FORECASTING EXTREME MONTHLY RAINFALL EVENTS IN REGIONS OF QUEENSLAND, AUSTRALIA USING ARTIFICIAL NEURAL NETWORKS

JOHN ABBOT[1,2,3] & JENNIFER MAROHASY[1,2]
[1]Climate Modelling Laboratory, Noosa, Queensland, Australia.
[2]Institute of Public Affairs, Melbourne, Victoria, Australia.
[3]James Cook University, Townsville, Queensland, Australia.

ABSTRACT
Extreme rainfall in Queensland during December 2010 and January 2011 resulted in catastrophic flooding, causing loss of life, extensive property damage and major disruption of economic activity. Official medium-term rainfall forecasts failed to warn of the impending heavy rainfall. Since the flooding, the Australian Bureau of Meteorology has changed its method of forecast from an empirical statistical scheme to the application of a general circulation model (GCM), the Predictive Ocean and Atmospheric Model for Australia (POAMA). Our previous studies demonstrated that more skilful monthly rainfall forecasts can be achieved using artificial neural networks (ANNs). This study extends those previous investigations focussing on the capacity of the forecast methodology to differentiate between extreme rainfall events and more average conditions, up to one year in advance. Sites within two geographical regions of Queensland are examined: (i) coastal Queensland using rainfall observations from Bingera, Plane Creek and Victoria Mill; (ii) a region of south-east Queensland, using rainfall observations from 54 weather stations, extending approximately 300 km northward along the Queensland coast, from the Gold Coast to Bundaberg, and approximately 200 km inland. For both regions, the capacity to differentiate between average conditions and impending extreme rainfall events up to one year in advance is demonstrated.
Keywords: artificial neural network, flood, forecast, rainfall, Queensland.

1 INTRODUCTION

Prolonged rainfall over large areas of Queensland led to flooding of historic proportions in December 2010, extending into January 2011 [1, 2]. Thirty-three people died in those 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 A$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].

In January 2011 Brisbane, the state capital of Queensland, experienced its second highest flood in over 100 years. Major flooding occurred throughout most of the Brisbane River catchment, with an estimated 18,000 properties inundated in metropolitan Brisbane, Ipswich and elsewhere in the Brisbane River Valley [2].

This event has been termed a "dam release flood" [2], suggesting that the sudden release of water from the Wivenhoe Dam was a principal cause of flooding. The decision to suddenly release water resulted because the capacity of the Wivenhoe dam did not allow for the extreme rainfall event, and the official medium-term rainfall forecasts available were inadequate.

The official rainfall forecasts issued by the Australian Bureau of Meteorology (BOM) did not provide sufficiently accurate rainfall forecasts at the regional or localised levels, or with sufficient lead times. The BOM forecasts are only provided to the public in the form of prob-

© 2017 WIT Press, www.witpress.com
ISSN: 1743-7601 (paper format), ISSN: 1743-761X (online), http://www.witpress.com/journals
DOI: 10.2495/SDP-V12-N7-1117-1131

abilities relative to the median, and do not differentiate between anticipated rainfall slightly above the median and extreme rainfall events such as occurred in January 2011 [3]. Until May 2013, official forecasts were based on a statistical scheme using an El Niño Southern Oscillation (ENSO) index as the primary predictor. The BOM switched to the use of the Predictive Atmospheric Model for Australia (POAMA), a general circulation model (GCM), in June 2013 [4, 5]. GCMs, however, do not generally perform well at forecasting rainfall, despite substantial efforts to enhance skill over three decades [6, 7].

Artificial neural networks (ANNs), a form of machine learning, provide an alternative technique for medium-term rainfall forecasting both in Australia [8–10], and in other regions of the world [11–13]. Our previously reported investigations applied neural networks to individual sites within Queensland with long rainfall records to generate forecasts of monthly rainfall with lead times up to 12 months [8–10]. A recent study [14] focussing on the sites of Harrisville and Gatton in south-east Queensland clearly demonstrated that superior monthly rainfall forecasts can be generated using the ANN approach when compared to results generated by POAMA. Forecasts generated using the ANN approach were consistently more skilful than climatology, as shown by lower root means square error (RMSE) and mean absolute error (MAE) values. In contrast, forecasts from POAMA generally had skill levels only about equivalent to climatology (the long term average).

The present study extends our previous investigations [14], focussing on the capacity of the ANN methodology to differentiate extreme rainfall events from more average conditions. Two regions of Queensland were considered: (i) the coastal region of Queensland extending from Mossman in the far north of the state, through central Queensland, to Maryborough in the south; (ii) the south-eastern region of Queensland, where 54 individual sites were used to generate regional isohyet maps, illustrating the capacity of the methodology to distinguish between extreme and more average rainfall conditions.

2 DATA AND METHODS

The first region of Queensland considered in this study is the coastal strip extending from Mossman in the far north to Maryborough in the south, a distance of about 1500 km, as shown in Fig. 1. The second region is an area of south-east Queensland enclosed by longitudes 151.0°E and 153.5°E, and latitudes 24.4°S and 28.5°S, as shown in Fig. 2. This region extends about 300 km southwards along the Queensland Coast from Bundaberg to the Gold Coast. This region extended inland from the coast to include the towns of Dalby, Inglewood and Mundubbera, a distance of about 200 km. The region considered is therefore approximately the same area as a single grid forecast area used by POAMA (250 km × 250 km) [5].

