagged values for the previous twelve months. Binary and ternary combinations of these unary sets were also used as inputs, Table 1. In this study a total of 62 combinations were tested, Table 1. Table 1 shows the range of Root Mean Square Error, RMSE, for unary, binary, ternary, quaternary and quinary inputs and also the specific combination that gave the lowest RMSE. RMSE gives a simple, transparent quantitative measure of the difference between the forecast rainfall and observed rainfall. The lower the RMSE the smaller the difference between the forecast and observed rainfall and therefore the more skilful the forecast.

unary data set was defined as the current monthly value of one of these input parameters, plus the twelve 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 MaxT (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, MaxT and Rain (SOI + MaxT + Rain).

In order to show the benefits of the larger data sets and lagging for the full 12 months, data sets that lagged for only 1, 2 and 3 months were also constructed and inputted into the neural network. Lagging for up to 3 months produced 4 input data columns for each unary set. These combinations are not shown in Table 1.



Table 1: Combinations of input variables tested in the neural network.

Input data sets	Number of Combinations	Range of RMSE (mm)	Combinations giving lowest RMSE values (mm)	
Unary	6	81.6 - 67.9	Rain	67.9
Binary	15	91.0 - 63.7	MaxT+IPO	63.7
Ternary	20	76.0 - 60.2	Rain+MinT+SOI	60.2
			Max+/MinT+SOI	61.9
			SOI+Nino+MinT	63.4
Quaternary	15	71.4 - 64.0	Rain+MinT+SOI+IPO	64.0
Quinary	6	72.4 - 55.4	Rain+MaxT+MinT+IPO+Nino	55.4
Total	62			

3 Results

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 $r = 0$ to 0.5 and show high variability both spatially and temporally [10]. Despite these poor correlations, single input variables, typically SOI, continue to be used by government agencies for seasonal rainfall forecasting in Australia. These forecasts are typically presented as shaded or coloured maps with a conditional probability of rainfall being greater than or less than a seasonal median value (e.g. fig 1). In order to generate these maps, numbers representing an actual rainfall forecast must first be generated. The government agencies, however, do not provide the actual numbers (i.e. the actual rainfall quantities) to the public or industry. That is, government agencies do not present their forecasts as time series charts (e.g. fig 5). In order to compare and contrast output from the neural network method of forecasting with output from the simple statistical models used by government agencies we begin by showing the likely output from government models as time series charts, figs. 2 and 3.

There is no reason to believe that output from a simple linear statistical model generating a rainfall forecast using only SOI, and lag-times of one, two or three-months as reported in the literature [e.g. 2], would give a rainfall forecast any better than output from a more sophisticated neural network. The rainfall forecast from the neural network using only SOI, the index favoured by the BoM, gives a poor representation of monthly rainfall for Nebo, fig. 2, reflected in the high values of RMSE of 78.8 mm and low value of $r = 0.26$.

The monthly rainfall forecast for Nebo is only improved marginally by introducing two additional climate indices as input, Nino and IPO, fig. 3, reflected in the slightly improved RMSE of 71.1 mm and $r = 0.50$. A rainfall

forecast for Nebo with a comparable level of skill can be derived from a neural network by simply using rainfall data as an input, fig. 4.

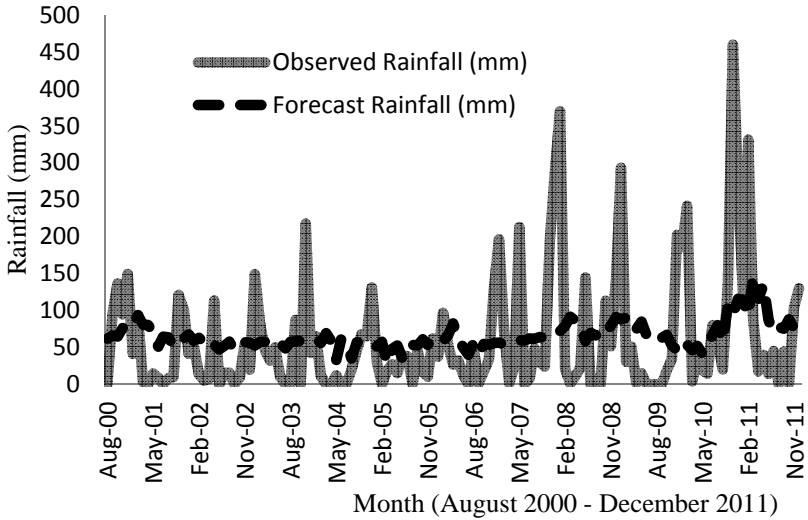


Figure 2: Monthly rainfall forecast for Nebo with one-month lead-time using one, two and three month lag-time for unary input data set SOI (RMSE 78.8 mm, $r = 0.26$).

