Monthly rainfall forecasting using neural networks for sugarcane regions in Eastern Australia

Ali Haidar and Brijesh Verma

ABSTRACT

Sugarcane is an important agricultural crop grown on the east coast of Australia. The timing and amount of rainfall is critical in determining both the yield of sugar and scheduling of harvesting operations. Rainfall forecasts issued through the Australian Bureau of Meteorology are based on general circulation models (GCMs) and have a poor skill levels. They are also limited in utility to end-users such as farmers as they cover very broad geographical areas and are only issued as probabilities above or below median. This paper presents an alternative approach for forecasting monthly rainfall with up to 12 month lead-time based on machine learning, in particular neural networks. Monthly rainfall forecasts have been developed for the eight locations in Eastern Australia at 3 and 12 month lead-time. The accuracy of the forecasts has been assessed relative to a skill scale with 0% representing climatology (the long term average) and 100% representing a perfect forecast (observation). On this scale, neural network forecasts are typically in the range 39.9–68% for all months using a single month optimization. This compares very favorably with forecasts using GCM from the Bureau that have skill levels only in the range −20% to 20%.

Key words | artificial neural networks, rainfall forecasting, water management

Ali Haidar (corresponding author)
Brijesh Verma
Centre for Intelligent Systems, School of Engineering and Technology,
Central Queensland University,
Sydney,
Australia
E-mail: *a.haidar@cqu.edu.au*

INTRODUCTION

Weather forecasting is vital for many aspects in agriculture (Shukla *et al.* 2011). Climate is a key driver for sugarcane productivity (Everingham *et al.* 2003). Everingham *et al.* reported that early knowledge about the climate could add value to production, harvest and marketing efforts (Everingham *et al.* 2003). Decision-making based on rainfall forecasts can increase profitability of sugarcane seasons. Delivering accurate seasonal and annual rainfall forecasts in sugarcane industry processes such as plantation, fertilization, irrigation, and harvest lead to larger yields, therefore higher profit. Through the season, there are a number of industry operations carried out that are timed to suit typical rainfall conditions (Everingham *et al.* 2012). Outputs from long lead forecasting models can be beneficial to industry planning early in the year (Everingham *et al.* 2007). The ability to predict the time and the amount of precipitation before the start of crop seeds allows growers to set schedules for planting, in addition to avoiding seed damage due to wet or dry weather after plantation.

Sugar cane is a plant that requires water to grow. Usually, based on the seasonal amount of precipitation, growers decide whether to include irrigation. Precise rainfall prediction gives growers the opportunity to effectively manage irrigation where they would buy water, identify the best suitable machines, and sign contracts for maintenance earlier if forecasts showed that the amount of rainfall through the season would not be sufficient. Therefore, growers would avoid additional fees for higher water prices during the season, and permit irrigation companies to be scheduled earlier and not overwhelmed. On the other side, if forecasts revealed a wet season, growers would avoid signing contracts with irrigation companies, and paying extra amounts for season insurance.

Fertilization is the process of adding chemicals and pesticides to sugar areas, in addition to spraying fields at

doi: 10.2166/ws.2016.099

different stages during the season. Growers must be careful before adding fertilizers to the ground. If heavy rainfall were detected after adding these chemicals to the fields, it would be dragged to river water and cause pollution. This may produce disastrous effects on river plants. Spraying time is crucial for growers. If there is heavy rainfall and wind, spraying may not lead to satisfiable consequences. Moreover, if there is a dry season, spraying may burn the plants if there is no irrigation.

In Australia, the sugarcane-harvesting season starts around June and continues up until late September. Information about the start day of harvest should be given no later than March (Clarke *et al.* 2010). Ability to identify the type and amount of rainfall during this part of the season contributes to managing this process. Growers would take a decision to start early if forecasts indicated a wet season, or to start first with paddocks prone to floods. In addition to these forecasts, encourage growers to sign maintenance contracts and setup transport chains earlier with higher numbers of containers to move crops. In the harvest season, wet yield can cause additional work in the mills to clean it up. Sugarcane is a perennial crop, where harvesting in wet weather may damage the sugarcane crop, therefore reducing its profitability during future seasons (Everingham *et al.* 2003). Wet conditions in the 1998 harvest season diminished the industry revenue by around 175 million dollars (Everingham *et al.* 2007). The loss was reported as being due to cane being left unharvested, reduced commercial cane sugar levels, and damage to paddocks from wet weather harvesting. On the other hand, if forecasts designated a dry season, growers would start harvesting the driest areas first to avoid loss of crops at the end of the season.

In addition to seasonal and annual, long decadal rainfall prediction enhances decisions related to investment on irrigation structure. Millers would benefit from data to improve long-term management of bagasse supplies. Furthermore, marketers could utilize this data to help choices related to buying expensive items as storage infrastructure.

Sugarcane is a plant that can regrow for five consecutive seasons before being ploughed out from the ground (Clarke *et al.* 2010). Australia is the third largest raw sugar supplier in the world. In Australia, 95% of sugarcane grows in Queensland, a state located in the north-east of the continent. Sugarcane areas are located along 2,100 km coastline extending from 16 to 29 °S (Everingham *et al.* 2003). Australian sugar areas are basically situated along Australia's eastern coastline, from Mossman in far north Queensland to Grafton in northern New South Wales. Sugarcane production in Australia occurs from tropical regions in northern Queensland to the sub-tropical environment of northern New South Wales. Sugar industry generates approximately 1 billion AUD in export earnings (Everingham *et al.* 2012).

Decision-making in water management is crucial, not only in the sugar industry but also for many different aspects such as dam management. Precipitation contributes in a high degree to water management. In 2010, disastrous floods occurred in Brisbane, the capital city of Queensland. It has been stated that the floods were 'dam-release' floods. Release of water from the Wivenhoe dam (located at the Brisbane River in South East Queensland, Australia) was the principle cause of flooding along the mainstream and tributaries of the Brisbane River downstream of the dam over the period 11–12 January 2011 (van den Honert & McAneney 2011). Taking into consideration the effect of rainfall on water level in Brisbane catchment, a different strategy would have been taken (van den Honert & McAneney 2011).

