Theor Appl Climatol
DOI 10.1007/s00704-015-1457-3

ORIGINAL PAPER

Spatiotemporal monthly rainfall forecasts for south-eastern and eastern Australia using climatic indices

Maryam Montazerolghaem[1] · **Willem Vervoort**[1] · **Budiman Minasny**[1] · **Alex McBratney**[1]

Received: 13 February 2014 / Accepted: 1 April 2015

Abstract Knowledge about future rainfall is important for agriculture management and planning in arid and semi-arid regions. Australia has complex variations in rainfall patterns in time and space, arising from the combination of the geographic structure and the dual effects of Indian and Pacific Ocean. This study aims to develop a forecasting model of spatiotemporal monthly rainfall totals using lagged climate indices and historical rainfall data from 1950–2011 for south-eastern and eastern Australia. Data were obtained from the Australian Bureau of Meteorology (BoM) from 136 high-quality weather stations. To reduce spatial complexity, climate regionalization was used to divide the stations in homogenous sub-regions based on similarity of rainfall patterns and intensity using principal component analysis (PCA) and K-means clustering. Subsequently, a fuzzy ranking algorithm (FRA) was applied to the lagged climatic predictors and monthly rainfall in each sub-region to identify the best predictors. Selected predictors by FRA were found to vary by sub-region. After these two stages of pre-processing, an artificial neural network (ANN) model was developed and optimized separately for each sub-region and the entire area. The results indicate that climate regionalization can improve a monthly spatiotemporal rainfall forecast model. The location and number of sub-regions were important for ranking predictors and modeling. This further suggests that the impact of climate variables on Australian rainfall is more variable in both time and space than indicated thus far.

✉ Maryam Montazerolghaem
maryam.montazerolghaem@sydney.edu.au

Willem Vervoort
willem.vervoort@sydney.edu.au

Budiman Minasny
budiman.minasny@sydney.edu.au

Alex McBratney
alex.mcbratney@sydney.edu.au

[1] Department of Environmental Sciences, Faculty of Agriculture and Environment, The University of Sydney, Sydney, Australia

1 Introduction

Knowledge about future rainfall can significantly benefit land, water resources, and agriculture management, as this assists with planning and management decisions (Anwar et al. 2007; Bannayan et al. 2011; Chiew et al. 2003; Sivakumar and Hansen 2007). In Australia, rainfall and climate patterns vary greatly in time and space. South-eastern and eastern Australia has distinctly different rainfall regimes and is agriculturally productive and home for more than half of Australia's population. However, topographic and geographic features of this region make developing a reliable and skilful rainfall forecasting system for this area complicated (Drosdowsky 1993; Murphy and Timbal 2008).

The first step in developing a forecast model is to investigate possible predictors. Several studies have suggested that concurrent and lagged broad-scale climate indices are important predictors for Australian rainfall and streamflows; for example, El Niño Southern Oscillation (ENSO) (Cai et al. 2001; Drosdowsky 1993; Drosdowsky and Chambers 2001; McBride and Nicholls 1983; Nicholls 1983; Piechota et al. 1998; Power et al. 1999; Wang and Hendon 2007), Southern Oscillation Index (SOI) (McBride and Nicholls 1983; Ropelewski and Jones 1987; Stone and Auliciems 1992), Indian Ocean Dipole (IOD) (Cai et al. 2009; Saji et al. 1999; Saji and Yamagata 2003), and Southern Annular Mode (SAM) (Cai et al. 2009; Hendon et al. 2007; Meneghini et al. 2007). Particularly, over the Pacific and Indian regions, lagged climate indices have been indicated as the most useful

Published online: 02 May 2015

predictors for forecasting rainfall at seasonal and monthly time scales over Australia at different times of the year and regions of the continent (Cai et al. 2012; Cai et al. 2001; Kirono et al. 2010; Murphy and Timbal 2008; Risbey et al. 2009; Ruiz et al. 2007; Schepen et al. 2012a). Seasonal rainfall forecasts based on sea surface temperature anomalies over the Indian and Pacific Ocean have been developed by the Australian Bureau of Meteorology (BoM) (Drosdowsky and Chambers 2001; Fawcett et al. 2005).

The strength of the effect of these climate indices is variable for different areas and seasons. Given this, many researchers have investigated this across the Australian continent (Ashok et al. 2003a; Cai et al. 2012; McBride and Nicholls 1983; Power et al. 1998; Risbey et al. 2009; Wang and Hendon 2007) or separately for different parts of Australia (Kirono et al. 2010; Mekanik et al. 2012; Nicholls 2010; Piechota et al. 1998; Shi et al. 2008). In some regions and seasons, climate indices did not show strong correlation with seasonal and monthly rainfall and were weak predictors for rainfall forecasting (Kirono et al. 2010; Schepen et al. 2012a). This might explain why the BoM has moved away from statistical models and now uses a regional circulation model (POAMA) to develop forecasts. However, statistical models can be more computationally efficient, can be developed for focus areas, and can highlight specific influences.

Most of the recent studies used forecasts models with fixed climate indices for all of continental Australia or part of Australia defined by conventional boundaries (Schepen et al. 2012a; Schepen et al. 2012b; Stone et al. 1996; Wang et al. 2012). However, what has not been comprehensively addressed is whether different climate indices at different time lags influence monthly rainfall variations at a finer spatial resolution. Understanding these relationships could improve rainfall forecasts at specific locations (Risbey et al. 2009).

Most studies apply traditional statistical analysis such as linear correlation and time series analysis methods to investigate the relationship between lagged climate indices and rainfall in Australia (Schepen et al. 2012a; Verdon and Franks 2005; Wang et al. 2012). Conventional statistical methods assume linear interaction between predictors and output, and therefore, are not very well able to deal with the large number of predictors.

Artificial neural networks (ANNs) can address problems involving complex nonlinear interactions without having prior knowledge, large number of predictors, and incomplete datasets (Tasadduq et al. 2002). Application of ANNs in atmospheric science started in 1986 (Gardner and Dorling 1998; Rumelhart and Mcclelland 1986). A comprehensive review of the application of ANNs to hydrology can be found in the ASCE Task committee's report (Govindaraju 2000a; b). However, one disadvantage is that ANN tends to converge on local minima as the number of inputs increase. In order to avoid this, input space reductions should still be considered as one of the primary steps (Lin et al. 1996).

This study aims to develop a spatiotemporal monthly rainfall forecasting model for a group of weather stations in southeastern and eastern Australia (Fig. 1) by applying artificial neural networks (ANN) to lagged climate and non-climate predictors such as number of months, monthly long-term mean rainfall, spatial coordinates, and altitude. This paper, therefore, assesses the ability of ANN in conjunction of climate regionalization to provide accurate forecasts for monthly rainfall. As a start, the study focuses on the spatial variations of the rainfall at the monthly time scale and the area is divided into different sub-regions with similar monthly rainfall patterns and intensity. A fuzzy ranking algorithm (FRA) is applied to different input datasets to identify significant predictors for monthly rainfall at different spatial resolutions. Finally, a spatiotemporal model is developed with the selected input variables. The performance of models is compared with statistical parameters and against average climatology to investigate whether spatial climate regionalization can improve the performance.

2 Materials and methods

2.1 Data

2.1.1 Rainfall data

Rainfall data were obtained from the Bureau of Meteorology (BoM) from 136 high-quality weather stations from the southeastern and eastern Australia with monthly rainfall records from 1879 to 2012. The area considered in this study includes mainland Australia south of 25° S and east of 137° E, including Victoria and New South Wales (NSW), part of South Australia (SA) and Queensland and covering the Murray-Darling River basin. Stations used in this study are shown in Fig. 1. There are at least four different rainfall patterns including arid, uniform, summer, and winter dominant, as defined by BoM and based on the Koppen-Geiger climate regionalization across the study area (Kottek et al. 2006; Peel et al. 2007).

2.1.2 Climate indices

Climatic indices, used in this study, are based on past literature on seasonal and monthly rainfall forecasting over Australia (Table 1). These indices are mean sea surface temperature (SST) of defined regions over Indian and Pacific Ocean, such as Niño 3, which is the mean SST covering the area 5° S–5° N and 150°–90° W over Pacific Ocean. All of these indices have been found to have high concurrent or lagged correlations with monthly rainfall over Australia and are at a monthly time scale.

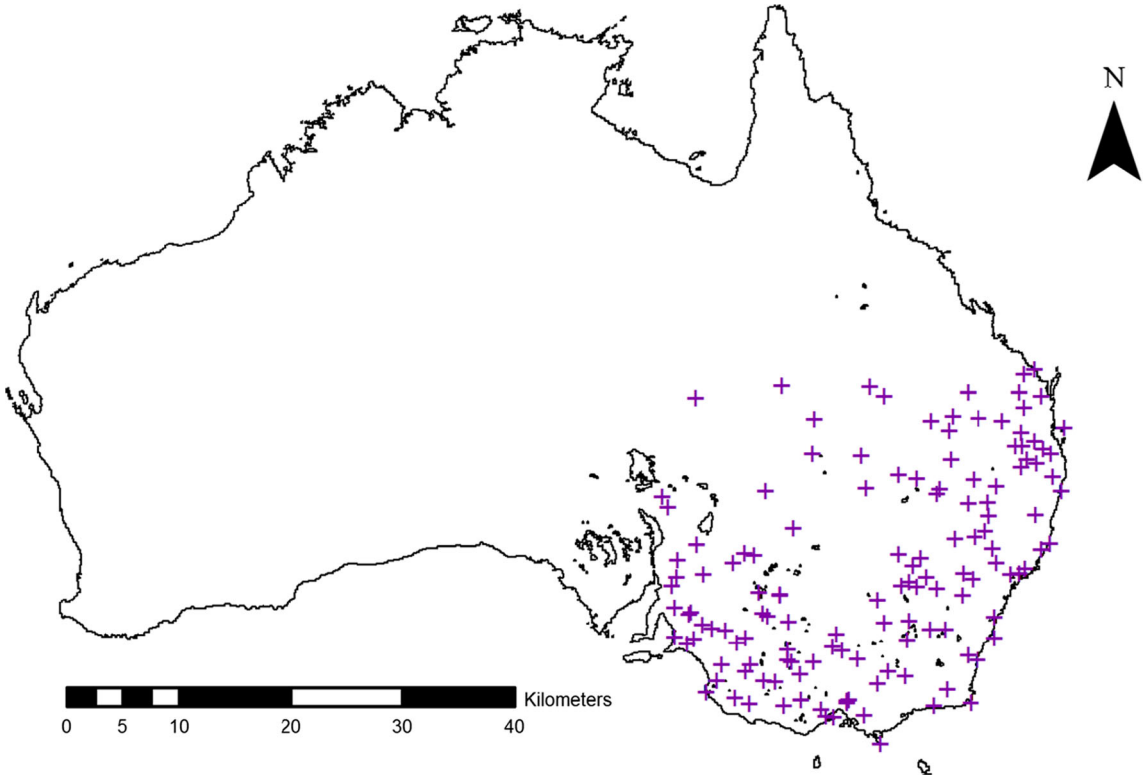

Fig. 1 Location map indicating high-quality stations used in this study

2.2 Method

2.2.1 Framework

The spatial complexity of the area was reduced by grouping together stations into sub-regions with monthly rainfall patterns. This was achieved by grouping the stations using PCA and K-means on the monthly rainfall time series to identify three to eight sub-regions. PCA reduced the dimensions of the monthly rainfall time series and removed correlation in the rainfall time series. K-means clustering was used on the PC scores describing 93 % of long-term monthly rainfall