The skill of a rainfall forecast using any statistical model, including ANNs, will depend on the quality and relevance of the data input provided to the model, with a longer time series generally giving superior forecasts. Our previously reported studies [14] focussed on Gatton and Harrisville have reported monthly rainfall forecasts using neural networks. These investigations have demonstrated that using combinations of inputs including both climate indices, and also local variables such as maximum and minimum temperatures [14] produces the best forecasts.

Monthly rainfall data were obtained from the Australian BOM's Climate Data Online for 8 sites of coastal Queensland and the 54 locations in south-east Queensland shown in Fig. 2. These sites were chosen on the basis of their geographic spread in the regions under consideration, and also the quality of data, that is the desirability of long data series with few missing values.

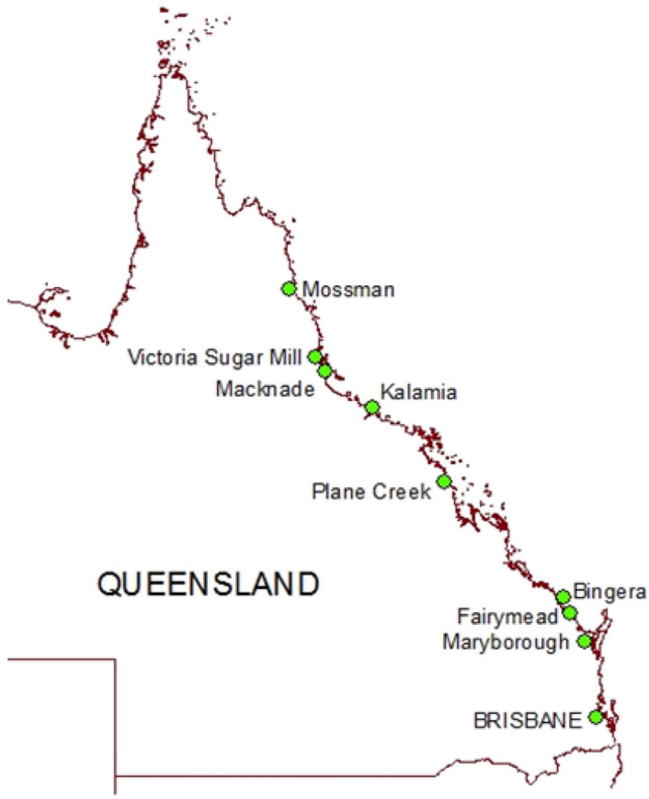

Figure1: Locations of rainfall observation sites for coastal Queensland, and Brisbane, the capital of Queensland.

Variations in rainfall in many parts of the world are associated with large-scale climate phenomena which can be described by climate indices. 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.

All four Niño indices were used in this study with values 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. Pressure differences are typically represented by the SOI, calculated as the difference between Tahiti and Darwin, with values for this study obtained from the BOM website. A sixth Climate Index, the IPO, was also used as input with values sourced directly from Chris Folland at the UK Met Office.

All of the above climate indices with local rainfall and temperatures were inputted into *Neurosolutions Infinity* software and used to build ANN models. Each attribute was lagged up to 12 months. The data were divided into training (75%), evaluation (15%) and test sets (10%).

Figure 2: Rainfall observation sites in south-east Queensland.

There is no established way to determine in advance which ANN model configuration will perform best for a specific problem. Some pre-selection of inputs can be undertaken, for example by examination of linear correlations between each input and the target output. However, as ANN models can accommodate non-linear relationships this procedure is not necessarily advantageous, and may eliminate useful inputs. The *Infinity* program is highly automated, sequentially testing hundreds of candidate neural network model configurations, and thousands of combinations of inputs, to select the optimised output. The *Infinity* program uses a pre-set formula incorporating RMSE, MAE and correlation coefficient (r) to evaluate the accuracy for each neural network model and a corresponding set of selected inputs tested. Based on this formula, the program determines which ANN model and set of inputs is optimal.

The present investigation used the same neural network methods previously described for Gatton and Harrisville [14], applied to the individual stations in the two regions under consideration. Each of these selected stations had rainfall record extending back until at least 1920.

Two approaches have previously been investigated for neural network optimization of monthly rainfall forecasting in Queensland [14]. With the first approach, designated as "all-month optimization", data for all 12 months of the year were included as input and optimised together, as in our previous studies [8–10]. With the second approach, designated as "single month optimisation", forecasts corresponding to each calendar month were performed individually [14]. The monthly rainfall forecast skill using the two approaches was previously compared for each month for two of the sites used in the present study, Gatton and Harrisville. For months with the heaviest rainfall, January and December the skill level was significantly higher, as shown with lower Mean absolute Error and Root Mean Square Errors. The single-month optimization method was therefore applied to sites used in the present investigation for sites in coastal Queensland for each month of the year and to generate December forecasts in south-east Queensland.

For the coastal Queensland study, monthly rainfall forecasts were generated for each month of the year at lead times of 3 months and 12 months for the sites of Bingera, Plane Creek and Victoria Mill. The results were plotted as a time series over ten years for a test period November 2004 to September 2014. This enabled comparisons of observed and forecast rainfall during periods of extreme rainfall events with periods of average, or below average rainfall.