LITERATURE REVIEW

Rainfall forecasts play a key role in water management. Rainfall forecasting models have been developed in an attempt to predict precipitation for a specific location. Operational climate forecasts in Australia have been available for more than a decade (Everingham *et al.* 2007). Various types of forecasting systems have been developed to improve knowledge and assist with decision-making in the sugar industry. These models predict for various durations of time. Decadal, annual, and seasonal forecasts would be helpful to the industry. The Bureau of Meteorology (BOM) is an executive agency of the Australian Government in charge of giving climate administrations to Australia and encompassing ranges. BOM uses general circulation models (GCMs) to produce official rainfall forecasts across Australia. GCMs are dynamic systems that use computer simulation for rainfall forecasting. These models failed to forecast rainfall variability in different periods of time, especially in 2010 where heavy rainfall occurred along different areas in Queensland. The Australian Bureau of Meteorology has developed a dynamical forecasting system, POAMA (Predictive Ocean-atmosphere model for Australia). A seasonal to inter-annual forecasting system based on a coupled model of the ocean and atmosphere (Cottrill *et al.* 2012), POAMA has been established to provide seasonal forecasts, and it was introduced officially in 2013. POAMA (version P2.4) is reset at the beginning of each month, and then forecasts are released for that month and the next 8 months (Hawthorne *et al.* 2013). These forecasts are probabilistic, where probability values are launched across large grid areas (250 km). They are considered insufficient for the sugarcane industry, as they give no indication about the magnitude, time and specific location of precipitation.

Researchers tried different techniques to setup models for rainfall forecasting. Climate indices have been used to build forecasting models. Scientists claim that El-Niño southern oscillation (ENSO) is a contributor to rainfall variability over the world. Clarke *et al.* developed a model using ENSO to predict weather for the sugar season, earlier before autumn (Clarke *et al.* 2010). The success of an Australian sugarcane-cropping season depends on rainfall and the ability to forecast it. Authors tried to determine if weather conditions in the second half of the harvest season can be predicted early in the year. This method of prediction is probabilistic, and above or below the median values were computed.

Artificial neural networks (ANNs) are massively parallel distributed processing systems representing a new computational technology built on the analogy to the human information processing system (Shukla *et al.* 2011). ANNs are computer algorithms that aim to recognise non-linear relationships between inputs. It receives input values and establishes a pattern between these values to release forecasts. ANN mimics the biological neurons inside the human brain. It consists of a number of simple and highly interconnected processing elements called neurons that process information by their dynamic state response to external inputs (Baawain *et al.* 2005). Neurons have additional parameters (weight, learning rate, momentum, etc.). An ANN has a great capability to learn by making proper adjustments of these parameters, in order to produce the desired output (Singh & Borah 2013). The basic structure of a neural network is formed of: input layer, hidden layer and output layer. Each layer consists of a set of neurons. Neurons are linked to the next layer, where each link has a weight that determines the strength and the relationship between the two connected neurons. Output of the neuron is multiplied by this weight while it is transferred to the neuron of the next layer. Neurons not belonging to the input layer may receive multiple inputs, which would be calculated based on a transfer function to produce output. The dataset is usually subdivided into three different sets: training set, validation set and testing set. ANN is trained using the learning process (Zou *et al.* 2008). During this process, inputs are given to the model, outputs are calculated and weights are modified to desired values (French *et al.* 1992). Usually, the weight of the interconnected nodes remains the same after finishing the training process. Indeed, to obtain successful training, a large number of known inputs and outputs are required to be added to the neural network (French *et al.* 1992). The second set is the validation set, which is used to validate the reliability of the trained network (Abbot & Marohasy 2012). The third set is called the test set and it is used to measure the accuracy of the developed model. It is basically a hind cast of data. If the performance during the testing is adequate, then the model can be used in predicting future data (Shukla *et al.* 2011).

ANN methods and approaches have been used across the world to forecast rainfall and weather variability (Haupt *et al.* 2009). Rani *et al.* have used an ANN model to forecast monthly rainfall for Andhra Pradesh in India (Rani *et al.* 2014). A teaching learning based optimization (TLBO) neural network that simulates teaching-learning phase in life was used in this study. A modification was applied to the learning phase of this model, in an attempt to enhance results. Monthly rainfall historical data were collected from the Indian Institute of Tropical Meteorology (IITM), where the dataset was composed of 1,692 monthly observations, representing the duration between 1871 and 2011. Three different samples were used: training sample, testing sample and hold out sample. A back propagation algorithm was developed for comparison with this model, and results showed that modified TLBO had better results in terms of root mean square error (RMSE) (Rani *et al.* 2014). Maqsood *et al.* used ANNs to forecast hourly rainfall for southern Saskatchewan in Canada (Maqsood *et al.* 2003). Temperature, wind speed and relative humidity were used as input for the neural network models applied in the study.

Neural networks have not been widely explored for forecasting Australian rainfall (Abbot & Marohasy 2013). Abbot and Marohasy predicted rainfall for the Brisbane catchment using ANNs (Abbot & Marohasy 2013). Historical rainfall and temperature datasets in addition to climate indices have been used as input in their study. Neurosolutions6 for Excel was used as the ANN software, where 85% of data were used for training and the remaining for testing. Output was assigned to 1, 2 and 3 months respectively. Results were compared to POAMA in two locations (Gatton and Harristville), having lower RMSE and WNDI for the ANN approach. Authors showed that climate indices and temperature are enhancing components for rainfall forecasting using ANN.

Abbot and Marohasy used ANN again to forecast rainfall for the Brisbane catchment in Australia for 12 month lead-time. Better forecasts have been recorded with this mechanism where correlation exceeded 0.85 for most of the tested months (Abbot & Marohasy 2015a).

He *et al.* proposed an approach using historical rainfall data and climate indices to develop a forecasting model for south Australia that is based on the integration of multi-resolution analysis and multi-linear regression model (He *et al.* 2014). Results were compared to a traditional multi-linear regression model where lower relative absolute error for most of the stations studied has been recorded with their mechanism.

In our study, we developed a rainfall forecasting mechanism that uses ANNs and climate attributes to forecast both seasonal (Lead 3) and annual (Lead 12) forecasts. This paper consists of six sections. The following section describes the dataset that has been used to perform the study. The subsequent two sections present the used methodology, and the results and the comparative analysis. The final section derives a conclusion and future directions.

DATA

Rainfall and temperature

Sugarcane production occurs along 2,100 km of coastline between Mossman in north Queensland and Grafton in northern New South Wales. For this study, monthly rainfall data were collected from the BOM for each of the 8 cities: Bingera, Mossman, Macknade, Plane Creek, Maryborough, Victoria Mill, Fairymead and Kalamia. Those sites were chosen to perform this study for their closeness to sugarcane paddocks and mills, in addition to availability of long historical records in each of these locations. Missing data were replaced by values found in the nearest weather station. Long historical datasets for more than 100 years were composed. Planecreek historical monthly rainfall dataset commenced in January 1909. Macknade, Maryborough, Mossman and Victoria plantation began in 1908. Bingera and Fairymead started in January 1900, while Kalamia holds the largest dataset dating back to the late 19th century, August 1892. Annual average rainfall for each of the following locations is represented in Table 1. Maximum and minimum monthly temperatures were collected from BOM for each site.