Table 1 Commonly used atmospheric and oceanic predictors of monthly rainfall in Australia with a brief explanation of each of the predictors and related references that have used these predictors before (van Ogtrop et al. 2011)

Predictors	Definition	Period	Source
NINO 3	Mean SST from 5° S–5° N and 150°–90° W	1950–2012	NCEP—SST anomalies http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices
NINO 4	Mean SST from 5° S–5° N and 160° E–150° W	1950–2012	NCEP—SST anomalies http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices
NINO 1+2	Mean SST from 0°–10° S and 90° W–80° W	1950–2012	NCEP—SST anomalies http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices
NINO 3.4	Mean SST from 5° S–5° N and 170°–120° W	1951–2012	NCEP—SST anomalies http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices
Indian Ocean West Pole (WPI)	Mean sea surface temperature anomaly over 50° E–70° E and 10° N–10° S (Saji et al. 1999)	1854–2008	NCAR, ERSST.v3 (Smith et al. 2008)
Indian Ocean East pole (EPI)	Mean sea surface temperature anomaly over 90° E–110° E and 0° N–10° S (Saji et al. 1999)	1854–2008	NCAR, ERSST.v3 (Smith et al. 2008)
Indonesia Index (II)	Mean sea surface temperature anomaly over 120° E–130° E and 0° N–10° S (Verdon and Franks 2005)	1854–2008	NCAR, ERSST.v3 (Smith et al. 2008)
Tasman sea Index (TSI)	Sea surface temperature anomalies in Tasman sea (Murphy and Timbal 2008)	1854–2008	NCAR, ERSST.v3 (Smith et al. 2008)

variations. Subsequently, a fuzzy ranking algorithm (FRA) was applied to the lagged climatic predictors and monthly rainfall to identify the best predictors. After these two stages of pre-processing, climate regionalization, and identification of important input variables, a neural network model was developed and optimized for each of the sub-regions and the entire area. The model for the entire area is based on the combination of sub-region prediction models. The algorithms and analysis were developed using MATLAB 2012a and Arc GIS 10.2.

2.2.2 Climate regionalization

Climate regionalization is the process of grouping highly correlated stations in terms of rainfall patterns and intensity into the same sub-region. This has been a useful pre-processing method in different climate studies (Gerstengarbe et al. 1999; Lund and Li 2009; Munoz-Diaz and Rodrigo 2004; Villar et al. 2009).

The main reason for using climate regionalization for this study is threefold: (1) it reduces spatial complexity and the dimensionality of the original dataset, to somewhat more homogenous data in terms of geographical and climatological characteristics (Dezfuli et al. 2010); (2) homogenous sub-regions might be affected in a more consistent way by the climate indices than the overall region; therefore, this allows a finer resolution for selecting important predictors; (3) it improves the prediction model and increases model accuracy. Therefore, climate regionalization is the primary method for identifying significant predictors at a fine spatial resolution. However, no specific clustering method is considered optimal in the literature for climate regionalization. Therefore, method selection could be a subjective decision (Dezfuli 2011).

To compare rainfall time series of stations point by point, dealing with missing data is a major issue. In this study, the approach followed two rules: First, months with missing values more than 25 % of stations were omitted from dataset. This resulted in the elimination of data before 1915. Secondly, missing values were replaced by the mean of the four closest stations in terms of geographical distance when less than 25 % of stations have missing data.

K-means clustering is one of the methods suggested for climate regionalization (Gong and Richman 1995; Wilson et al. 1992). Using principal component analysis (PCA) in conjunction with K-mean clustering can improve results (Corte-Real et al. 1998). Therefore, PCA (Hotelling 1933; Pearson 1901) was applied to the rainfall time series to project the variability of the rainfall time series into fewer dimensions. There are several methods recommended for choosing the number of principal components that present more or less the same information as the original data (Rogers 1990; Rogers and McHugh 2002; Valle et al. 1999). PCs which account for 93 % of the total temporal variations of rainfall were selected (Johnson 1998). Therefore, dimensionality of observations was reduced from the original 1457 monthly rainfall time series to 50 variables.

Subsequently, K-means clustering was applied to the PC scores with data dimensions of 136×50 (number of stations× selected PCs). Stations with similar patterns and rainfall intensity were grouped together. Dynamic time warping (DTW) and Euclidian distance are the most common methods for measuring similarity in time series clustering and sub-region

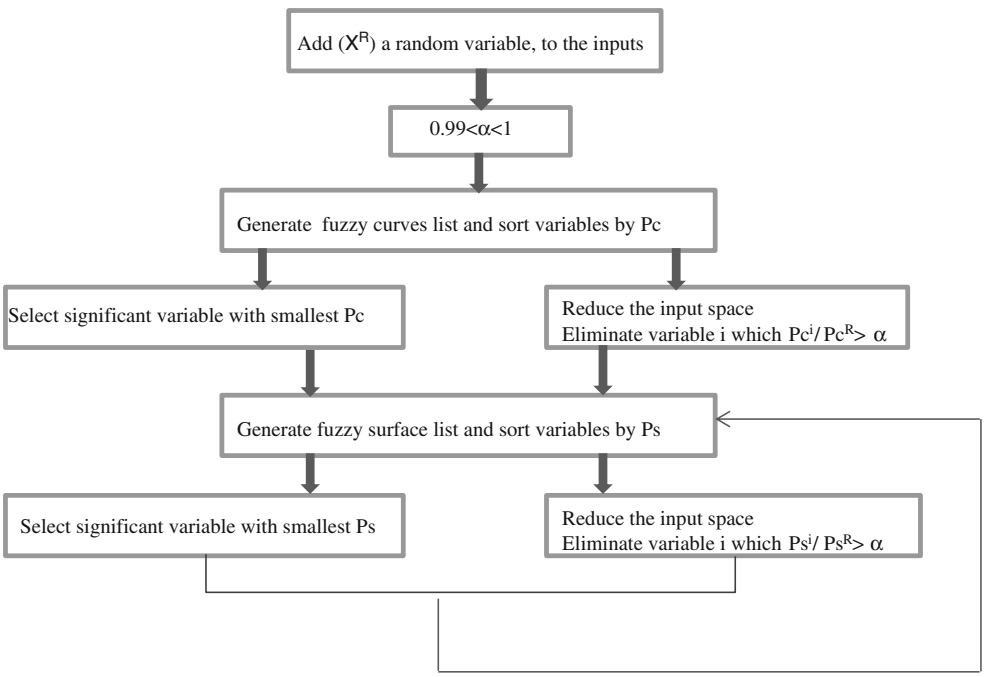

Fig. 2 Fuzzy ranking algorithm flow diagram

applications (Keogh and Kasetty 2003). Because the objective in this study is to group stations based on their real-time rainfall pattern similarity, Euclidian distance rather than DTW was used for time series distance measurements. Stations with a higher degree of rainfall similarity at the monthly scale were divided into several (three to eight) sub-regions.

2.2.3 Identification of significant predictors

Lagged climatic indices (listed in Table 1) from 1 to 6 months were considered as rainfall predictors. As a result, the dataset has hundreds of input variables and thousands of data points. Climatic indices are usually noisy and frequently cross correlated. Nonlinear modeling techniques such as neural networks and genetic algorithms can deal with such a high dimensional input space. The main issue in these methods is the convergence on local minima as the number of inputs and data points increases. Therefore, input variable selection and ranking methods that seek to identify reliable predictors were considered as a primary step to improve the performance of the model. There are several input space reduction techniques for nonlinear systems applications such as forward selection and backward elimination. This study chose a fuzzy ranking algorithm (FRA), which can identify input variables that better predict output for nonlinear methods such as genetic algorithms and ANN (Lin et al. 1996).

FRA develops two stages of fuzzy curves and surfaces between the input variables and outputs (Fig. 2). First, a random variable is added to the dataset. The performance of each input is compared with performance of the random variable as an input to make sure that there is no random relation between input and output. First and second stage fuzzy curves are developed to rank input variables, based on the idea that certain inputs could estimate the output better than other inputs. First stages of fuzzy surfaces are based on the assumption that two independent inputs together can approximate the output better than single dependent input variables. This allows correlated inputs to be eliminated. Second stage fuzzy curves and surfaces are identified based on the fact that the local estimate of the variance of the output will differ from the global variance of the output at many points. More detailed definitions and equations for first and second fuzzy curves and surface can be found in Lin et al. (1996).

First and second stages of fuzzy curves were generated for all input variables. The performance index of the fuzzy curve of input variables is calculated to rank the inputs based on their ability to predict the output variables. The performance index for fuzzy curve (Pc^i) of input X^i for output y is:

$$Pc^i = \frac{Py_c^i}{I + Pv_c^i} \tag{1}$$

where Py_c^i and Pv_c^i are the performance indices of the first stage and second stages of fuzzy curves for the ith input, respectively. The first input variable is selected by a normalized performance index of the fuzzy curves. To select the next important input variables, the fuzzy surface performance index is used:

$$Ps^{i,j} = \frac{Py_s^{i,j}}{1 + Pv_s^{i,j}} \tag{2}$$

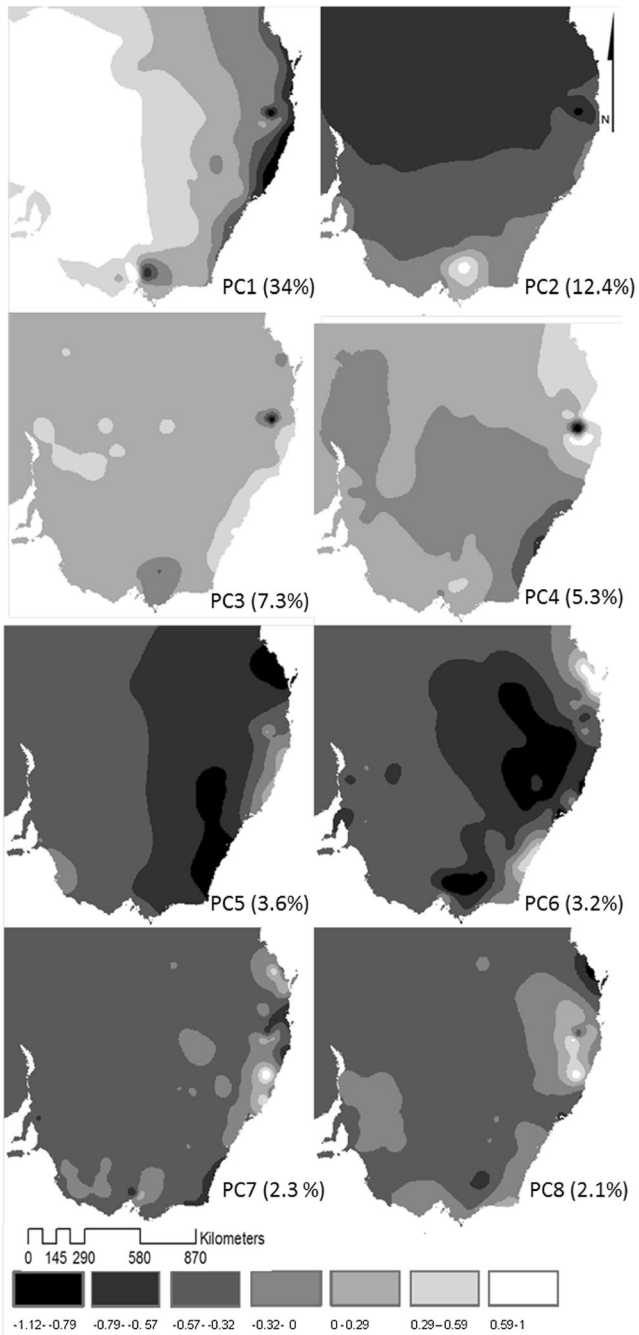

Fig. 3 Spatial variation of the first eight PCs describes together 70 % of long-term (temporal) rainfall variations

The performance index of each individual input variable is compared with the performance index of a random variable. Input variables that do not show better performance to predict output than a random variable, at any stage of the algorithm, will be eliminated from the input dataset. Where $Py_s^{i,j}$ and $Pv_s^{i,j}$ are the performance indices of the first and second stages of fuzzy curves for the ith input, respectively.