In order to measure the relative reduction of the error of the forecasts over a climatology-based forecast an RMSE skill score can be calculated according to eqn (1):

$$SS = (RMSE_Climatol. - RMSE_model)/RMSE_Climatol. \times 100\% \quad (1)$$

If a skill score SS is positive, then the RMSE of the forecast is smaller than the RMSE of the corresponding climatology forecast, and therefore the ANN forecasts provide better accuracy than simply using the long-term mean monthly rainfall. Conversely a negative skill score indicates a forecast accuracy below climatology. Therefore a skill score of 0% is equivalent to climatology, whereas a score of 100% represents a perfect forecast, equivalent to the observed value. Negative scores represent skill levels below climatology. Equation (1) is analogous to an equation used in studies where a skill score has been calculated for forecasts based on POAMA [5, 15].

For the coastal Queensland study, skill scores were generated for the locations of Bingera, Plane Creek and Victoria Mill at 3 months and 12 months lead time.

For the south-east Queensland study, monthly rainfall forecasts for December were obtained for each of the 54 stations with a lead time of 12 months. The forecast results were examined in detail for the selected years of 2005 and 2010, representing examples of average and extremely high December rainfall respectively across the region. The results were displayed using both regional rainfall isohyets and also bar charts representing observed and forecast rainfall amounts at each site, using *TeraPlot* software. This enables visual comparisons to be made between observed and forecast rainfall for each year, and also between years.

3 RESULTS AND DISCUSSION

3.1 Coastal Queensland region

Figure 1 shows the locations of 8 sites in coastal Queensland where long rainfall records are available. Rainfall patterns in this region are characterised by high variability in annual rainfall, as illustrated in Fig. 3. Large variations in rainfall are also observed throughout the year,

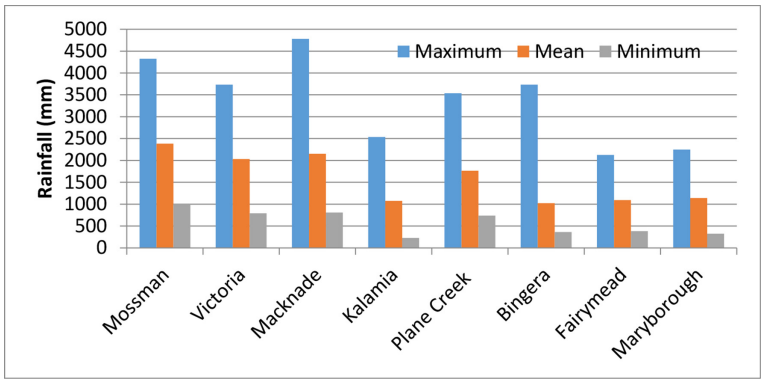

Figure 3: Variation in annual rainfall at locations in coastal Queensland.

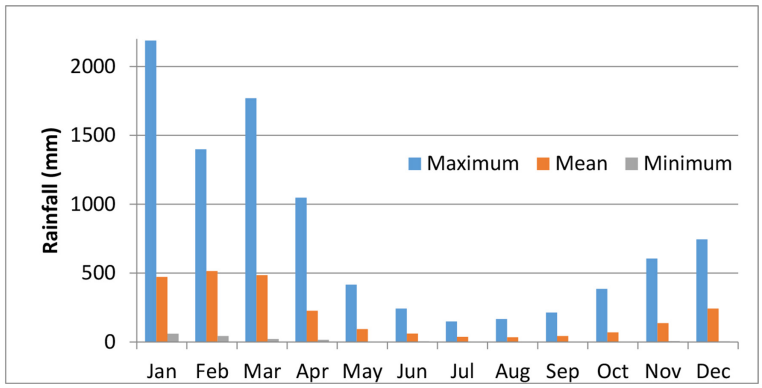

Figure 4: Variation in monthly rainfall at Plane Creek.

with heavy rainfall in the summer months and low rainfall during the winter, as illustrated in Fig. 4 for Plane Creek.

Table 1 shows forecast skill scores calculated using eqn (1) at lead times of 3 months and 12 months for three of the locations of coastal Queensland region. A skill score of 0% is equivalent to climatology, while 100% represents a perfect forecast. Skill scores for the same region generated using POAMA fall in the range between –20% to 20% [5] averaging about the same level as climatology (0%) at long lead times. The skill score in Table 1 shows all values above 25%, ranging up to 84%. Clearly the skill of the methodology using ANNs to forecast monthly rainfall is superior to that demonstrated by a GCM [5].

Figure 5a and b shows comparisons of observed and forecast monthly rainfall with a three month lead time for Bingera over a test period of about 10 years. Figure 5a and b shows a similar comparison for Plane Creek with a lead time of 12 months. These figures illustrate the superior forecast methodology using the single-month optimisation as opposed to all-month optimisation [14], illustrated with higher coincidence between observed and forecast lines in Figs 5b and 6b compared to Figs 5a and 6a respectively. These results clearly show that the neural network forecast methodology is capable of differentiating between extreme rainfall events that typically occur during the summer periods that have average, or below average, monthly rainfall.

Table 1: Skill scores for Bingera, Plane Creek and Victoria Mill at 3 and 12 months lead times.