Climate indices

Multiple studies have been applied showing evidence that climate indices are potential predictors of seasonal and annual rainfall (Pasini & Langone 2010; Schepen *et al.* 2012). The ENSO is an important climate phenomenon that affects primarily the atmospheric conditions of the tropical Pacific region including the local climate and weather in Australia (Tularam 2010). The Southern Oscillation Index (SOI) and Nino3.0 are the two most widely used indicators to represent ENSO (Baawain *et al.* 2005). SOI is the monthly averaged pressure difference between Darwin and Tahiti (Baawain *et al.* 2005), while the Nino3.0 index represents the sea surface

Table 1 | Annual average rainfall for each of the eight sites used in the study

City	Annual average (mm)
Bingera	1,023
Kalamia	1,076
Fairymead	1,094
Maryborough	1,141
Plane Creek	1,764
Victoria Plantation	2,032
Macknade	2,151
Mossman	2,385

temperature (SST) anomalous averaged over the region bounded by 5 °N to 5 °S and 90 °W to 15 °W (Baawain *et al.* 2005). Climate anomalies are expected to be an effect of ENSO, which may lead to several consequences such as high crop loss. ENSO affects multiple countries of the world and influences agricultural activities in these countries (Schepen *et al.* 2012). According to Fawcett, ENSO should be taken into account when developing forecasting systems for Australia (Fawcett & Stone 2010). It has been added as an input to our ANN model. Data for the SOI were taken from the BOM, while Nino3.0 values were collected from the Royal Netherlands Meteorological Institute Climate Explorer, a web application that analyses climate data statistically.

Nino3.4 is the average SST anomaly in the region bounded by 5 °N to 5 °S, from 170 °W to 120 °W. Small temperature deviation in the Nino3.4 region significantly increases or decreases the chance of rainfall in Australia. This temperature deviation is frequently referred to as the value of the Nino3.4 index (Everingham *et al.* 2007). Nino3.4 values and additional Nino values (Nino4.0, Nino1.2) that can be used as a measurement for the ENSO phenomena were collected from KNMI climate explorer and added to the dataset.

Indian Ocean Dipole (IOD) is a climate index that recent studies have shown to contribute to rainfall variability across Australia (Risbey *et al.* 2009). It is an ENSO-like coupled ocean–atmosphere phenomenon in the equatorial Indian Ocean. The Dipole Mode Index (DMI) is a measure of the IOD defined as the difference in SST between the tropical western Indian Ocean and the tropical south-eastern Indian Ocean. DMI based in HadISST1 was used in this study (Abbot & Marohasy 2012). This climate index was collected as daily values from KNMI climate explorer, and converted to monthly representations for each month.

The inter-decadal Pacific Oscillation (IPO) measures pressure and temperature changes in the Pacific Ocean (Abbot & Marohasy 2015a). IPO records were added to datasets. IPO data were collected from the UK Met Office as monthly values.

Sunspots

The sun is the main driver of the Earth's climate (Steinhilber *et al.* 2009). Sunspot records were used in this study. Monthly values were collected from KNMI climate explorer.

PROPOSED METHOD

Research on seasonal climate forecasts for the Australian sugar industry is steadily on the rise (Jaffres & Everingham 2005). Twelve lagged values were created and added for each of the historical monthly rainfall values, maximum and minimum monthly temperature values, climate indices and sunspots. Data were manipulated so that records begin at the same time as monthly rainfall data values. For each of the eight cities, two types of rainfall forecasting were targeted: seasonal and annual. The first type is to forecast seasonally for the next 3 months (Lead 3), while the second is to forecast rainfall up to a year. Neurosolutions infinity (NeuroDimensions Inc., Florida) was the neural network software used to perform this study. For each type of forecasts, two approaches to rainfall forecasting as described below have been proposed. Figure 1 represents the two proposed approaches.

All months optimization

The first method uses all the historical data as one dataset and input this set into the neural network. For each of the eight sites, data for all the months were combined and added as an input to the neural network; 75% were used for training, 15% for validation and the remaining 10% were used for testing. For each site, two runs were applied: 3 month lead-time, and 12 month lead-time. The 3 month lead-time is to forecast three values for the next coming 3 months, while 12 month lead-time is to forecast up to 1 year. As a result, 16 runs have been applied for the eight sites. Testing data contained at least 10 years of data.

Single month optimization

A single month optimization is a technique for inputting data into the neural network, where instead of adding all the monthly records as a single set, each month's data (rainfall, temperature, sunspots and climate indices) are added to the neural network as a dataset. The 75% of data were used for training, 15% data were used for validation and 10% data were used for testing. For each of the eight sites, 24 runs have been conducted: 12 to forecast seasonally (Lead 3) and the remaining to forecast annually (Lead 12). Lead 3 means

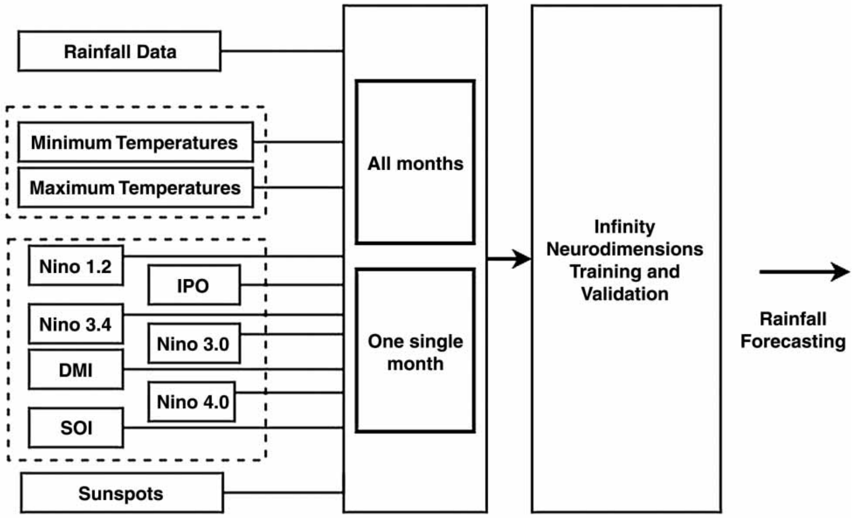

Figure 1 | Two approaches followed, all months and one single month optimizations.

each month forecasts the next third month, e.g. June data forecast September, while Lead 12 is to forecast the amount of rainfall for the same month in the next year, e.g. June values predict one single value of June for next year. Using this approach, 192 runs were conducted. At running single month optimization, 10 records representing the latest 10 years of monthly values were used for testing.