To investigate significant predictors, FRA was applied with monthly rainfall as an output variable and lagged climate indices (Table 1) from 1 to 6 months as input variables. Input variables considered in this study are SST anomalies over the Indian (WPI, EPI, II) and Pacific Ocean (NINO 3, NINO 4, NINO 1+2, NINO 3.4) and Tasman Sea indices (TSI) with 1–6-month lags. Other variables used to train the model are longitude, latitude, altitude, and number of the month.

2.2.4 Spatiotemporal rainfall forecasting model using ANN

Two different approaches were taken to develop a monthly rainfall forecast model using FRA selected climate input variables:

1. Determine an optimal neural network for the entire area, this basically means there are no sub-regions.
2. Based on the climate regionalization results, determine an optimum neural network for each sub-region. An artificial neural network (ANN) was developed for each of the sub-regions. As a result, if K determines the number of sub-regions, then K ANNs are trained. A rainfall prediction model for the entire area was subsequently developed, based on a combination of sub-region models.

The accuracy of the forecasting based on different numbers of sub-regions was compared to select the optimal number of sub-regions for a forecast model for the area.

Feed forward backpropagation (FFBP) was considered as the structure of the ANN in this study. One hidden layer is required to approximate most of the continuous functions (Hagan and Menhaj 1994; Reusch and Alley 2002). The Levenberg-Marquardt training method was selected due to the efficiency of this method in the optimization (Hagan and Menhaj 1994).

The optimal size of hidden layers and hidden neurons is dependent on both the complexity of the forecasting problem and the selected architecture (Reusch and Alley 2002). The tan-sigmoid transfer function in the hidden layer and a linear function in the output layer were used. All input variables were normalized within a range of (−1, 1) to maximize the efficiency of the ANN.

The number of input neurons depends on the number of selected predictors from lagged climatic indices (Table 1) by FRA. Setting the number of hidden neurons is a critical issue for developing an accurate ANN. In this study, the number of hidden layers was fixed to one; the number of hidden neurons was increased from 6 to 50. The best number of hidden neurons was selected by comparing the performance of different models using the root mean square error (RMSE). In order to check the accuracy of the ANN model, the dataset was randomly partitioned into 70, 15, and 15 % as a training, test, and validation dataset, respectively. The ability to make accurate predictions of the optimal network was checked using validation set. The model structure that has the smallest average error on the validation set is selected.

2.2.5 Model accuracy and forecast skill assessment

To select the model and the number of sub-regions that improve forecast performance, common statistical parameters such as coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were calculated (Stanski et al. 1989).

3 Results

3.1 Spatiotemporal rainfall variability

As highlighted earlier, the majority of variance (93 %) of monthly rainfall was explained by the first fifty PCs. The first

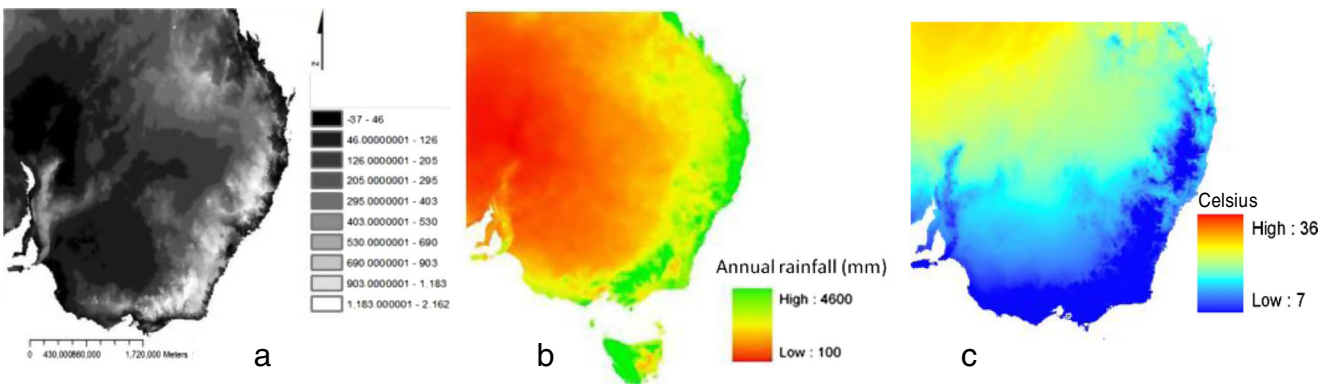

Fig. 4 **a** Digital elevation map of the area of the study. **b** and **c**, respectively, mean annual rainfall and Australia maximum temperatures adopted from BOM

eight PCs, explaining 70.2 % of the variation were mapped using kriging at a resolution of 1 × 1 km as a way to explain the main spatial variation in the long-term pattern of monthly rainfall (Fig. 3). Locations with higher scores in each of the PCs have a higher contribution to the temporal variations in monthly rainfall described by that PC. The first principal component (PC1) accounts for 34 % of long-term monthly rainfall variations (Fig. 3a). The first PC is uniform north to south, but varies from west to east, with negative scores in the western part of the area and positive scores in the eastern part (Fig. 3a). Comparing the pattern of PC1 with the long-term mean annual rainfall (Fig. 4b) and digital elevation map (DEM) indicates that the first PC represents the annual rainfall pattern variability of the area (Fig. 4a). Higher elevations are an important variable that causes extreme rainfall variations in space and time. The East Australian Cordillera, which runs along the eastern coastal area, has a strong influence on rainfall variations in this area. Complex interactions between wind, distance from the coastal area, higher altitudes, and their shadow effect make the investigation of the relationship between altitude and rainfall patterns hard (Goovaerts 2000).

The spatial variations in PC2, which represent 12.4 % of the total variation in the long-term monthly rainfall pattern, show a uniform west-east pattern. The spatial variability of PC2 is mainly in the north-south direction, with positive loading values in northern part and negative loading values in southern part. The spatial pattern of this map is similar to the spatial pattern of maximum and mean temperature in the area

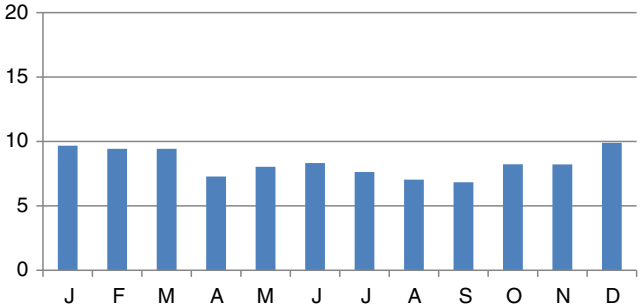

Fig. 6 Annual cycle of the monthly rainfall index for all 136 stations

(Fig. 4c). Also, the rainfall temporal variation of the area is mostly in the north–south direction, from moist temperate with rain in summer in the north to moist temperate with rain in winter in the south. Furthermore, 7.3 and 5.3 % of the temporal variations of long-term rainfall are explained by PC3 (Fig. 3c) and PC4 (Fig. 3d), and both these PCs and PC5 and PC6 show some spatial variation across the area. However, these lower order PCs do not seem to relate to any clear climate patterns. They are hard to interpret as there are complex relationships between geographical features and large-scale climate indices with rainfall patterns and variability. However, it seems that these PCs mainly represent coastal fringe rainfall patterns. PC7 and PC6 illustrate weak or similar scores over almost the entire area, and there is no clear spatial variation patterns associated with these PCs.

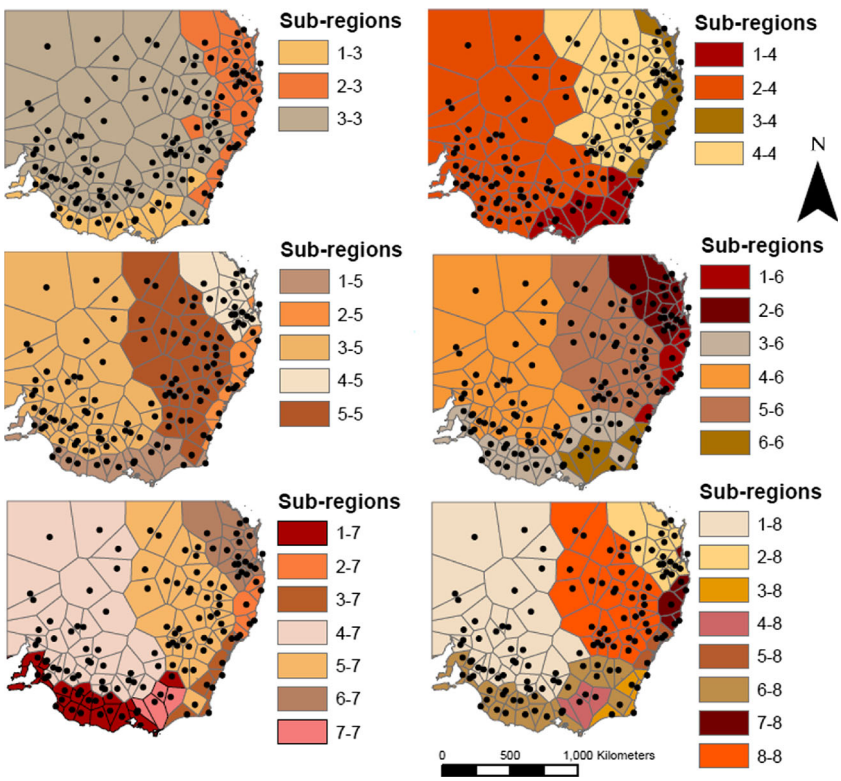

Fig. 5 Maps of location of weather stations including different number of climatic zones have been determined by similarity of rainfall patterns. The area of the study were divided into three (**a**), four (**b**), five (**c**), six (**d**), seven (**e**), and eight (**f**) different major sub-regions