	Skill score (%)					
Month	Lead 3 months			Lead 12 months		
	Bingera	Plane Creek	Victoria Mill	Bingera	Plane Creek	Victoria Mill
January	79	34	28	68	47	81
February	78	34	48	64	87	87
March	51	42	79	57	31	79
April	77	45	38	44	73	71
May	50	82	40	67	57	58
June	77	57	78	27	43	59
July	65	68	38	53	54	48
August	40	50	57	74	57	74
September	67	52	25	22	23	48
October	26	47	84	42	92	59
November	78	61	38	25	74	70
December	41	50	53	31	39	74

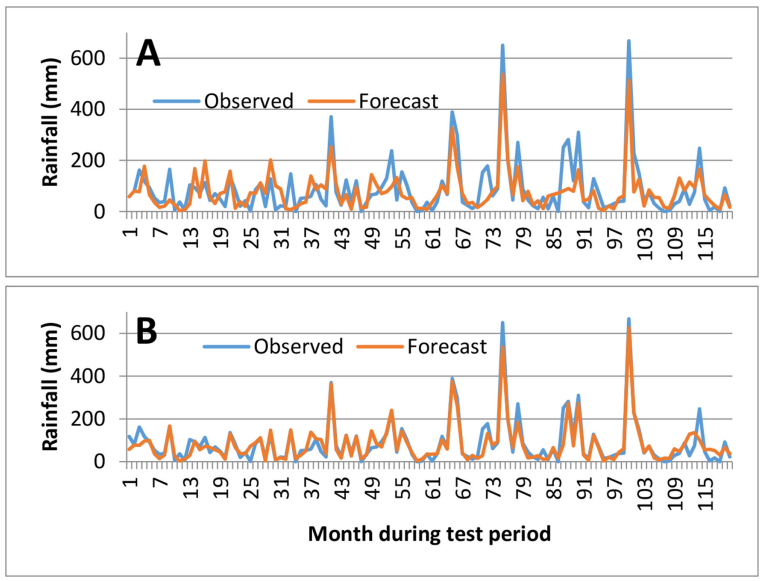

Figure 5: Observed monthly rainfall for Bingera and forecast monthly rainfall at 3 months lead time for test period November 2004 to September 2014. A: all-month ANN optimization; B: single-month ANN optimization.

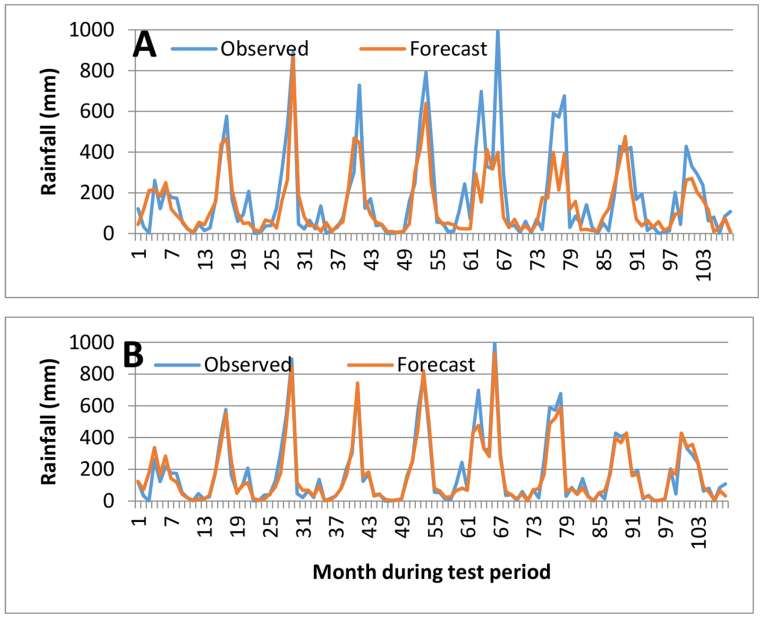

Figure 6: Observed monthly rainfall for Plane Creek and forecast monthly rainfall at 12 months lead time for test period November 2005 to September 2014. A: all-month ANN optimization; B: single-month ANN optimization.

3.2 South-east Queensland region

Figure 2 shows the locations of 54 sites in south-east Queensland where long rainfall records are available. Rainfall patterns in this region are characterised by alternating extended periods of comparative drought followed by flooding rains. This high variability in rainfall is generally reflected in wide variations in monthly rainfall. For example, this is illustrated in Fig. 7 for the town of Gympie, where observed December rainfall is shown for the period

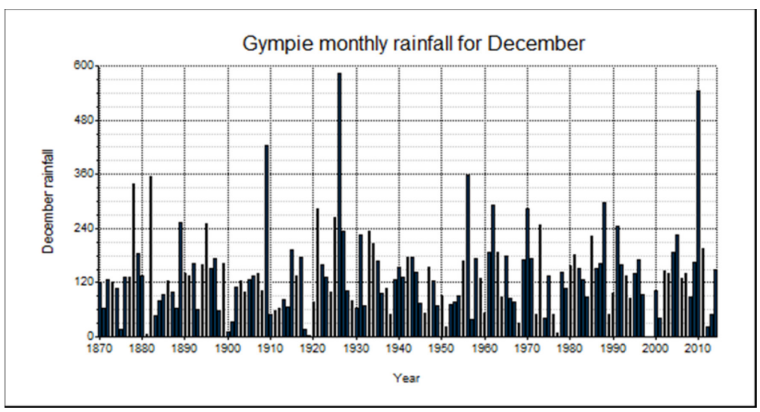

Figure 7: December monthly rainfall for Gympie between 1870 and 2014.

from 1870 and 2014. The lowest recorded December rainfall for Gympie was 1.7 mm in 1919, and the highest 584.5 mm in 1926, with a mean value of 137.4 mm. It is important to develop forecasting systems with long lead times that can differentiate between extreme rainfall events and more average rainfall for a particular selected month, for example to give adequate warning of impending flood conditions.