DISCUSSION AND ANALYSIS

To compare the two approaches, testing data were collected from all the runs. RMSE, mean absolute error (MAE) and correlation coefficient (r) are statistical techniques used to compare results of forecasts. These statistical techniques have been deployed in water forecasting (Shukla *et al.* 2011). RMSE measures the difference between the forecast rainfall and observed rainfall quantitatively (Abbot & Marohasy 2015b). RMSE values are positive, where the closer the number is to zero, the more skilful the forecasts are. To achieve precise comparison, a deconstruction of data was performed. Deconstruction is used to divide the 10% of testing data in the first approach (all month optimization) into 12 datasets where each set contains the values of a specific month across all the years used in testing. MEA, RMSE, and *r* values were calculated for each month for two sites, Bingera (Table 2) and Mossman (Table 3) using the two approaches.

Negative correlations were obtained for several months in the all months optimization for annual forecasts (Lead 12). Using a single month optimization in both seasonal and annual forecasts revealed high correlations ($r > 0.8$) for all the months in both locations. A correlation average of 0.96 was reported for both types (Lead 3 and Lead 12) in Bingera using one single month optimization, in addition to Mossman Lead 3. A slightly lower average correlation was found in Mossman Lead 12 using one single month optimization. Lower correlation averages were gained using the all months optimization approach in both seasonal and annual forecasts in Bingera and Mossman. Mossman obtained an average RMSE of 69.71 mms and 58.91 mms for Lead 3 and Lead 12 using a single month optimization. Higher RMSE values were acquired using the all months approach for the two types respectively in most of the months. The same applied for Bingera, where lower RMSEs were obtained using a single month approach. The all months optimization revealed lower RMSE in 2 months in December Bingera Lead 3 and Mossman January Lead 3. Tables clearly verify that a single month approach reveals better forecasts than all months approach in both locations.

For all months optimization, 10% of the data represents the 10 years between October 2004 and September 2014 (before deconstruction). For a single month optimization, 10% of the data represents 1-month values between October 2004 and September 2014. To demonstrate better representation, single

Table 2 | MAE, RMSE and *r* values for Bingera monthly rainfall forecasts between October 2004 and September 2014, seasonal and annual forecasts using all months optimization and one single month optimization

Type	Seasonal (Lead 3)						Annual (Lead 12)					
Approach	All-months			One-month			All-months			One-month		
Month	MAE (mm)	RMSE (mm)	r	MAE (mm)	RMSE (mm)	r	MAE (mm)	RMSE (mm)	r	MAE (mm)	RMSE (mm)	r
January	75.43	94.18	0.883	27.37	38.67	0.980	81.60	127.64	0.886	17.39	22.57	0.995
February	66.67	83.26	0.806	20.36	27.47	0.977	88.05	109.55	0.508	35.65	44.40	0.961
March	74.22	85.49	0.771	37.65	56.88	0.925	65.95	83.23	0.824	43.23	49.16	0.959
April	31.96	41.73	0.113	7.69	9.31	0.977	27.96	41.11	− 0.077	16.46	23.07	0.974
May	31.92	37.13	0.376	14.23	20.95	0.856	30.99	37.89	0.232	7.02	13.81	0.934
June	43.17	62.42	0.385	9.88	13.96	0.973	46.89	65.01	0.458	29.26	38.24	0.932
July	29.34	35.59	0.503	12.16	14.78	0.946	32.85	40.67	0.201	16.51	19.69	0.986
August	31.72	45.23	0.487	20.60	28.44	0.906	34.83	55.25	− 0.240	10.94	12.53	0.964
September	28.38	46.38	0.439	9.39	16.10	0.990	23.84	31.85	0.803	9.14	14.33	0.996
October	46.99	54.14	0.120	23.31	28.33	0.984	28.31	34.59	0.356	18.14	22.40	0.810
November	35.40	46.58	0.355	6.41	8.27	0.985	28.17	35.71	− 0.165	22.58	23.73	0.975
December	50.94	73.23	0.938	57.80	105.34	0.988	74.83	124.07	0.842	62.41	110.62	0.997
Average	45.51	58.78	0.51	20.57	30.71	0.96	47.02	65.55	0.39	24.06	32.88	0.96

Table 3 | MAE, RMSE and r values for Mossman monthly rainfall forecasts between October 2004 and September 2014, seasonal and annual forecasts using all months optimization and one single month optimization

Type	Seasonal (Lead 3)						Annual (Lead 12)					
Approach	All-months			One-month			All-months			One-month		
Month	MAE (mm)	RMSE (mm)	r	MAE (mm)	RMSE (mm)	r	MAE (mm)	RMSE (mm)	R	MAE (mm)	RMSE (mm)	r
January	153.16	198.33	0.738	196.27	212.38	0.931	117.22	158.22	0.691	76.86	99.28	0.925
February	182.64	221.22	0.884	116.42	174.66	0.878	199.88	279.86	0.522	202.63	241.85	0.904
March	169.72	193.28	0.670	72.06	80.21	0.963	123.37	154.09	0.798	76.26	93.29	0.935
April	174.53	245.61	− 0.189	39.63	63.06	0.963	99.64	120.61	0.841	57.52	88.35	0.984
May	65.56	118.57	− 0.042	35.56	59.77	0.986	75.14	125.18	− 0.164	14.17	21.33	0.998
June	43.93	53.71	0.016	21.78	34.82	0.859	48.88	63.06	− 0.173	11.56	21.16	0.926
July	21.92	25.77	0.041	9.62	11.67	0.962	29.85	35.83	− 0.076	3.39	4.49	0.989
August	29.90	32.07	− 0.058	9.35	12.05	0.963	19.72	26.30	0.605	1.70	2.10	0.998
September	32.42	48.95	0.371	8.98	11.61	0.982	29.29	49.85	0.317	13.05	15.65	0.965
October	54.12	82.50	0.180	15.04	19.22	0.976	72.22	99.51	0.000	13.90	18.20	0.992
November	100.53	123.78	− 0.309	52.35	63.23	0.965	84.78	118.53	0.099	39.67	52.48	0.969
December	91.40	107.58	0.583	76.06	93.78	0.867	72.65	93.42	0.706	44.61	48.76	0.958
Average	93.32	120.95	0.24	54.43	69.71	0.94	81.05	110.37	0.35	46.28	58.91	0.96