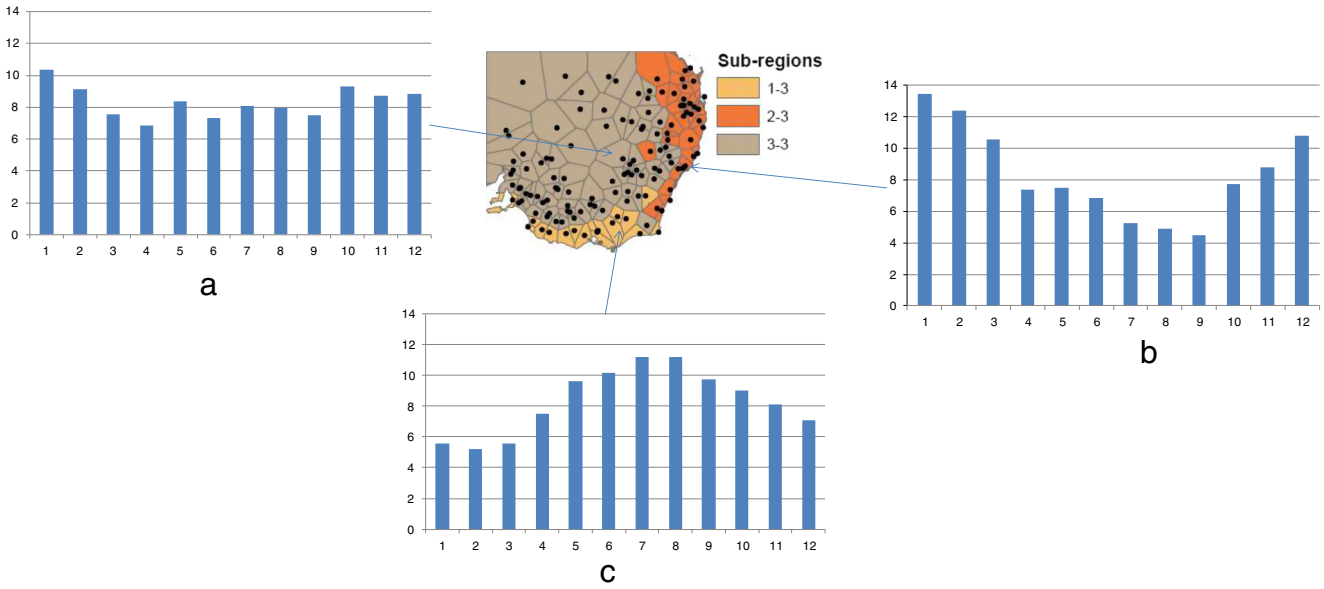

Fig. 7 Annual cycle of monthly rainfall index for three sub-regions

3.2 Climate regionalization

The results of the climate regionalization with three to eight clusters are mapped in Fig. 5. The results are homogenous from a climate and geographic perspective. Sub-regions are geographically well distributed over the area. Each of these maps is considered a climate definition in this study, resulting in six definitions with different numbers of sub-regions.

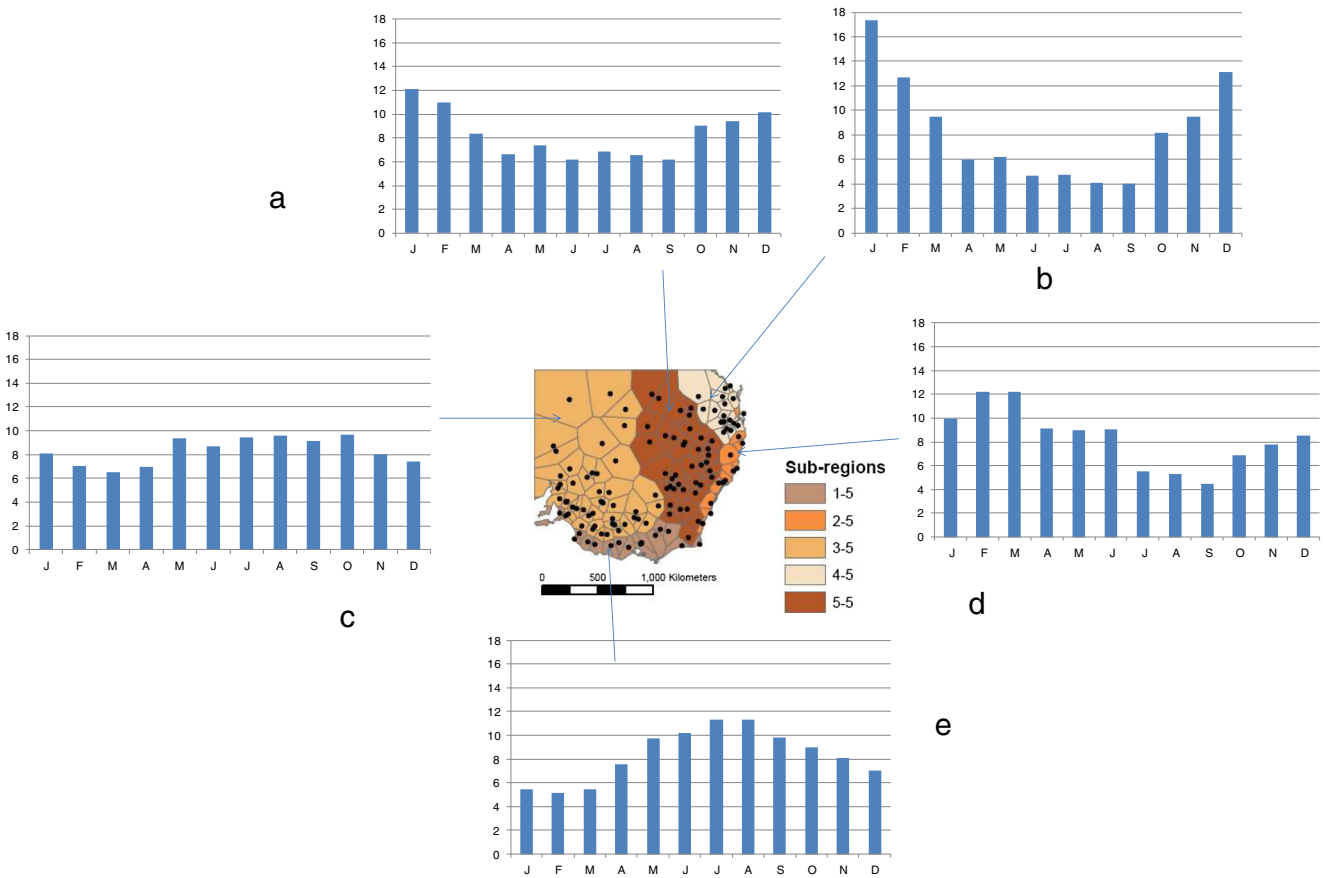

Fig. 8 Annual cycle of monthly rainfall index for five sub-regions (map (c)—Fig. 9)

Table 2 Significant climate variables for each climate definition

Climate regionalization Fig. 7	Sub-region 1	Sub-region 2	Sub-region 3	Sub-region 4	Sub-region 5	Sub-region 6	Sub-region 7	Sub-region 8
6	EPI (−2)	NINO 1+2 (−1)	II (−1)	[a]	NINO 1+2 (−6)	[a]	NINO 4 (−2)	NINO 1+2 (−3)
Map (f)		NINO 3 (−4)			NINO 4 (−1)			NINO 1+2 (−6)
		NINO 1+2 (−4)						NINO 4 (−6)
		NINO 3.4 (−3)						
5	NINO 1+2 (−3)	NINO 1+2 (−3)	TSI (−2)	NINO 1+2 (−6)	NINO 1+2 (−6)	NINO 1+2 (−1)	[a]	
Map (e)	NINO 3.4 (−1)	NINO 1+2 (−6)		NINO 4 (−1)				
		II (−3)						
4	NINO 1+2 (−3)	EPI (−2)	NINO 1+2 (−3)	NINO 1+2 (−6)	NINO 1+2 (−5)	NINO 3.4 (−5)		
Map (d)	NINO 1+2 (−6)		NINO 3.4 (−1)	NINO 3.4 (−1)	EPI (−2)			
					NINO 3.4 (−2)			
					NINO 4 (−4)			
					NINO 3.4 (−3)			
					EPI (−6)			
					NINO 3.4 (−4)			
					II (−3)			
					II (−2)			
					TSI (−4)			
3	NINO 1+2 (−4)	NINO 1+2 (−3)	NINO 1+2 (−6)	EPI (−2)	NINO 1+2 (−5)			
Map (c)	NINO 3.4 (−1)	NINO 1+2 (−6)	NINO 4 (−1)	NINO 1+2 (−1)	TSI (−1)			
		II (−2)			NINO 3.4 (−2)			
2	NINO 3.4 (−5)	NINO 1+2 (−6)	NINO 4 (−2)	EPI (−2)				
Map (b)		NINO 1+2 (−3)						
		NINO 1+2 (−1)						
1	TSI (−4)	EPI (−2)	NINO 1+2 (−6)					
Map (a)	NINO 3 (−3)	WPI (−1)	NINO 3.4 (−1)					
		NINO 4 (−4)						
The entire area	NINO 1+2 (−6)							

Z (−N): N is lagged months of climate indices Z, for example, NINO 1+2 (−3) indicated 3-month-lagged NINO 1+2

[a] Fuzzy ranking did not find a significant variable for this sub-region, means that none of the input variables satisfy the FRA criteria's

Increasing the number of clusters results in more sub-regions in the eastern and coastal part of the area. However, the sub-regions in the central part of the study area remain quite stable.

3.3 Spatiotemporal rainfall variability

K-means gathers together stations with similar rainfall patterns and regimes. The different inter-annual regimes of climate sub-regions can be further analyzed based on a monthly rainfall index. The rainfall monthly index I_i is calculated within each sub-region to evaluate and compare inter-annual rainfall variations within different sub-regions:

$$I_i = \frac{R_i}{R_t} \times 100 \tag{3}$$

Where I_i is the monthly index for the month i, R_i is total rainfall for month i, and R_t is the total annual rainfall of stations in the sub-region. Figure 6 shows the annual cycle of the monthly rainfall index (I_i) for the entire area (without climate regionalization). The monthly rainfall index is evenly distributed across all months, with a slightly higher monthly rainfall index during the Australian summer.

Table 4 Statistical parameters between observed and predicted rainfall for test dataset

Climate definition	R^2	RMSE (mm)	MAE (mm)
(1) Map a	0.66	44.97	29.48
(2) Map b	0.64	51.7	34.25
(3) Map c	0.64	49.78	31.3
(4) Map d	0.68	38.5	25.48
(5) The entire area	0.59	49.23	31.7
(6) The null model	0.39	60.11	38.07

Figure 7 shows the monthly rainfall index for each of the sub-regions when the area is divided into three sub-regions. A clear contrast in rainfall regimes occurs between the eastern part (sub-region 2–3) and southern part (sub-region 1–3) as shown in Fig. 7b, c. There is a clear contrast in the rainfall pattern between these two areas following the Austral winter and summer. However, Fig. 7a shows that the monthly rainfall index for sub-region 3–3 has an approximately uniform annual pattern with a slightly higher monthly rainfall index during the Austral summer.