Figures 8 and 9 show results for observed and forecast rainfall respectively, for the period of very heavy rainfall in December 2010, at 12 months lead time. These results are presented

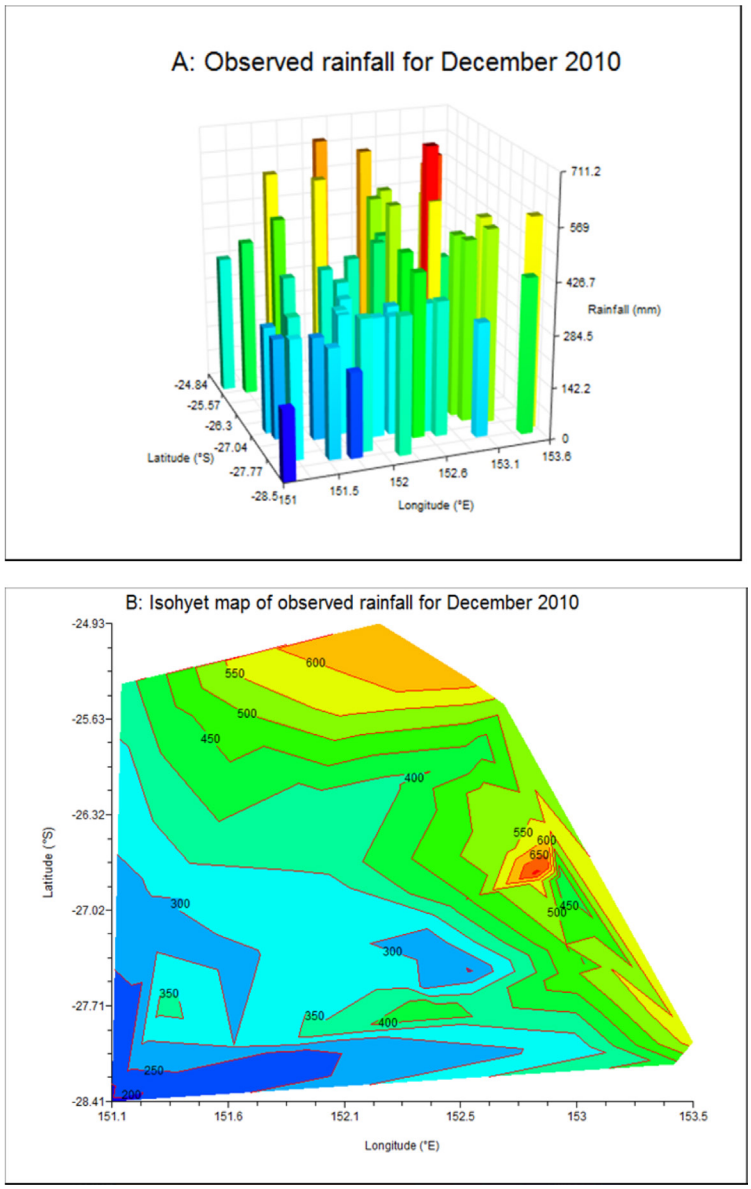

Figure 8: Observed rainfall (mm) for December 2010 for the south east Queensland region. A: bar chart for individual sites; and B: isohyet map with 50 mm interval spacing.