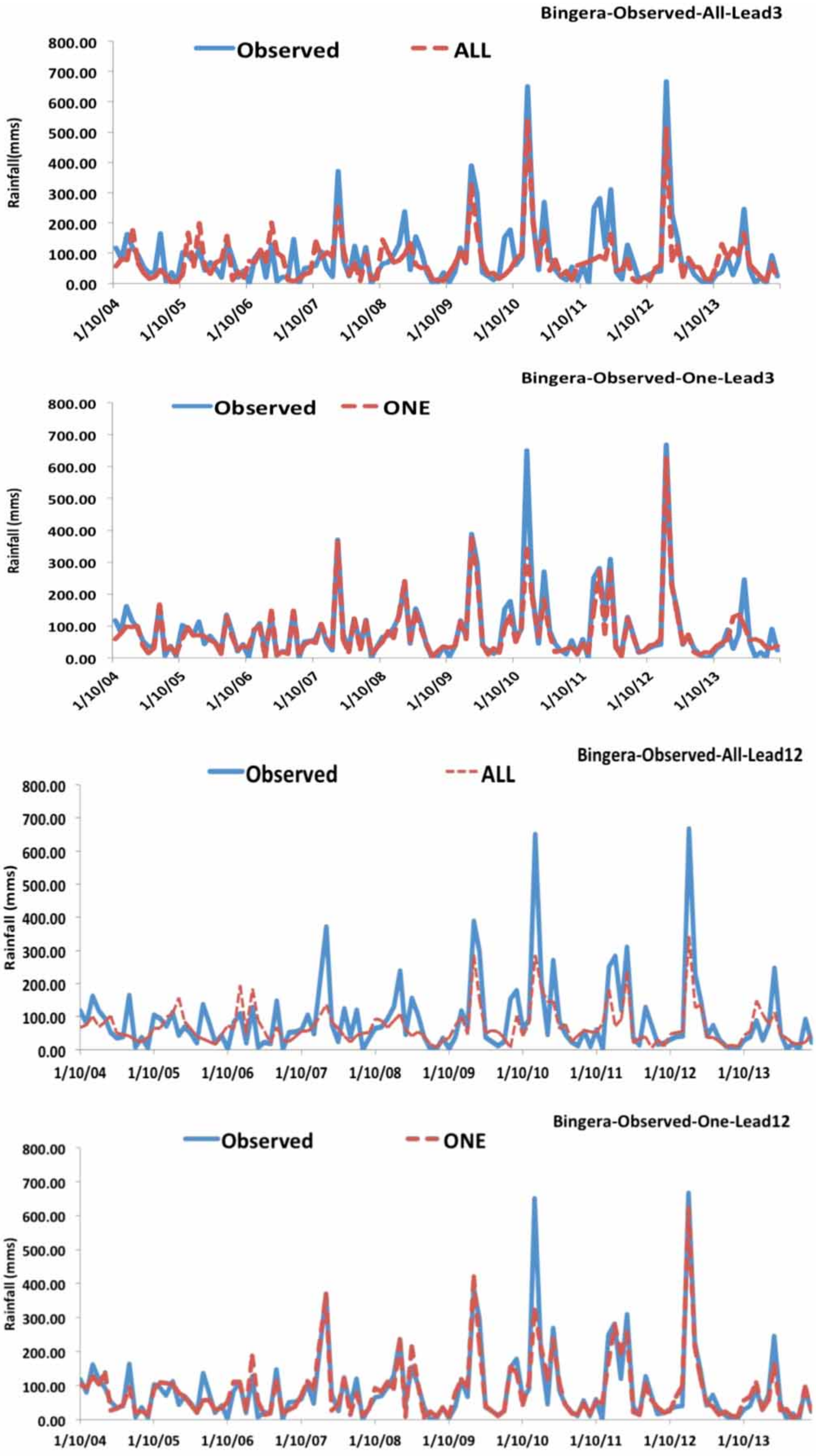

Figure 2 | Bingera observed and forecasted rainfall representation: (a) all months optimization Lead 3, (b) single month optimization Lead 3, (c) all months optimization Lead 12, (d) single month optimization Lead 12.

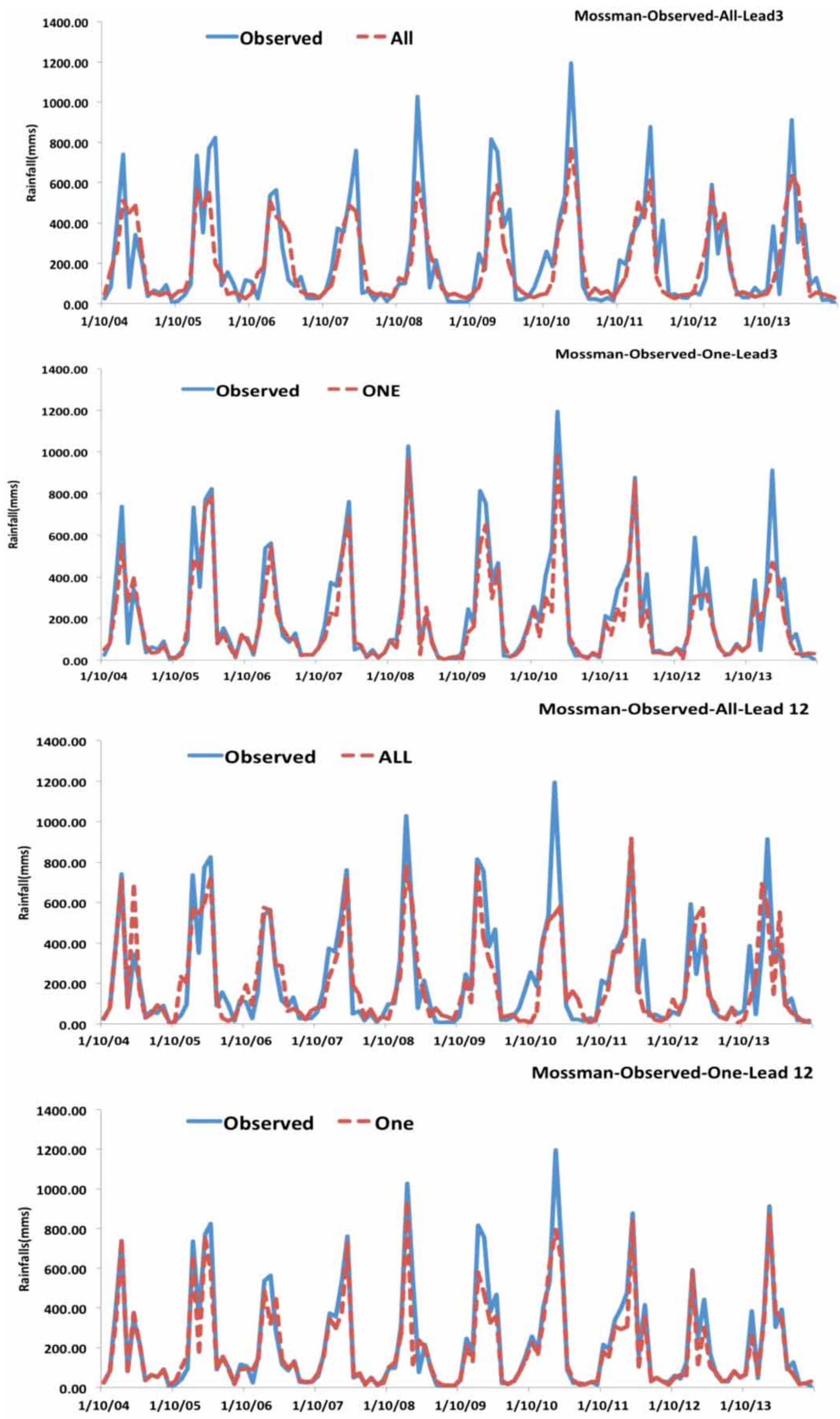

Figure 3 | Mossman observed and forecasted rainfall representation: (a) all months optimization Lead 3, (b) single month optimization Lead 3, (c) all months optimization Lead 12, (d) single month optimization Lead 12.