Visually, the difference between I_i in different sub-regions becomes more significant, if the number of sub-regions is increased to five (Fig. 8). There is strong austral summer dominated regime in the northern part of the area (Fig. 8b) and austral winter dominated rainfall regimes in southern part

Table 3 Statistical parameters between observed and predicted rainfall in each of the climate regionalization definitions

Climate definition	Sub-region	Model structure	Train		Validation		Test	
			RMSE	R^2	RMSE	R^2	RMSE	R^2
The entire area	–	5–40–1	47.43	0.62	47.74	0.62	49.58	0.61
(1) Map a	1–3	6–49–1	41.02	0.76	46.46	0.72	44.10	0.73
	2–3	7–48–1	64.78	0.68	71.29	0.64	66.88	0.67
	3–3	6–46–1	31.66	0.63	31.10	0.63	32.37	0.62
(2) Map b	1–4	5–48–1	24.78	0.70	25.69	0.69	24.93	0.69
	2–4	7–50–1	79.16	0.65	82.34	0.64	83.09	0.62
	3–4	5–43–1	48.76	0.62	50.63	0.61	52.11	0.61
	4–4	5–44–1	76.98	0.68	86.00	0.64	84.99	0.61
(3) Map c	1–5	6–42–1	23.25	0.67	23.29	0.65	23.72	0.65
	2–5	7–45–1	60.16	0.67	61.49	0.62	73.45	0.61
	3–5	6–41–1	36.28	0.72	38.87	0.69	37.15	0.70
	4–5	6–39–1	27.72	0.77	28.34	0.75	28.90	0.76
	5–5	7–47–1	63.14	0.63	70.05	0.60	68.35	0.63
(4) Map d	1–6	6–43–1	28.3	0.75	30.6	0.72	29.8	0.73
	2–6	5–20–1	64.43	0.61	66.32	0.61	76	0.58
	3–6	6–44–1	29.8	0.73	32	0.7	31.47	0.71
	4–6	6–46–1	25.82	0.52	26	0.52	26.7	0.5
	5–6	14–46–1	36.57	0.7	40.8	0.67	38.94	0.66
	6–6	5–35–1	62	0.68	69.65	0.65	67.99	0.65

(Fig. 8e) in this climate regionalization. By increasing number of sub-regions to more than five, some of the sub-regions showed similar annual cycles.

3.4 Identification of significant input variables

There are six climate regionalization definitions for the area (Fig. 5) based on the number of sub-regions. Table 2 indicates the ranked input variables for each sub-region in these six climate regionalizations and for the entire area. In this table, climate regionalization (1) is related to map (a) in Fig. 5, where the area is divided into three sub-regions. Different significant predictors are found for long-term monthly rainfall for each of sub-regions. For example, significant predictors for sub-region 1–3 (Fig. 5-map a) are 4-month-lagged TSI and 3-month-lagged NINO 3. On the other hand, significant input variables for sub-region 2–3 in the same climate regionalization (Fig. 7—map (a)) are 2-month-lagged EPI, 1-month-lagged WPI, and 4-month-lagged Niño 4. In the same climate regionalization, significant input variables for sub-region 3 are NINO 1+2 and NINO 3 with 6- and 1-month lag, respectively. The same type of variation occurred across all different climate regionalization definitions. For the entire area, 6-month-lagged NINO 1+2 was found as the most important predictor. The main outcome here is that this confirms earlier research that there is significant spatial variation in climate predictors across Australia (Kirono et al. 2010).

Subsequently, these identified inputs are used to develop a forecasting model using an artificial neural network. A forecast model was developed for climate regionalization definitions for which FRA found at least one significant input variable for all sub-regions. Therefore, a forecasting model was developed for the first four climate definitions. For climate regionalization definitions five and six as in sub-region 7 and sub-region 4 and 5, respectively, none of inputs could satisfy the FRA criteria and these were therefore omitted from the modeling part.

3.5 Modeling results

This covers the comparison between the result of the entire area and the climate regionalizations with different numbers of clusters.

The performance and structure of the optimized models is described in Table 3, which presents the coefficient of determination (R^2) and RMSE between observed and predicted rainfall for the training, validation, and test datasets. For each network, the optimized structure was selected based on model performance described by a lower RMSE and a higher R^2. Performance of models improved using climate regionalization. The best result was obtained for climate regionalization 4 with six sub-regions having the lowest RMSE (38.5) and highest R^2 (0.68) (Table 4). The second best result was obtained for climate regionalization 1 with three sub-

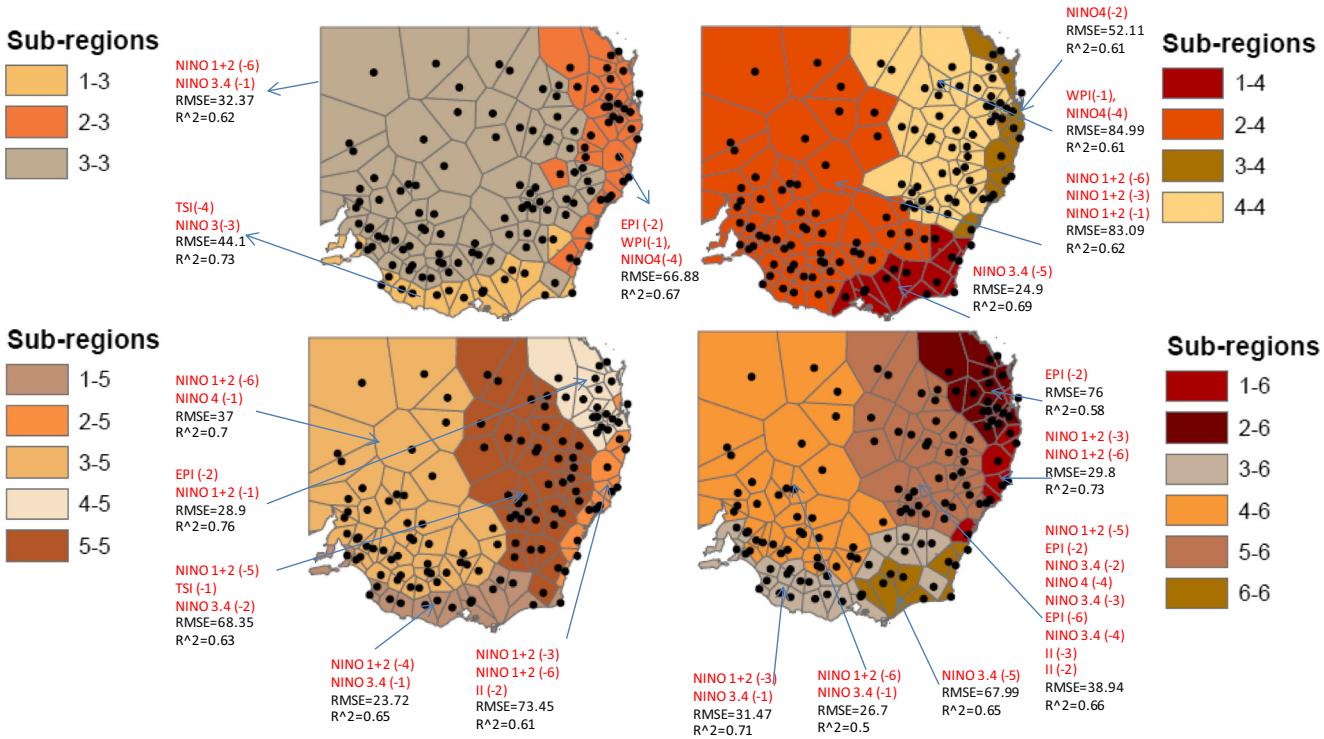

Fig. 9 Maps of four selected climate definitions with **a** three sub-regions, **b** four sub-regions, **c** five sub-regions, **d** six sub-regions. Significant predictors and performance of model for each sub-region are indicated

regions, compared to a forecast for the entire area (Table 4). Model performances can also be compared to the null model (mean rainfall of each station/each month), which indicates a lower R^2 (0.39) and higher RMSE (60.11). This indicates that ANN models perform better than the null model. Finally, box plots of monthly rainfall prediction residuals for all the models, based on the validation dataset, are shown in Fig. 10. It is clear that the null model always under predicts the actual rainfall, while the ANN models have more equal under and over predictions, with an overall a lower RMSE.

Similar to the predictor analysis (FRA), the ANN results indicate strong spatial variability in the performance of the models, with some regions clearly performing better. For example, in climate regionalization (1) (Fig. 9a), internal subregions and on the Victorian south coast tend to perform better than in sub-region 2–3 covering east coastal area. Comparing the result of modeling for different climate regionalization is however difficult as different predictors are used for each subregion Fig. 10.

A multiple comparison of the predicted values of the models for the test dataset was run to test the hypothesis that the means of model prediction values are significantly different, against the alternative that they are the same (with 95 % level of significance). Table 5 shows the results of multiple comparisons of the predicted values of the models for the test dataset. Table 5 includes one row per comparison and five columns; columns 1–2 are the indices of the two samples being compared. Columns 3–5 are a lower bound, estimate, and upper bound for each pair of models difference. For example, the first row of Table 5 indicates that the differences of

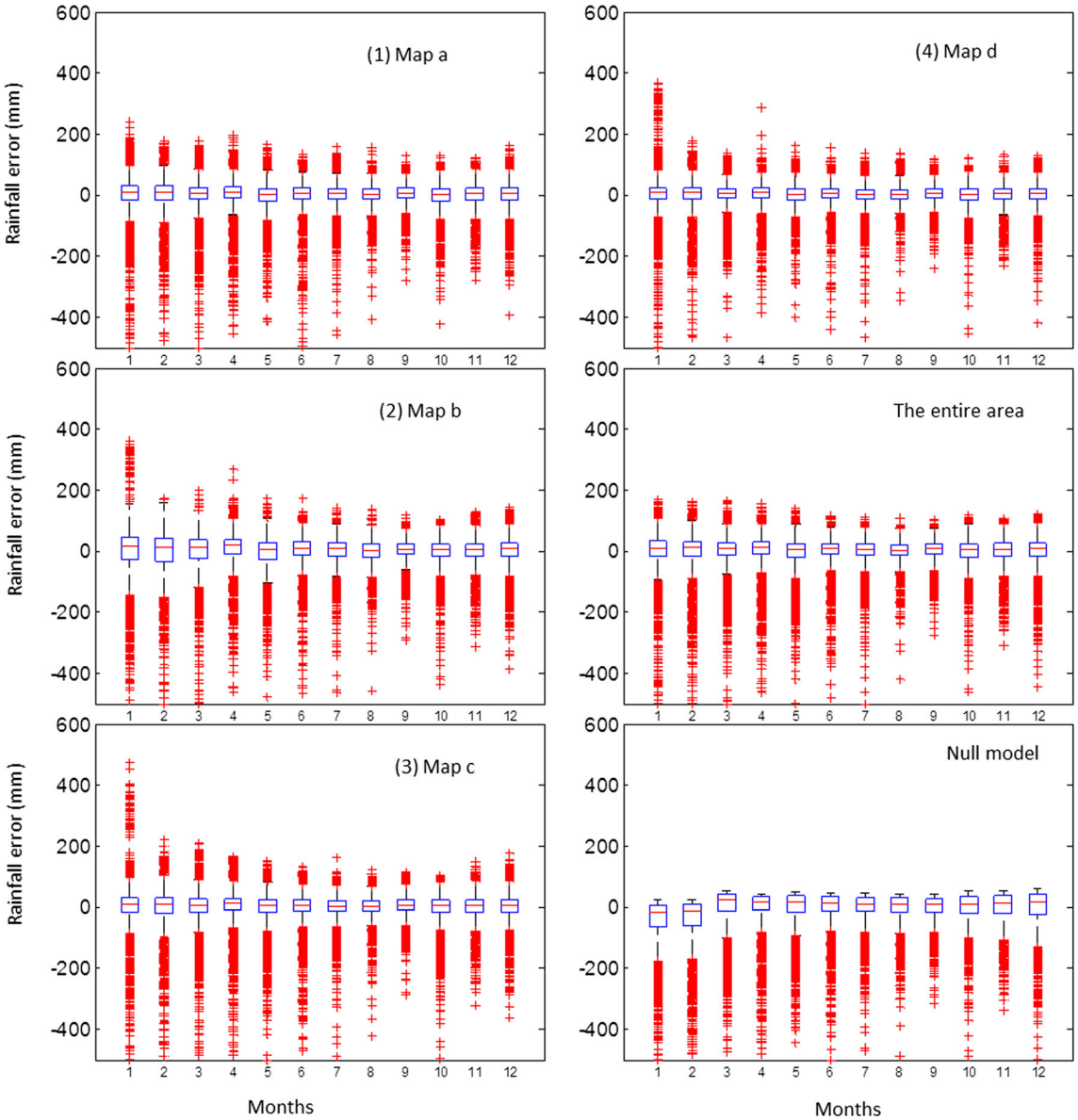

Fig. 10 *Box plots* of monthly rainfall prediction error to compare the different models