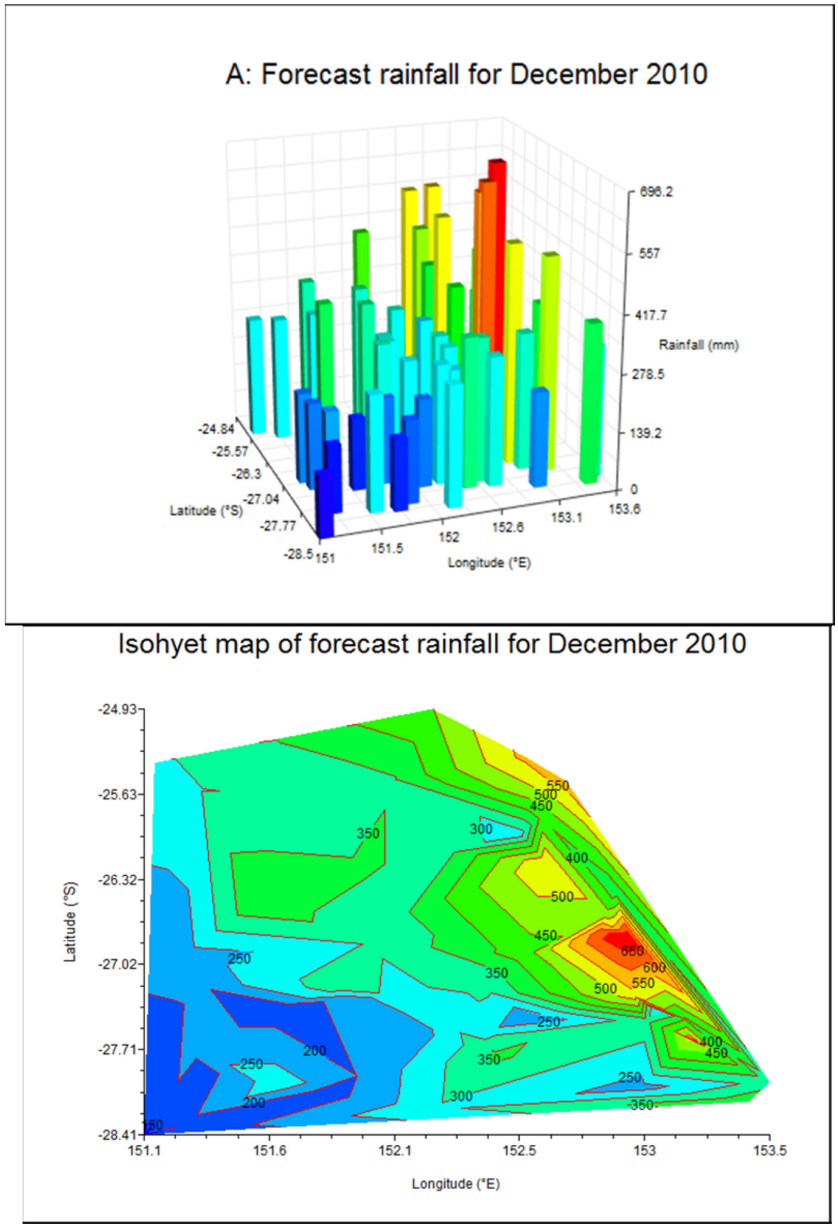

Figure 9: Forecast rainfall (mm) for December 2010 for the south east Queensland region. A: bar chart for individual sites; and B: isohyet map with 50 mm interval spacing.

as A: bar charts with rainfall values for individual sites represented by a single coloured bar; and also B: regional isohyet maps derived from the bar charts using *TeraPlot* software. Visual inspection of Figs 8 and 9 shows that the heaviest rainfall is concentrated along the coast, with a maximum observed and forecast values above 600 mm. For both observed and forecast rainfall, the highest rainfalls were concentrated around the locations of Mapleton and Maleny.

Moving inland from the coast, levels of observed and forecast rainfall decreased 350 mm or less.

Figure 10 shows comparisons between observed and forecast monthly rainfall for each of the 54 individual sites considered for December 2010. Over this region, the difference between observed and forecast monthly rainfall values ranged between 1% and 40% with a mean of 18.7%.

Figures 11 and 12 show results for observed and forecast monthly rainfall respectively, for December 2005, again at 12 months lead time, representing a period of average rainfall for the region. Visual inspection of these figures shows that heavier rainfall is again concentrated along the coast, but with a maximum observed and forecast values in the 200–300 mm range. Inland, both forecast and observed values are typically in the 50–100 mm range.

Comparison between Figs. 9 and 11 show the significant difference between observed rainfall across the region when considering a period of extremely heavy rainfall and a period of more average rainfall. Across the 54 stations considered the rainfall increased by a factor of 4.8 on average when comparing 2010 and 2005. Comparison of Figs. 9 and 12 shows that this difference is also successfully reflected in the forecast, with the corresponding factor calculated as 3.7. The neural network forecasts tend to somewhat underestimate the quantitative amount of rainfall during the period of extremely heavy precipitation. Nevertheless, this forecast method does provide a warning of very heavy rainfall one year in advance and provides rainfall at the localised level across the region.

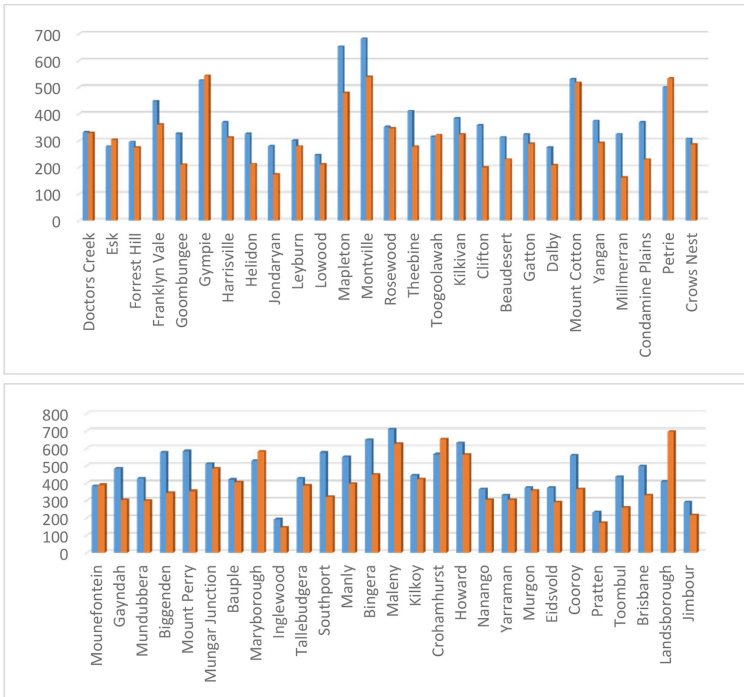

Figure 10: Observed (blue) and forecast (red) monthly rainfall for December 2010 for 54 sites in south-east Queensland.