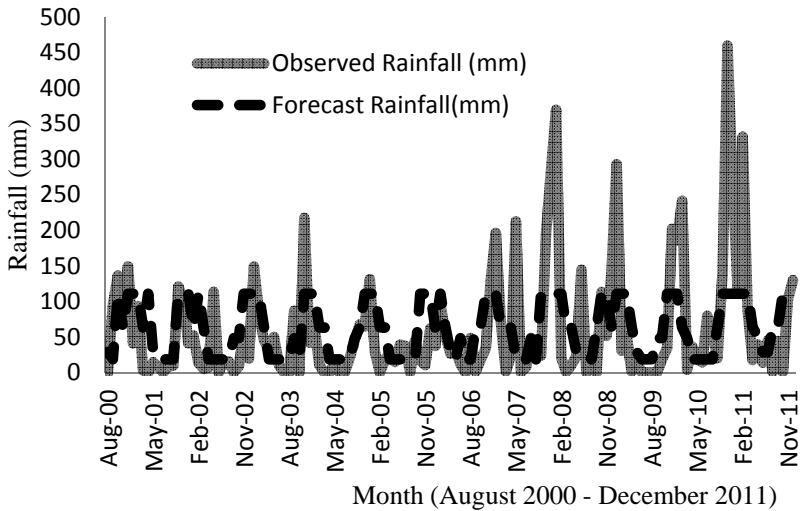


Figure 3: Monthly rainfall forecast for Nebo with one-month lead-time using one, two and three month lag-time for ternary inputs SOI, IPO & Nino (RMSE 71.1 mm, $r = 0.50$).



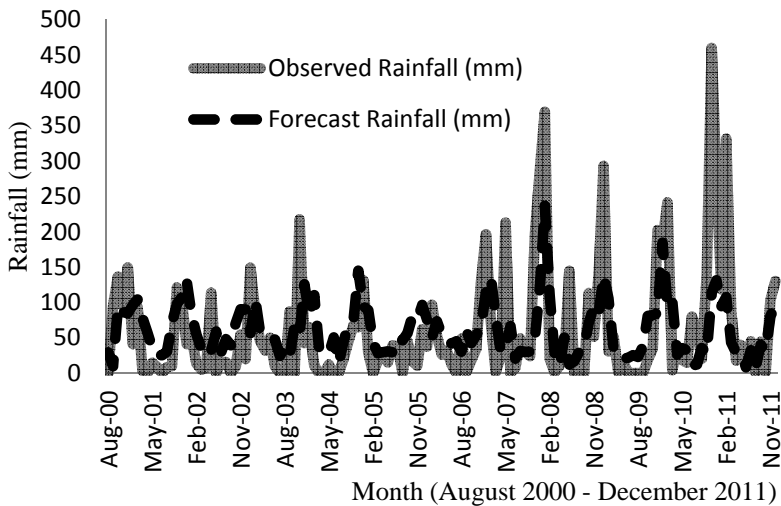


Figure 4: Monthly rainfall forecast for Nebo with one-month lead-time using only rainfall including lagged values to 12 months, without climatic indices (RMSE 67.9 mm, $r = 0.43$).

Comparing output from the neural network generated using just monthly SOI values as input, fig. 2, with the output generated using just monthly rainfall values as input, fig. 4, would suggest there are more useful patterns in the historical rainfall data than in the SOI index. Sensitivity analysis of the output for the forecast based only on rainfall data indicated that rainfall 11 and 12 months previously (i.e. 11 and 12 month lagged inputs) significantly influenced the forecast. It is notable that his output did not accurately forecast the flooding rains of January 2011.

The neural network approach has the advantage that it 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 an input 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.

It is possible to generate a much more skilled rainfall forecast for Nebo by inputting historic rainfall, local minimum atmospheric temperatures and also SOI reflected in the lower RMSE of 60.2 mm and $r = 0.67$, Table 1. Sensitivity analysis of this forecast indicated lagged minimum temperature had a significant influence on the forecast.