Table 5 Multiple comparison test for the means of the predicted value of the models for the test dataset

Model	Model	Lower bound	Estimate	Upper bound	*P* value
1	2	−8.53	−8.07	−7.60	9.78e-10
1	3	−2.6	−2.18	−1.74	1.2139e-05
1	4	2.99	3.42	3.84	8.9582e-06
1	5	−0.67	−0.24	0.18	0.0023
1	6	6.45	6.88	7.3	8.9502e-06
2	3	5.42	5.88	6.35	3.0232e-10
2	4	11.04	11.49	11.95	0
2	5	7.365	7.82	8.28	1.8687e-194
2	6	14.49	14.95	15.41	1.8687e-194
3	4	5.173	5.605	6.03	0.00045
3	5	1.49	1.94	2.37	0.00032
3	6	8.62	9.064	9.503	0
4	5	−4.09	−3.66	−3.24	0.0034 e-06
4	6	3.034	3.45	3.88	0.000483
5	6	6.696	7.12	7.55	0

the mean of predicted results of model 1 and 2 are estimated to be −8.07, and a 95 % confidence interval (CI) for this difference is [−8.53, −7.60]. This interval does not contain 0. This indicated that means of Model 1 and Model 2 predicted values for test dataset are different. The only confidence interval in Table 5 including 0 was in case of comparing Model 1 (Map (a)) and Model 5 (the model for the entire area). This shows that the means of these two models are not significantly different. Figure 11 displays the estimated means of predicted values of each models (Tables 4 and 5) with comparison intervals around them. Red points indicate models that are not significantly different and blue points are significantly different (with 95 % of level of significant). Conclusion of the results of multiple comparisons, presented in Table 5 and Fig. 11, is that the model results are significantly different except Map (a) and the model for the entire area.

3.6 Spatiotemporal analysis and validation of forecast skill

Box plots of the monthly rainfall and model error (observed value minus predicted value) are shown in Fig. 12 for six sub-regions in climate regionalization 4. Comparing the distribution of forecast values to the distribution of the error values shows that regions and times (months) with higher variability of rainfall have a higher absolute error, suggesting a level of randomness in the rainfall variability, which is not well explained by the climate-based predictors. Model error is mostly negative which shows that model is predicting less than expected rainfall.

Table 6 shows statistical parameters of model performance of climate definition 4, as the best performance, in seasonal time scale for each sub-region in each season. There is a relationship between the amount of rainfall and the absolute error in space and time. Variations in total rainfall between sub-regions take into account by calculating normalized RMSE values (NRMSE) for each sub-region over a season which is the ratio of RMSE to the mean monthly rainfall of the corresponding season (Table 6). The smallest NRMSE (0.45) was obtained during winter in sub-region 1–6. The smallest NRMSE indicates better performance of the model. The

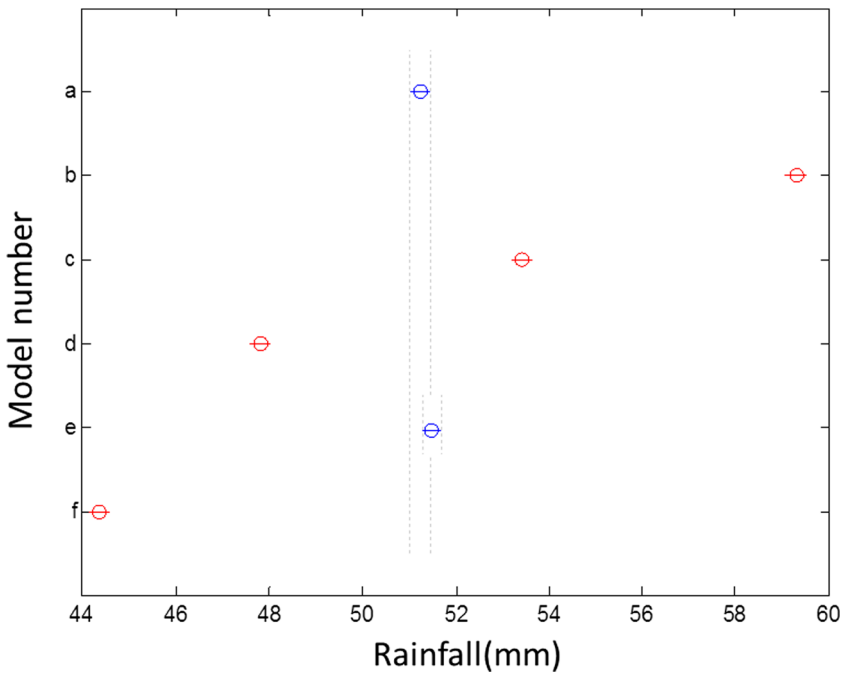

Fig. 11 Multiple comparison of the mean predicted values of the models. *Red points* indicate models that are not significantly different, and *blue points* are significantly different (with 95 % of level of significant)

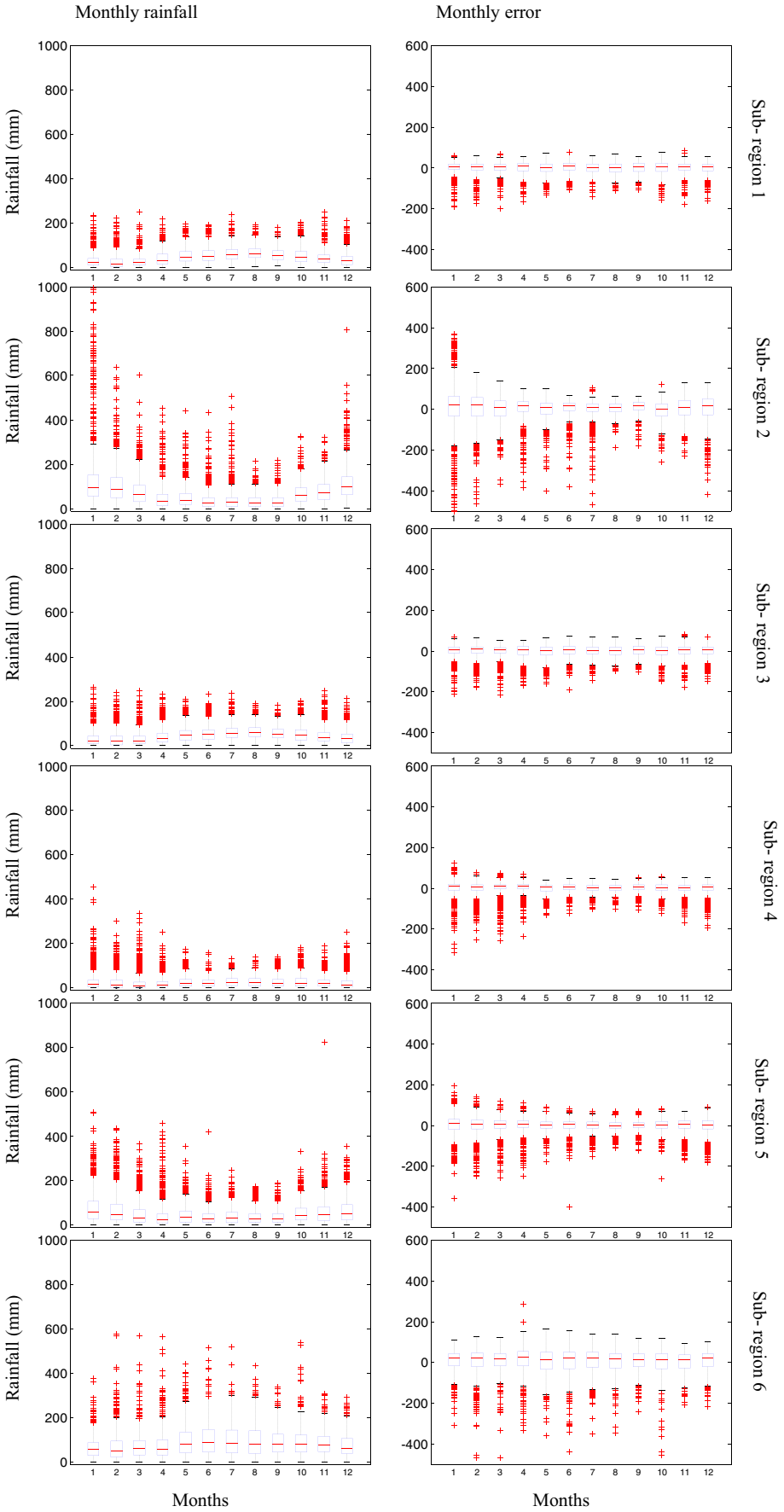

Fig. 12 *Box plots* of the monthly rainfall (*left*) and error (*right*) related to each sub-region

Table 6 Statistical parameters between observed and predicted rainfall for each season in climate definition 4

Location	Season	RMSE (mm)	R^2	NRMSE	Mean rainfall (mm)	Standard deviation
Sub-region 1–6	DJA	29.1	0.6	0.91	31.7	31.7521
	MAM	29.6	0.7	0.71	41.4	32.93
	JJA	28.48	0.8	0.45	61.99	32.6246
	SON	28.54	0.8	0.56	50.88	31.9016
Sub-region 2–6	DJA	100	0.63	0.83	121.1	115.914
	MAM	56.94	0.56	0.94	60.4	61.29
	JJA	42.2	0.45	1.1	37.9	42.4949
	SON	44.03	0.67	0.72	60.7	47.959
Sub-region 3–6	DJA	31.83	0.58	0.9	35	34.8746
	MAM	31.62	0.66	0.75	41	34.5231
	JJA	28.8	0.81	0.48	59.2	32.7669
	SON	28.9	0.77	0.56	51.1	32.5657
Sub-region 4–6	DJA	31.67	0.45	1.22	25.76	34.4208
	MAM	26.48	0.48	1.138	23.25	28.6820
	JJA	18.51	0.69	0.71	25.77	21.1780
	SON	28.9	0.77	0.56	51.1	25.4063
Sub-region 5–6	DJA	42.7	0.78	0.61	69.7	61.7054
	MAM	32.92	0.73	0.75	43.69	47.5069
	JJA	24.29	0.74	0.67	35.95	31.5903
	SON	29.35	0.78	0.61	47.89	41.1653
Sub-region 6–6	DJA	61.64	0.6	0.83	73.75	63.9337
	MAM	68.5	0.64	0.79	89.31	76.4021
	JJA	67.68	0.71	0.66	101.36	76.6017
	SON	58.19	0.71	0.64	89.99	62.4090

model error changed in space and time, whereby a higher amount of rainfall resulted in a higher corresponding error (Table 6). The model produced for sub-region 2–6, resulted in higher RMSE in summer and autumn (100 and 56.94 mm/month), compared to the same area in winter and spring (42.2 and 44.03 mm/month), on account of higher rainfall in the summer. Sub-region 4–6 with less rainfall in comparison to other sub-regions, especially during winter and spring has the smallest RMSE. In sub-region 2–6 with higher rainfall during summer (DJA), higher RMSE was obtained.