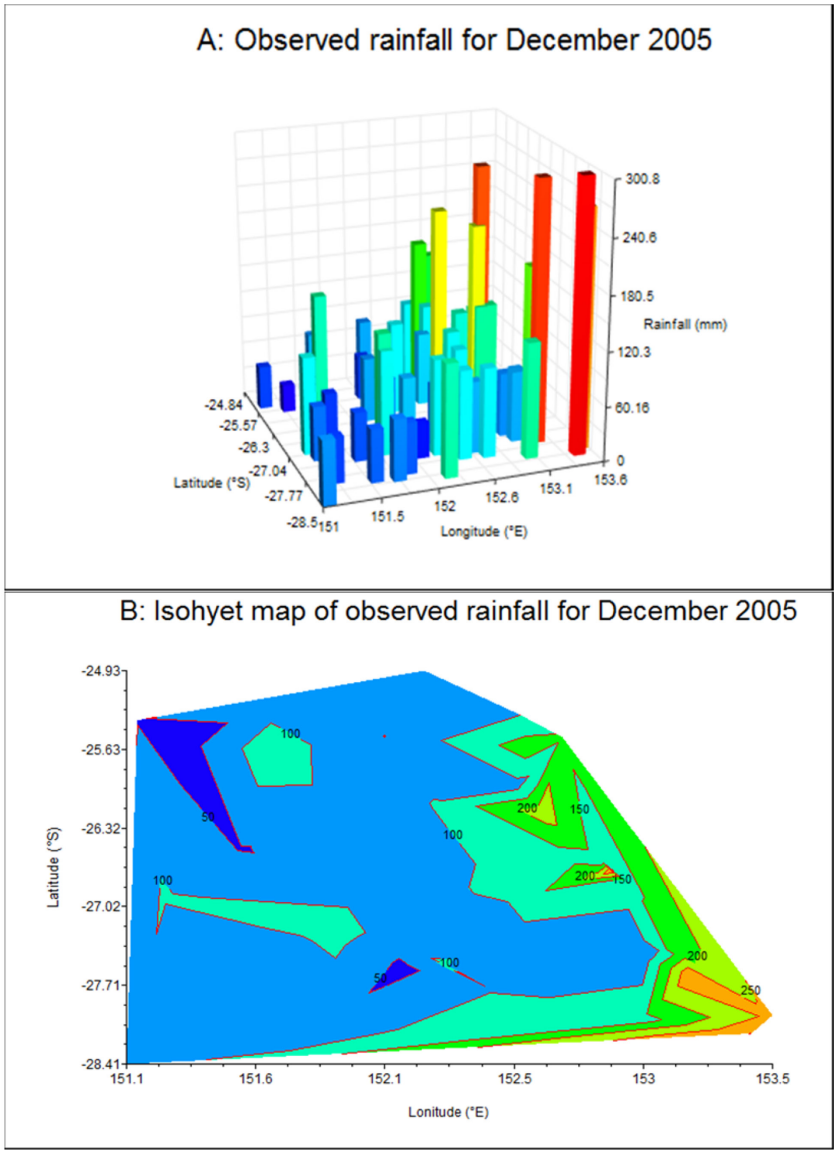

Figure 11: Observed rainfall (mm) for December 2010 for the south east Queensland region. A: bar chart for individual sites; and B: isohyet map with 50 mm interval spacing.

The isohyet maps also enable the generation of forecasts for locations within the geographic region where there are either no weather stations, or the historical records are too short to provide adequate data to train a neural network. This approach could thus be used to provide total rainfall estimates over specific catchment areas, providing important information for dam and water storage operations.

Since January 2010, official forecasts have been based on general circulation models rather than statistical models. However, GCMs generate rainfall forecasts at relatively coarse

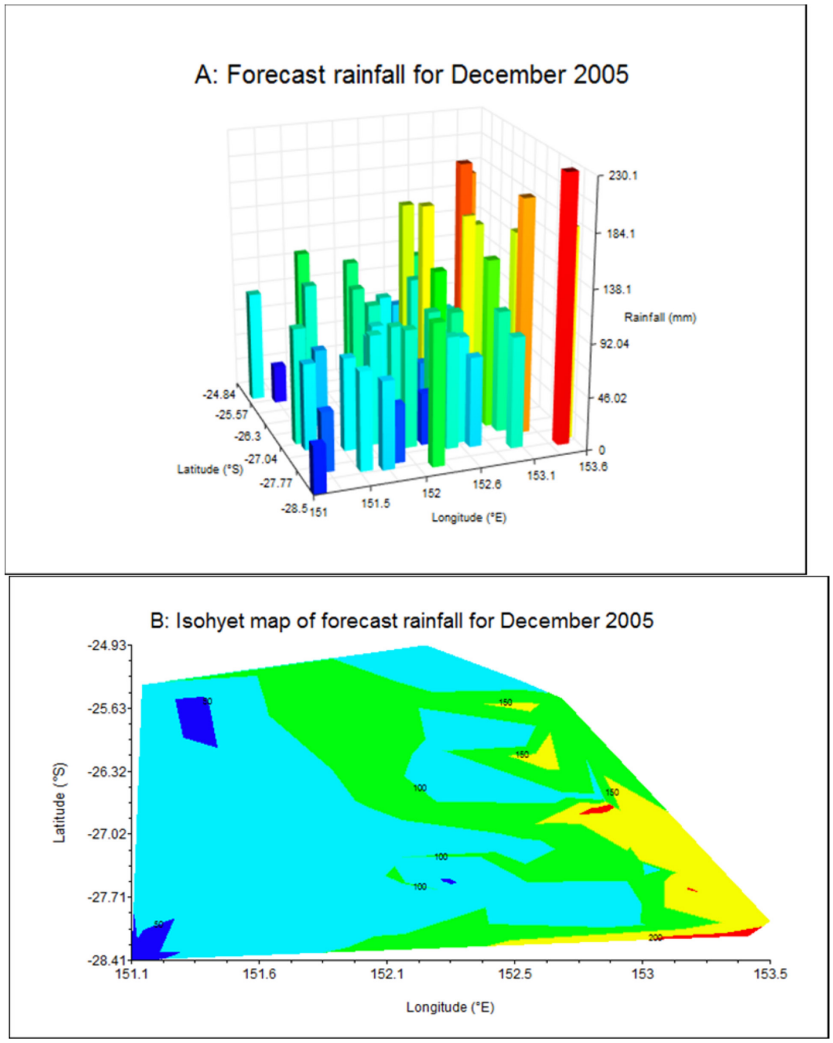

Figure 12: Forecast rainfall (mm) for December 2005 for the south east Queensland region. A: bar chart for individual sites; and B: isohyet map with 50 mm interval spacing.