month testing data results were combined in each of the two locations Bingera and Mossman. Ten records from each run (12 runs in total for each approach and each type) were merged. The formed dataset represents 10 years of data (October 2004 to September 2014). Figures 2 and 3 show testing data results for two locations Bingera and Mossman from each approach on behalf of two types of forecasts, seasonal (Lead 3) and annual (Lead 12). Using two approaches, ANN had the ability to capture patterns in rainfall, with a difference in terms of accuracy and amount of rainfall. All months optimization was capable of forecasting seasonally December 2010 rainfall, where all months optimization Lead 12, single month Lead 3 and single month Lead 12 failed to predict this variation in terms of rainfall. Although better results were shown with single month in most of the locations as in December 2010, this does not mean that we can only rely on single month. Single month optimization predicted correctly rainfall for Mossman January 2014 with Lead 12, while single month optimization was not able to perform with the same level of accuracy in seasonal forecasts (Lead 3).

Ideal point error (IPE), which is calculated by identifying the ideal point to a multi-dimensional space that each model should be evaluated against (Malamos & Koutsoyiannis 2016), was used to inter-compare the performance of the two approaches. IPE is formed by normalising the statistical measurements so that it ranges between 0 and 1, where zero is the best and 1 is the worst (Domínguez *et al.* 2011). In this study, the IPE is calculated by up to three statistical measurements MAE, RMSE and determination of coefficient (R^2). R^2 is simple the square of correlation of coefficient presented in the previous Tables 2 and 3. The IPE formula is represented in the following:

$$IPE3 = \left[0.33\left(\left(\frac{RMSE_i}{\max RMSE}\right)^2 + \left(\frac{R_i^2 - 1}{\min R^2 - 1}\right)^2 + \left(\frac{MAE_i}{\max MAE}\right)^2\right)\right]^{\frac{1}{2}}$$

where i represents a case for each month in each specific approach. The following Table 4 summarizes the IPE values for both Bingera and Mossman. The results show that 1 month optimization revealed lower values in most of the cases leading to higher accuracy.

There is a wide range of approaches used to look at the aptitude of precipitation conjectures (Abbot & Marohasy 2014). The skill score of a model is calculated to get an indication about the performance of forecasts for future events. In an attempt to assess the accuracy of forecasts, skill score

Table 4 | IPE for Bingera and Mossman for the two forecasting approaches

	Bingera				**Mossman**			
Type	**Seasonal (Lead 3)**		**Annual (Lead 12)**		**Seasonal (Lead 3)**		**Annual (Lead 12)**	
Month	**All-months**	**One-month**	**All-months**	**One-month**	**All-months**	**One-month**	**All-months**	**One-month**
January	0.822	0.355	0.793	0.199	0.718	0.863	0.556	0.553
February	0.746	0.269	0.870	0.421	0.783	0.799	0.914	0.995
March	0.805	0.576	0.600	0.492	0.768	0.441	0.519	0.501
April	0.674	0.133	0.631	0.211	0.969	0.370	0.414	0.284
May	0.600	0.602	0.607	0.235	0.670	0.328	0.652	0.066
June	0.706	0.170	0.623	0.399	0.604	0.623	0.589	0.449
July	0.535	0.268	0.622	0.189	0.581	0.214	0.582	0.069
August	0.576	0.462	0.640	0.169	0.585	0.212	0.372	0.013
September	0.589	0.135	0.295	0.113	0.519	0.158	0.533	0.224
October	0.752	0.286	0.559	0.610	0.613	0.206	0.644	0.077
November	0.642	0.101	0.613	0.256	0.674	0.381	0.665	0.254
December	0.596	0.814	0.761	0.812	0.539	0.702	0.404	0.309
Average	0.670	0.348	0.634	0.342	0.668	0.441	0.570	0.316

Table 5 | Skill score results for the eight cities used in this study, all months optimization seasonal forecasts

	City							
Month	**Bingera**	**Fairymead**	**Kalamia**	**Macknade**	**Maryborough**	**Mossman**	**Plane Creek**	**Victoria Plantation**
January	49.2	70.1	7.6	56.9	52.2	22.6	−47.8	49.4
February	32.1	51.0	39.9	33.2	35.8	30.7	11.5	59.9
March	25.6	43.7	18.1	25.6	13.7	23.2	31.4	19.0
April	−2.0	23.3	−5.8	35.2	11.7	−6.9	−38.5	−13.2
May	10.9	1.2	−45.3	−29.0	−13.4	−8.1	−11.0	−18.4
June	−2.8	2.9	−31.3	−14.3	−13.5	−2.5	−3.4	12.5
July	14.9	−1.6	−40.3	−30.3	12.0	−6.5	−64.3	9.0
August	4.2	−9.7	10.0	−36.5	−4.1	−5.7	−49.0	−1.1
September	4.6	53.0	0.9	12.9	−19.3	2.1	36.4	−1.7
October	−40.8	−8.6	−22.8	−3.2	−16.9	−1.0	−7.5	−4.9
November	−26.2	−6.1	6.6	−4.1	−24.3	−14.3	12.8	6.8
December	58.9	66.8	−1.8	38.0	51.6	18.5	22.2	−46.3

Table 6 | Skill score results for the eight cities used in this study, one single month optimization seasonal forecasts