3.7 Forecast validation

To check the ability of model to forecast rainfall, data from 2010 to 2011 for all stations were separated for specific validation. Models were developed using the remaining data for each sub-region. This period is a La Niña period (2010–2011), with record-breaking rainfall over Australia. Figure 13 shows the rainfall forecast for selected stations for each sub-region. Table 7 shows the statistical parameters of the model forecast for monthly rainfall over 2010–2011 for some random stations.

4 Discussion

A clear outcome of this study is that model performance for rainfall forecasts with a 1-month lead time for south-eastern and eastern Australia can be improved by using climate regionalization (Tables 3, 4 and Fig. 9). The number of sub-regions considered in the clustering plays an important role in the results of FRA and the performance of models. By changing the number of clusters, even though some sub-regions seem unaffected, the results of FRA and consequently model performance are quite different (Fig. 9). For example, in Fig. 9, sub-region 1 in map a, b, and c is approximately located in a similar location with differences for some of the stations; however, the results of the FRA in map a, b, and c are [NINO 3 (−3), TSI (−4)], [NINO 3.4 (−5)], [NINO 1+2 (−4)], [NINO 3.4 (−1)], respectively. This may be a drawback of FRA whereby the distribution of the sampling dataset has an influence on the outcome of the method (Hou et al. 2007).

The results support the notion that Australian rainfall is more sensitive to the variation of the central and western Pacific Ocean SST anomalies compared to the eastern Pacific Ocean SST anomalies (Wang and Hendon 2007), as NINO 1+2, NINO 3, and NINO 3.4 were selected more

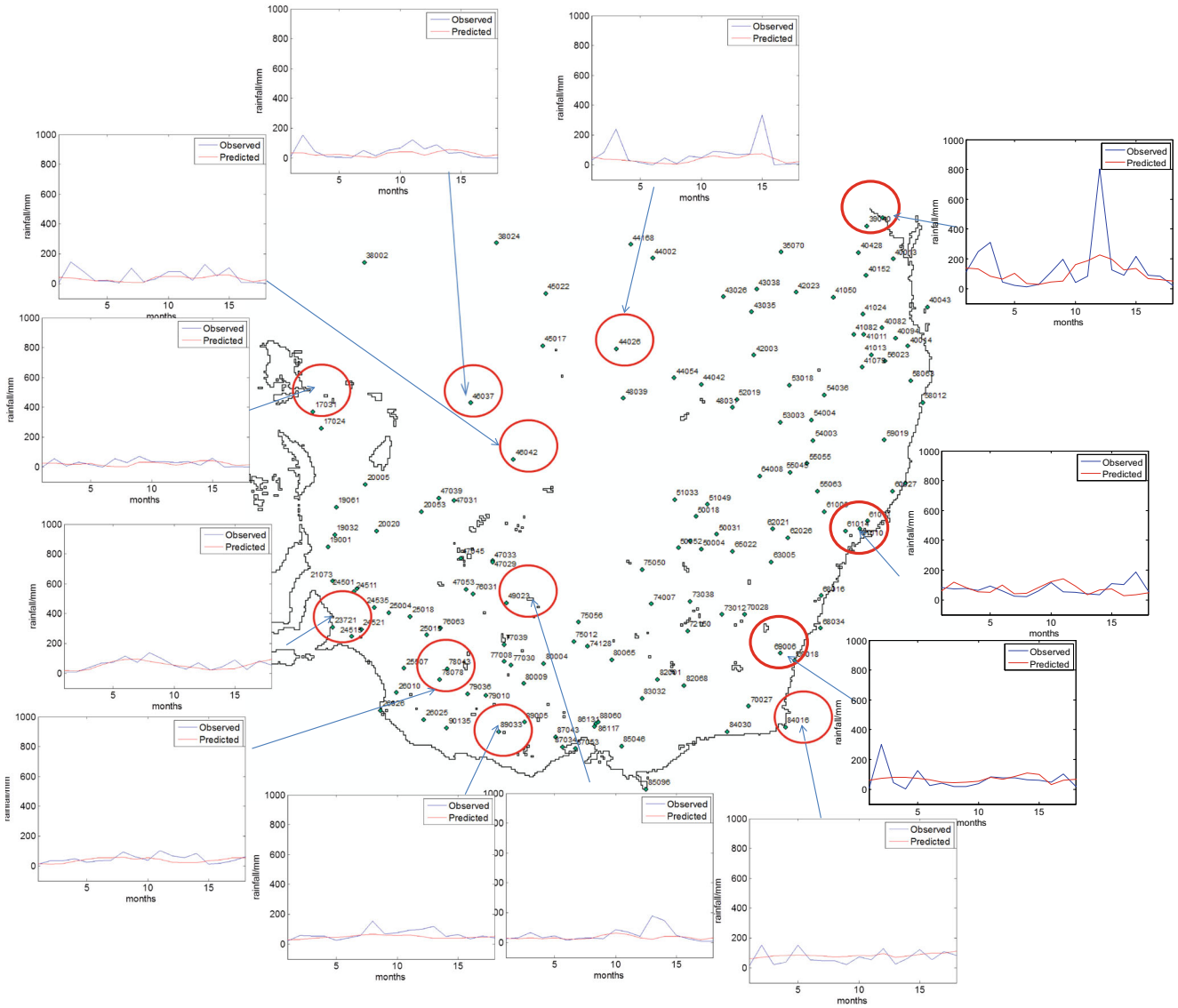

Fig. 13 Rainfall prediction for selected stations over the period of 2010–2011

frequently, compared to NINO 4. The best result was obtained from climate regionalization 4 (Fig. 9d) with six sub-regions in comparison to other models and the null model (Table 4). The model's errors are different in time and space following the rainfall patterns (Table 6). The model results, as for the FRA results, varied depending on the number of clusters included. Monthly rainfall forecast models still retain the tendency to produce high prediction error, but as identified here, the error is higher in particular regions, especially for sub-regions and seasons with higher rainfall.

Previous studies investigating relationships between climate indices and rainfall for the development of rainfall forecast models used common boundaries within Australia without considering the similarity in rainfall patterns and regimes (Anwar et al. 2008; Drosdowsky 2002; Drosdowsky and Chambers 2001; Kirono et al. 2010; McIntosh et al. 2007; Stone and Auliciems 1992). Comparison of the findings between the current study and these studies is difficult due to the different time and space scales considered. However, the results of the current study are comparable to two recent studies by Kirono et al. (2010) and Schepen et al.(2012a).

Kirono et al. (2010) investigated the relationship between 12 atmospheric predictors for rainfall over the south-eastern Australia and showed how this relationship changed in each season. It was observed that south-eastern Australian spring rainfall is strongly correlated with climate indices. The relationship between lagged climate indices and spring (SON) rainfall was significant through south-eastern Australia but not consistently for other seasons. NINO 4 [sea surface temperature (SST) in western Pacific] and thermocline (2nd EOF of 20 °C isotherm of Pacific Ocean) were the best predictor for spring rainfall over the south-eastern and eastern part of

Table 7 Statistical parameters for performances of model for random stations (2010–2011)

Station ID	Sub-region	RMSE	MAE	R^2
89033	1	31.3	22.7	0.78
23721	1	24	19.3	0.88
84016	6	40.39	35.77	0.74
39040	2	145.6	86.07	0.5
46037	4	42.07	30.60	0.48
69006	6	62.71	41.57	0.51
61014	5	47.62	33.37	0.68
38024	4	71.63	36.51	0.43
17031	4	20.15	16.21	0.53
78078	1	27.01	21.73	0.71
44026	5	36.13	25.54	0.70
49023	4	47.44	26.43	0.50
46042	4	40.72	27.41	0.56

Australia based on Kirono's results. Also, spring rainfall is lagged correlated with SOI, NINO 3, II, IOW (index of west pole), and SST1 (the first principal component of SST over the Indian and Pacific Ocean). In summer time, significant predictors are SOI (Southern Oscillation Index), thermocline, NINO 3, NINO 4, and SST1. In autumn and winter, thermocline was identified as the most important indicator. In the current study, important predictors for monthly rainfall over a year selected for specific area using FRA. This makes comparison of this study with Kirono et al. (2010) difficult. Although, selected predictors in the current study show up as being promising in their study as well. For example, SST indices over Pacific Ocean (NINO 1+2, NINO 3.4 and NINO 4) were selected predictors in the current study and in the area with uniform regime of rainfall and summer regime, SST indices from Indian Ocean were selected as important predictors.

Schepen et al. (2012a) developed forecasts using a single predictor Bayesian model averaging (BMA) and validated the forecasting ability of each predictor over Australia. The selection of predictors was based on the ability to forecast seasonal rainfall in different regions. The result of their study indicated that there is strong evidence supporting using NINO 4 and NINO 3.4; although NINO 1+2, which was selected as a predictor for the entire area and for most of sub-regions in the current study, was not considered in Schepen's analysis. Higher forecast skill scores from their study were obtained by using NINO 3.4 in most seasons.

The relationship and impact of the Pacific Ocean-atmospheric circulation relative to the Indian Ocean circulation are a challenge for Australian climate studies (Ashok et al. 2003b; Meyers et al. 2007). Schepen et al. (2012a) discovered that DMI (WPI-EPI) is able to predict rainfall in south-eastern Australia in Sep-Oct-Nov and Oct-Nov-Dec. Also, they found that NINO 4 and NINO 3 provide strong season predictability in these seasons. They were suggested further analysis to check the independency of DMI form Indian Ocean and NINO indices. As the structure of FRA is based on removing the correlated and dependent variables to the first significant predictor method (Hou et al. 2007), this method can address independent predictors selection requirement. In the current study, the area representing uniform rainfall patterns and summer regime has predictors from both Indian and Pacific Ocean which will be due to different activation times of Pacific and Indian Oceans events. Climate indices used for all sub-regions as predictors (input of FRA) were fixed; therefore, the pattern of rainfall (output) in each sub-region plays an important role in FRA step. This shows that there is an interaction between rainfall regime and active time of Indian and Pacific Ocean (Zhao and Hendon 2009). For example, in the coastal area in the east with a mixed uniform and summer rainfall regime sub-region 2 (Fig. 9a), selected predictors are from climate indices over both Indian and Pacific Ocean that are EPI (−2), WPI (−1), and NINO 4 (−4), in comparison to selected predictors for sub-region 1, with winter rainfall regime, that are NINO 3 (−3) and TSI (−4).