spatial scales, in the order of a few hundred kilometres and are unable to resolve the effects of sub-grid scale features such as topography and land use [16]. These outputs cannot be directly used in catchment scale studies, which require hydroclimatic data at fine spatial resolutions [17]. The scale mismatch between the GCM outputs and the hydroclimatic required at the catchment level is a major obstacle in climate studies of hydrology and water resources [17].As a solution to the scale mismatch between the GCM'soutputs and the information required at the catchment scale, downscaling techniques have beendeveloped [16, 17]. However, results from using this approach applied to catchments in eastern Australia [15] demonstrate little improvement over climatology, a major limitation being the low level of skill in the monthly rainfall forecasts from the general circulation model at the course grid scale [5].

This study provides a methodology where forecast results from a large number of individual stations are combined in an upscaling process to provide a high level of accuracy for a catchment or regional areas. The approach should enable rainfall forecasts that provide sufficient information to avoid the scenario of the catastrophic Wivenhoe dam release flooding in January 2011.

4 CONCLUSION

Previously reported studies using ANNs to forecast monthly rainfall in Queensland [14], presented the results for individual sites with long rainfall records. The present study extends this approach to present monthly rainfall forecasts over extended geographical regions along coastal Queensland, and also in south-east Queensland, with up to 12 months lead time. In both regions the ANN approach enables a clear differentiation between extreme rainfall events and more average conditions. This is illustrated with reference to extreme rainfall events such as occurred in December 2010, and more average rainfall such as occurred in December 2005 in south-east Queensland. Forecast skill using the ANN methodology is significantly better than reported for monthly rainfall forecasts at long lead times using a general circulation model. Skill scores fall in the range 25% to 80% are achievable using ANN methodology. In contrast, skill levels are reported between –20% and 20% for a GCM, and are generally only about equivalent to climatology. The lack of adequate official forecast capacity prior to the extreme rainfall in December lead to poor decisions regarding operations of the Wivenhoe Dam and exacerbated the destruction caused through inadequate storage capacity being available to accommodate extreme rainfall. This scenario could potentially have been significantly mitigated if better forecasting methodology had been applied at that time.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Queensland Flood Commission of Enquiry, Final Report 2012. http://www.floodcommission.qld.gov.au/__data/assets/pdf_file/0007/11698/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. https://doi.org/10.3390/w3041149

[3] Abbot, J. & Marohasy, J., The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia. *Water Resources Management VII Book Series: WIT Transactions on Ecology and the Environment,* **171**, pp. 287–297, 2013. https://doi.org/10.2495/WRM130261

[4] 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. https://doi.org/10.1175/MWR-D-13-00248.1

[5] 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. https://doi.org/10.1002/wrcr.20453

[6] 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.
https://doi.org/10.1002/hyp.7916

[7] Shao, Q. & Li, M., An improved statistical analogue downscaling procedure for seasonal precipitation forecast. *Stoch Environ Res Risk Assess,* **27,** pp. 819–830, 2013.
https://doi.org/10.1007/s00477-012-0610-0

[8] Abbot, J. & Marohasy, J., Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences,* **29**(4), pp. 717–730, 2012.
https://doi.org/10.1007/s00376-012-1259-9

[9] 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.
https://doi.org/10.1016/j.atmosres.2013.11.002

[10] Abbot, J. & Marohasy, J., Forecasting of monthly rainfall in the Murray Darling Basin: Miles as a case study. *River Basin Management*, **VIII**, pp.149–159, 2015.
https://doi.org/10.2495/RM150141

[11] 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.
https://doi.org/10.2478/s11600-011-0044-y

[12] Dahamsheha, A. & Aksoy, H., Artificial neural network models for forecasting intermittent monthly precipitation in arid regions. *Applied Meteorology,* **16**, pp. 325–337, 2009.
https://doi.org/10.1002/met.127

[13] 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.
https://doi.org/10.1016/j.atmosres.2011.06.013

[14] Abbot, J. & Marohasy, J., Improving monthly rainfall forecasts using artificial neural networks and single-month optimisation: a case study of the Brisbane catchment, Queensland, Australia. *Water Resources Management VIII: WIT Transactions on Ecology and the Environment*, pp. 3–13, 2015.

[15] Schepen, A. & Wang, Q.J., Ensemble forecasts of monthly catchment rainfall out to long lead times by post-processing coupled general circulation model output. *Journal of Hydrol*ogy, **519,** pp. 2920–2931, 2014.
https://doi.org/10.1016/j.jhydrol.2014.03.017

[16] Rashid, M.M., Beecham, S. & Chowdhury, R.K., Statistical downscaling of CMIP5 outputs for projecting future changes in rainfall in the Onkaparinga catchment. *Science of the Total Environment,* **530,** pp. 171–182, 2015.
https://doi.org/10.1016/j.scitotenv.2015.05.024

[17] Willems, P. & Vrac, M., Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology*, **402,** pp. 193–205, 2011.
https://doi.org/10.1016/j.jhydrol.2011.02.030

[18] 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.