	Area							
Month	**Bingera**	**Fairymead**	**Kalamia**	**Macknade**	**Maryborough**	**Mossman**	**Plane Creek**	**Victoria Plantation**
January	79.1	18.0	27.8	59.6	56.0	17.14	33.9	27.7
February	77.6	69.5	55.0	53.0	33.4	45.3	34.1	48.4
March	50.5	68.9	71.0	50.9	20.9	68.2	41.7	78.7
April	77.2	40.3	57.9	70.3	35.5	72.5	44.8	38.2
May	49.7	28.1	34.5	72.9	57.4	45.5	81.5	39.8
June	77.0	85.5	41.2	70.0	74.0	33.5	57.3	77.7
July	64.7	42.6	81.8	66.3	52.1	51.8	67.6	38.4
August	39.8	69.5	72.4	49.6	15.0	60.3	50.4	56.7
September	66.9	69.8	62.6	72.4	69.2	76.8	52.2	24.6
October	26.3	60.3	31.9	75.3	51.2	76.5	47.2	83.7
November	77.6	72.5	46.3	88.7	55.2	41.6	61.1	38.1
December	40.9	60.0	78.6	71.6	47.5	29.0	50.2	52.8

has been calculated for each month of the eight different sites. Skill score is a technique that is used to demonstrate the accuracy of a specific method in prediction. Skill score assesses the accuracy of a model based on a reference model. Our model was compared and evaluated against climatology. Climatology is a technique for forecasting rainfall based on the average of months, e.g. rainfall forecast for a specific month is the average of the rainfall values collected from the previous same month back to a specific time. Formula of skill score:

$$SS = \frac{(RMSEref - RMSE)}{RMSEref} * 100$$

where *RMSEref* is the *RMSE* value calculated for the reference forecasts (climatology), and RMSE is related to the

Table 7 | Skill score results for the eight cities used in this study, all months optimization annual forecasts

	Area							
Month	**Bingera**	**Fairymead**	**Kalamia**	**Macknade**	**Maryborough**	**Mossman**	**Plane creek**	**Victoria Plantation**
January	31.2	17.5	53.8	44.9	48.4	36.4	−15.2	12.1
February	10.7	17.7	62.7	18.3	43.3	10.1	43.1	14.7
March	27.6	12.1	14.6	4.1	60.7	38.1	18.7	39.3
April	−0.5	10.7	−4.8	−0.3	−0.2	46.7	−4.6	16.5
May	9.0	5.2	−39.7	−25.7	9.5	−14.9	−24.9	−24.1
June	−7.1	−21.6	−2.2	−4.0	−25.1	−21.0	−9.8	−16.8
July	2.8	16.7	−19.1	2.7	−4.4	−48.2	−11.0	−2.0
August	−17.0	−5.1	−1.9	29.6	1.5	13.1	−0.6	−27.5
September	34.5	−12.8	−8.5	11.6	4.6	−0.2	−3.2	−3.8
October	10.0	30.2	−49.4	−8.9	12.0	−25.8	16.4	−10.4
November	3.3	−1.4	−8.2	60.3	1.2	−12.2	16.2	1.8
December	30.4	18.6	45.8	65.8	37.4	22.5	−5.3	2.5

model to be assessed (ANNs model). Skill Score can be either positive or negative. A 0 skill-score means that the tested model has no skill in prediction; 100 means perfect forecasts. Skill score values were generated for each approach. Climatology forecasts were collected from BOM. Table 5 represents skill scores for seasonal forecasts with the all months optimization approach, Table 6 represents skill scores for seasonal forecasts with single month optimization, Table 7 represents annual forecasts with all months optimization, and Table 8 represents annual forecasts skill score results for annual forecasts with single month optimization.

Our results confirm that ANNs can be used to perform both seasonal and annual forecasts. In their paper, Hawthorne *et al.* (Hawthorne *et al.* 2013) compared their model that is based on GCM to climatology, where skill

Table 8 | Skill score results for the eight cities used in this study, one single month optimization annual forecasts

	Area							
Month	**Bingera**	**Fairymead**	**Kalamia**	**Macknade**	**Maryborough**	**Mossman**	**Plane creek**	**Victoria Plantation**
January	87.8	69.6	34.1	42.4	72.0	60.1	42.8	80.5
February	63.8	91.7	62.1	38.8	46.6	22.3	34.7	87.4
March	57.2	40.5	86.4	46.0	77.4	62.5	31.2	79.0
April	43.6	57.9	72.3	60.8	46.6	61.0	73.2	71.4
May	66.8	48.5	56.0	86.9	74.8	80.4	57.5	57.5
June	37.1	30.8	67.4	39.1	64.0	59.4	42.7	59.1
July	53.0	53.1	51.7	33.9	34.6	81.4	53.8	48.2
August	73.5	47.7	69.0	38.0	23.9	93.1	56.9	74.3
September	70.5	69.7	47.2	33.8	40.0	68.5	29.3	47.8
October	41.7	61.9	67.0	62.0	85.2	77.4	92.4	59.1
November	35.7	60.6	67.8	56.7	53.9	51.0	75.6	69.7
December	38.0	40.4	58.6	69.7	55.6	62.8	38.7	74.3

scores varied between $-20 \rightarrow +20\%$ (Hawthorne *et al.* 2013). Using a single month optimization revealed higher skill scores when compared to climatology. Skill score values for all months optimization varied between -49 and 65% for both seasonal and annual forecasts. Single month optimization showed a skill score range between 15.0 and 93.1%. Only three values were obtained with a skill score less than 20% (Fairymead January Lead 3, Mossman January Lead 3, and Maryborough August Lead 3). This does not mean that forecasts are always better with a single month optimization. For some of the months, higher skill scores were obtained with all months optimization as in Fairymead Lead 3 forecasts (70.1%). The average skill score for each month in the eight locations using a single month optimization to forecast seasonally (annually) was reported as 55.1 (58.1). Based on those results, growers would be able to trust seasonal and annual forecasts following this approach.

CONCLUSION

In this paper, an ANN for predicting rainfall for sugarcane regions has been investigated. Two approaches have been proposed. The first approach was to forecast using all the datasets and the second was to forecast using a single optimization. Single month optimization delivered accurate forecasts for the two forecasting types being targeted: 3 months and 12 months, and those results correlates with (Abbot & Marohasy 2015a). The results were compared to climatology in terms of skill scores. The analysis showed 39.9%–68% better accuracy than climatology using a single month optimization. Growers can rely on this type of forecasts for water management and decision making related to the sugarcane industry. These results are promising and further explorations are to be deployed. Studies revealed that the variation in the Australian rainfall is linked to several climate indices such as ENSO, IOS, PDO, SAM and STR but its usefulness in rainfall forecasting is still uncertain (He *et al.* 2014). ENSO has been investigated in this model, but new climate indices can be added and explored in future studies. Input data can be investigated and multiple approaches for missing data replacement is to be studied. In addition, new ANN models are to be developed and compared in terms of accuracy.

REFERENCES

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

Abbot, J. & Marohasy, J. 2013 The application of artificial intelligence for monthly rainfall forecasting in the Brisbane Catchment, Queensland, Australia. *River Basin Management VII* (C. A. Brebbia, ed.). WIT Press, Southampton, pp. 125–135.