The most skilled rainfall forecast for the neural network was achieved with the quinary input data set comprising rainfall, maximum and minimum temperature, Niño 3.4 and IPO, Table 1. This represents the forecast with lowest value of RMSE for the 62 examples tested, Table 1. This forecast was further

improved through post-processing of the output, for example by combining two output profiles by computing a weighted average of forecast rainfall at each time period, fig. 5. In particular, a linear combination of the outputs shown from the ternary input data set (Rain + MinT + SOI) and the quinary input set (Rain + MaxT + MinT + Nino + IPO) that gave the forecast with the lowest RMSE. The outputs were combined in the ratio 0.7:1.3 and then the result divided by a factor of 2. The RMSE of 54.7 mm for the combination is lower than for either of the individual outputs.

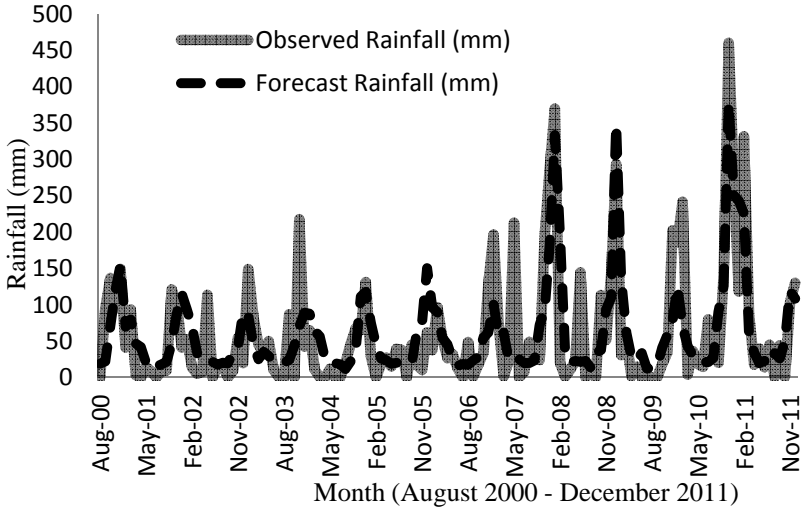


Figure 5: Monthly rainfall forecast for Nebo with one-month lead-time generated by combining two forecasts and post processing. (RMSE 54.7 mm, $r = 0.74$).

4 Discussion and conclusion

The basic method used to forecast seasonal rainfall in Queensland has not changed significantly for two decades [3]. The Queensland government and the BoM still use climate indices based on ENSO and simple linear statistical models to compute a conditional probability of rainfall being greater than or less than a seasonal median. The forecast, even if it is accurate for the 3-month period of the designated season, is largely meaningless for mine scheduling, because it gives no indication of the likely magnitude or distribution of rainfall within the season [1, 5]. For example, catastrophic flooding occurred in the Bowen Basin in December 2010 because rainfall was concentrated in time even though the overall seasonal rainfall may have been only marginally above the median

The situation at the Hail Creek mine was summarised in the Queensland Government Commission of Inquiries report:

“At the end of December 2010 at Hail Creek mine, water storages were calculated to be approximately 98 percent full; by the end of January 2011, the storage facilities were at 105 per cent capacity. The mine was storing approximately seven gigalitres of water on-site in dams and pits. Water was continually pumped from high priority areas into low priority areas around the site in an effort to maintain some operations. Rio Tinto sought, and was granted, authorisation by DERM [Department of Environment and Resource Management] to release water into surrounding watercourses in January 2011. All sale contracts were suspended by reason of the wet weather from 24 December 2010. Pits used for the purpose of storing water were unable to be mined until dewatering activities had begun; supplies of explosives were delayed; and resources were being deployed to address the water located in the pits. The sales suspension was lifted on 12 May 2011. As at September 2011, the mine was still not operating at full production on a sustained basis.”

The same report recommends that mine owners and government agencies need to complete risk assessments in preparation for wet seasons and that these assessments need to take place before 1 November each year [5]. However, central to any such assessments is access to a reliable and appropriate rainfall forecast. None currently exists, though it is assumed that the BoM and Queensland government seasonal forecasts are somehow adequate.

The big investments by the Australian government in computing power and technology in the climate science area have been in the development of general circulation models. These models, however, are even less reliable for rainfall forecasting than the simple statistical models used to advise the Queensland public and mining industry of the probability of 50-55% of above median rainfall for the summer of 2010-11.