Abbot, J. & Marohasy, J. 2014 Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmospheric Research* **138**, 166–178. doi: 10.1016/j.atmosres.2013.11.002.

Abbot, J. & Marohasy, J. 2015a Improving monthly rainfall forecasts using artificial neural networks and single-month optimisation: a case study of the Brisbane catchment, Queensland, Australia. *WIT Transactions on Ecology and the Environment* **196**, 3–13.

Abbot, J. & Marohasy, J. 2015b Using artificial intelligence to forecast monthly rainfall under present and future climates for the Bowen Basin. *International Journal of Sustainable Development and Planning* **10** (1), 66–75.

Baawain, M. S., Nour, M. H., El-Din, A. G. & El-Din, M. G. 2005 El Nino southern-oscillation prediction using southern oscillation index and Nino3 as onset indicators: application of artificial neural networks. *Journal of Environmental Engineering and Science* **4** (2), 113–113. doi: 10.1139/s04-047.

Clarke, A. J., Gorder, S. V. & Everingham, Y. 2010 Forecasting long-lead rainfall probability with application to Australia's northeastern coast. *Journal of Applied Meteorology and Climatology* **49** (7), 1443. doi: 10.1175/2010JAMC2373.1.

Cottrill, A., Hendon, H. H., Lim, E.-P., Langford, S., Kuleshov, Y., Charles, A. & Jones, D. 2012 *Seasonal Climate Prediction in the Pacific using the POAMA Coupled Model Forecast System*. National Library of Australia Cataloguing-in-Publication Entry.

Domínguez, E., Dawson, C. W., Ramírez, A. & Abrahart, R. J. 2011 The search for orthogonal hydrological modelling metrics: a case study of 20 monitoring stations in Colombia. *Journal of Hydroinformatics* **13** (3), 429. doi: 10.2166/hydro.2010.116.

Everingham, Y. L., Muchow, R. C., Stone, R. C. & Coomans, D. H. 2003 Using southern oscillation index phases to forecast sugarcane yields: a case study for Northeastern Australia. *International Journal of Climatology* **23** (10), 1211–1218. doi: 10.1002/joc.920.

Everingham, Y. L., Clarke, A. J., Chen, C. C. M., Gorder, S. V. & McGuire, P. 2007 *Exploring the Capabilities of a Long Lead Climate Forecasting System for the NSW Sugar Industry*. Australian Society of Sugar Cane Technologists, Mackay, Australia.

Everingham, Y. L., Stoeckl, N. E., Cusack, J. & Osborne, J. A. 2012 Quantifying the benefits of a long-lead ENSO prediction model to enhance harvest management–A case study for the Herbert sugarcane growing region, Australia. *International Journal of Climatology* **32** (7), 1069–1076. doi: 10.1002/joc.2333.

Fawcett, R. J. B. & Stone, R. C. 2010 A comparison of two seasonal rainfall forecasting systems for Australia. *Australian Meteorological and Oceanographic Journal* **60** (1), 15–23.

French, M. N., Krajewski, W. F. & Cuykendall, R. R. 1992 Rainfall forecasting in space and time using a neural network. *Journal of Hydrology* **137** (1), 1–31. http://dx.doi.org/10.1016/0022-1694(92)90046-X.

Haupt, S. E., Pasini, A. & Marzban, C. 2009 *Artificial Intelligence Methods in the Environmental Sciences*. Springer, New York, NY.

Hawthorne, S., Wang, Q. J., Schepen, A. & Robertson, D. 2013 Effective use of general circulation model outputs for forecasting monthly rainfalls to long lead times. *Water Resources Research* **49** (9), 5427–5436. doi: 10.1002/wrcr.20453.

He, X., Guan, H., Zhang, X. & Simmons, C. T. 2014 A wavelet-based multiple linear regression model for forecasting monthly rainfall. *International Journal of Climatology* **34** (6), 1898–1912. doi: 10.1002/joc.3809.

Jaffres, J. & Everingham, Y. 2005 An exploratory investigation of the relationship between decadal rainfall and climate indices.

Malamos, N. & Koutsoyiannis, D. 2016 Bilinear surface smoothing for spatial interpolation with optional incorporation of an explanatory variable. Part 2: application to synthesized and rainfall data. *Hydrological Sciences Journal* **61** (3), 527–540. doi: 10.1080/02626667.2015.1080826.

Maqsood, I., Khan, M. & Abraham, A. 2003 Weather Forecasting Models Using Ensembles of Neural Networks. In: *Intelligent Systems Design and Applications*, Vol. 23 (A. Abraham, K. Franke & M. Köppen, eds). Springer, Berlin, Heidelberg, pp. 33–42.

Pasini, A. & Langone, R. 2010 Attribution of precipitation changes on a regional scale by neural network modeling: a case study. *Water* **2** (3), 321–332. doi: 10.3390/w2030321.

Rani, B. K., Srinivas, K. & Govardhan, A. 2014 Rainfall Prediction with TLBO Optimized ANN. *Journal of Scientific & Industrial Research* **73** (10), 643–647.

Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C. & Hendon, H. H. 2009 On the remote drivers of rainfall variability in Australia. *Monthly Weather Review* **137** (10), 3233–3253. doi: 10.1175/2009MWR2861.1.

Schepen, A., Wang, Q. J. & Robertson, D. 2012 Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *Journal of Climate* **25** (4), 1230.

Shukla, R. P., Tripathi, K. C., Pandey, A. C. & Das, I. M. L. 2011 Prediction of Indian summer monsoon rainfall using Niño indices: a neural network approach. *Atmospheric Research* **102** (1), 99–109. doi: 10.1016/j.atmosres.2011.06.013.

Singh, P. & Borah, B. 2013 Indian summer monsoon rainfall prediction using artificial neural network. *Stochastic Environmental Research and Risk Assessment* **27** (7), 1585–1599. doi: 10.1007/s00477-013-0695-0.

Steinhilber, F., Beer, J. & Fröhlich, C. 2009 Total solar irradiance during the Holocene. *Geophysical Research Letters* **36** (19), L19704. doi: 10.1029/2009GL040142.

Tularam, G. A. 2010 Relationship between El Niño Southern Oscillation Index and rainfall (Queensland, Australia). *International Journal of Sustainable Development and Planning* **5** (4), 378–391.

van den Honert, R. C. & McAneney, J. 2011 The 2011 Brisbane floods: causes, impacts and implications. *Water* **3** (4), 1149–1173. doi: 10.3390/w3041149.

Zou, J., Han, Y. & So, S.-S. 2008 Overview of artificial neural networks. *Methods in Molecular Biology (Clifton, N.J.)* **458**, 15.

First received 28 January 2016; accepted in revised form 31 May 2016. Available online 20 June 2016