Neural network based on artificial intelligence, like the simple statistical models, use historical data and climate indices, but neural networks represent a leap forward in computational power.

In this study 62 combinations of input variables were tested, Table 1. A quinary combination of rainfall, maximum and minimum temperatures, IPO and Niño 3.4 gave the most skilled forecast, Table 1. This forecast was further improved by combining it with output from a ternary input data set, fig. 5. This forecast differentiated to some extent between the recent wet summer months, and dryer summer months during the extended drought, fig. 5. The forecast, however, failed to forecast the wetter summer in 2009-10, fig. 5. It is likely that further refining of the methodology based on neural networks can improve the correspondence between forecast and observed rainfall profiles.



The output charted in figure 5 could be regarded as comprising multiple signals, oscillations on the y-axis with regards to time. In particular it may be useful to consider rainfall as composed of an underlying signal, e.g. figure 4, with forecast significantly improved by superimposing combinations of signals from climate indices and atmospheric temperature, figure 5. Improvements in understanding may be achieved by decomposition of signals at both the input and output stage.

Sensitivity analysis suggests that the inputs important to the model are not intuitively obvious. For example lagged-values to 10 and 11 month for minimum temperature were important to Figure 5.

The results presented in this study are considered preliminary with potential for significant improvement in terms of improving the skill of the monthly forecast one month in advance. Other studies of flood forecasting, specific to south-east Queensland, suggest that the neural network technique can be easily extended to a forecast with a three month lead-time (i.e. a forecast three months in advance) without a significant deterioration in skill [11].

In conclusion this study demonstrates that superior seasonal rainfall forecasts can be achieved by deploying appropriate modern mathematical techniques such as neural networks. Neural networks can accommodate the complexities of multiple non-linear relationships inherent in rainfall data. There is reason to be optimistic that the forecasts so far achieved using neural networks can be further improved, and provide a means for the mining industry in the Bowen Basin to have access to improved rainfall forecasts reducing the risk of significant economic losses associated with exceptional flood events.

Acknowledgement

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

References

- [1] Sharma, V., van de Graaff, S., Loechel, B. & Franks, D., Extractive resource development in a changing climate: Learning the lessons from extreme weather events in Queensland, Australia, National Climate Change Adaptation Research Facility, Gold Coast, pp. 110. 2013.
- [2] Schepen, A., Wang, Q. & Robertson J., Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rain, *Journal of Climate* 25(4) 1230-1246, 2012.
- [3] Fawcett, R.J.B., & Stone, R.J. A comparison of two seasonal rainfall forecasting systems for Australia, *Australian Meteorological and Oceanographic Journal* 60, pp. 15-24, 2010.
- [4] Queensland Climate Change Centre of Excellence. Monthly Climate Statement, November 2010, <http://www.longpaddock.qld.gov.au/seasonalclimateoutlook/qccecclimatedstatement/2010/20101112.html>



- [5] Queensland Government Flood Commission of Inquiry. Chapter 13 Mining 2012, <http://www.floodcommission.qld.gov.au/publications/final-report>
- [6] Abbot, J., & Marohasy, J., Application of Artificial Neural Networks to Rain Forecasting in Queensland. Australia. *Advances in Atmospheric Sciences* 29(4), pp. 717-730, 2012.
- [7] Verdon, D. C., & S. W. Franks, Long-term behavior of ENSO: Interactions with the PDO over the past 400 years inferred from paleoclimate records. *Geophysical Research Letters*, 33(6), L06712, 2006.
- [8] Power, S., T. Casey, C. Folland, A. Colman, & V.Mehta,V. Interdecadal modulation of the impact of ENSO on Australia. *Climate Dynamics*, 15, pp. 319–324, 1999.
- [9] Cai, W., & van Rensch, P., The 2011 southeast Queensland extreme summer Rain: A confirmation of a negative Pacific Decadal Oscillation phase? *Geophysical Research Letter* (39), pp. L08702, 2012.
- [10] Risbey, J.S., Pook, M.J. & Mcintosh, P.C., On the Remote Drivers of Rain Variability in Australia. *Monthly Weather Review* 137, pp. 3233- 3253, 2009.
- [11] Abbot, J., & Marohasy, J., The Application of Artificial Intelligence for Monthly Rainfall Forecasting in the Brisbane Catchment, Queensland, Australia. 7th International Conference on River Basin Management, 22-24 May 2013, New Forest.