Schepen et al. (2012a) indicated that forecast skill score was low using 1-month-lag TSI as predictor, apart from western Australia. In the current study, TSI was only a predictor in the southern part of the area with a 4-month lag. Wang et al. (2012) developed a BJP model based using several predictors for the entire Australia territory. Selection of predictors in this study was based on the best forecast skill score over Australia. Wang et al. (2012) examined the contributions from the Pacific, Indian, and extratropical groups. The forecast skill score was low for the first half of the year, and for second part of the year, a more positive forecast skill score was achieved (higher is better). Overall, they found Pacific SST indices resulted in higher skill score model. Indian and extratropical groups also produced useful and sometimes distinct skill. The difference with this study and the method employed here is that multiple predictors were ranked as the predictors for each sub-region with the similar rainfall patterns and then models were developed based on selected predictors. The outcome is that the model in our study is working differently for each season in each sub-region.

Further development of the model is underway to produce different lead times (a week, a season, and a year ahead) for predicting rainfall. As important input variables varied for similar sub-regions by changing the number of clusters (Fig. 9), finding a robust method for identification predictors is a further aim for future model development. The result of climate regionalization by K-means is relatively coarse. In addition, changing the size of sub-regions resulted in a changed performance of the model in some regions. A flexible clustering method (for example, fuzzy clustering) for

grouping stations together which gives more freedom to change the size of the sub-regions will also be considered in future study.

5 Conclusions

This study attempted to predict monthly rainfall for stations in South-eastern and eastern part of Australia using lagged climatic indices. The performance of model for forecasting 1-month lead time rainfall for south-eastern and eastern Australia improved with using climate regionalization. The number of sub-regions played an important role in the result of the predictor selection using FRA and the performance of model. The best result was achieved when the area was divided to six sub-regions. The model residuals vary in time and space, which follows the rainfall patterns. In general, higher residuals are correlated with higher rainfalls, suggesting some level of smoothing by the model. In general, it may be concluded that climate regionalization improved the accuracy of spatiotemporal monthly model and could be considered for other time scales and other climatological and meteorological studies.

Acknowledgments The authors gratefully acknowledge funding for project provided by the Australian Research Council and Grain Growers Limited. The authors would like to thank to Dr. Floris Van Ogtrop for valuable discussions and advice.

References

Anwar MR, O'Leary G, McNeil D, Hossain H, Nelson R (2007) Climate change impact on rainfed wheat in south-eastern Australia. Field Crop Res 104(1–3):139–147

Anwar MR, Rodriguez D, Liu DL, Power S, O'Leary GJ (2008) Quality and potential utility of ENSO-based forecasts of spring rainfall and wheat yield in south-eastern Australia. Aust J Agric Res 59(2):112–126

Ashok K, Guan ZY, Yamagata T (2003a) Influence of the Indian Ocean Dipole on the Australian winter rainfall. Geophys Res Lett 30(15)

Ashok K, Guan ZY, Yamagata T (2003b) A look at the relationship between the ENSO and the Indian Ocean Dipole. J Meteorol Soc Jpn 81(1):41–56

Bannayan M, Lotfabadi SS, Sanjani S, Mohamadian A, Aghaalikhani M (2011) Effects of precipitation and temperature on crop production variability in northeast Iran. Int J Biometeorol 55(3):387–401

Cai W, Whetton PH, Pittock AB (2001) Fluctuations of the relationship between ENSO and northeast Australian rainfall. Clim Dyn 17(5-6): 421–432

Cai W, Cowan T, Sullivan A (2009) Recent unprecedented skewness towards positive Indian Ocean Dipole occurrences and its impact on Australian rainfall. Geophys Res Lett 36(11), L11705

Cai W, van Rensch P, Cowan T, Hendon HH (2012) An asymmetry in the IOD and ENSO teleconnection pathway and its impact on Australian climate. J Clim 25(18):6318–6329

Chiew FHS, Zhou SL, McMahon TA (2003) Use of seasonal streamflow forecasts in water resources management. J Hydrol 270(1-2): 135–144

Corte-Real J, Qian BD, Xu H (1998) Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation. Int J Climatol 18(6):619–635

Dezfuli AK (2011) Spatio-temporal variability of seasonal rainfall in western equatorial Africa. Theor Appl Climatol 104(1-2):57–69

Dezfuli AK, Karamouz M, Araghinejad S (2010) On the relationship of regional meteorological drought with SOI and NAO over southwest Iran. Theor Appl Climatol 100(1-2):57–66

Drosdowsky W (1993) An analysis of Australian seasonal rainfall anomalies 1950-1987: spatial patterns. Int J Climatol 13(1):1–30

Drosdowsky W (2002) SST phases and Australian rainfall. Aust Meteorol Mag 51(1):1–12

Drosdowsky W, Chambers LE (2001) Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. J Clim 14(7):1677–1687

Fawcett RJB, Jones DA, Beard GS (2005) A verification of publicly issued seasonal forecasts issued by the Australian Bureau of Meteorology: 1998-2003. Aust Meteorol Mag 54(1):1–13

Gardner MW, Dorling SR (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 32(14-15):2627–2636

Gerstengarbe FW, Werner PC, Fraedrich K (1999) Applying non-hierarchical cluster analysis algorithms to climate classification: some problems and their solution. Theor Appl Climatol 64(3-4): 143–150

Gong XF, Richman MB (1995) On the application of cluster analysis to growing season precipitation data in north-America east of. Rockies Journal of Climate 8(4):897–931

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228(1-2):113–129

Govindaraju RS (2000a) Artificial neural networks in hydrology. I: preliminary concepts. J Hydrol Eng 5(2):115–123

Govindaraju RS (2000b) Artificial neural networks in hydrology. II: hydrologic applications. J Hydrol Eng 5(2):124–137

Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993

Hendon HH, Thompson DWJ, Wheeler MC (2007) Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode. J Clim 20(11):2452–2467

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441

Hou Y, Zurada JA, Karwowski W, Marras WS, Davis K (2007) Identification of key variables using fuzzy average with fuzzy cluster distribution. IEEE Trans Fuzzy Syst 15(4):673–685

Johnson DE (1998) Applied multivariate methods for data analysts. Duxbury Press

Keogh E, Kasetty S (2003) On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Min Knowl Disc 7(4):349–371

Kirono DGC, Chiew FHS, Kent DM (2010) Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. Hydrol Process 24(10):1237–1247

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Koppen-Geiger climate classification updated. Meteorol Z 15(3): 259–263

Lin YH, Cunningham GA, Coggeshall SV (1996) Input variable identification—Fuzzy curves and fuzzy surfaces. Fuzzy Sets Syst 82(1): 65–71

Lund R, Li B (2009) Revisiting climate region definitions via clustering. J Clim 22(7):1787–1800

McBride JL, Nicholls N (1983) Seasonal relationships between Australian rainfall and the southern oscillation. Mon Weather Rev 111(10):1998–2004

McIntosh PC, Pook MJ, Risbey JS, Lisson SN, Rebbeck M (2007) Seasonal climate forecasts for agriculture: towards better understanding and value. Field Crop Res 104(1-3):130–138

Mekanik F, Imteaz MA, Ieee (2012) Forecasting Victorian spring rainfall using ENSO and IOD: a comparison of linear multiple regression and nonlinear ANN. 2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering, 86-89 pp

Meneghini B, Simmonds I, Smith IN (2007) Association between Australian rainfall and the Southern Annular Mode. Int J Climatol 27(1):109–121

Meyers G, McIntosh P, Pigot L, Pook M (2007) The years of El Nino, La Nina, and interactions with the tropical Indian ocean. J Clim 20(13): 2872–2880

Munoz-Diaz D, Rodrigo FS (2004) Spatio-temporal patterns of seasonal rainfall in Spain (1912-2000) using cluster and principal component analysis: comparison. Ann Geophys 22(5):1435–1448

Murphy BF, Timbal B (2008) A review of recent climate variability and climate change in southeastern Australia. Int J Climatol 28(7):859–879

Nicholls N (1983) The southern oscillation and Indonesian sea surface temperature

Nicholls N (2010) Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958-2007. Clim Dyn 34(6): 835–845

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2(7-12):559–572

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Koppen-Geiger climate classification. Hydrol Earth Syst Sci 11(5): 1633–1644

Piechota TC, Chiew FHS, Dracup JA, McMahon TA (1998) Seasonal streamflow forecasting in eastern Australia and the El Nino Southern Oscillation. Water Resour Res 34(11):3035–3044

Power S et al (1998) Australian temperature, Australian rainfall and the Southern Oscillation, 1910-1992: coherent variability and recent changes. Aust Meteorol Mag 47(2):85–101

Power S, Casey T, Folland C, Colman A, Mehta V (1999) Inter-decadal modulation of the impact of ENSO on Australia. Clim Dyn 15(5): 319–324

Reusch DB, Alley RB (2002) Automatic weather stations and artificial neural networks: improving the instrumental record in West Antarctica. Mon Weather Rev 130(12):3037–3053

Risbey JS, Pook MJ, McIntosh PC, Wheeler MC, Hendon HH (2009) On the remote drivers of rainfall variability in Australia. Mon Weather Rev 137(10):3233–3253

Rogers JC (1990) Pattern of low-frequency monthly sea-level pressure variability (1899-1986) and associated wave cyclone frequencies. J Clim 3(12):1364–1379

Rogers JC, McHugh MJ (2002) On the separability of the North Atlantic oscillation and Arctic oscillation. Clim Dyn 19(7):599–608

Ropelewski CF, Jones PD (1987) An extension of the Tahaiti-Darwin Southern Oscilation Index. Mon Weather Rev 115(9):2161–2165

Ruiz JE, Cordery I, Sharma A (2007) Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor. J Hydrol 341(3-4):156–164

Rumelhart DE, Mcclelland JL (1986) Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations, Medium: X; Size: Pages: 564pp

Saji NH, Yamagata T (2003) Possible impacts of Indian Ocean Dipole mode events on global climate. Clim Res 25(2):151–169

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. Nature 401(6751):360–363

Schepen A, Wang QJ, Robertson D (2012a) Evidence for using lagged climate indices to forecast Australian seasonal rainfall. J Clim 25(4): 1230–1246

Schepen A, Wang QJ, Robertson DE (2012b) Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. J Geophys Res Atmos 117(D20): 20107. doi:10.1029/2012JD018011

Shi G et al (2008) Variability and trend of North West Australia rainfall: observations and coupled climate modeling. J Clim 21(12): 2938–2959

Sivakumar MVK, Hansen J (2007) Climate prediction and agriculture: summary and the way forward. Climate Prediction and Agriculture 1–13. doi:10.1007/978-3-540-44650-7_1

Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). J Clim 21(10):2283–2296

Stanski HR, Wilson LJ, Burrows WR (1989) Survey of common verification methods in meteorology. World weather watch tech. Rep. 8, WMO/TD No. 358

Stone R, Auliciems A (1992) SOI phase relationships with rainfall in eastern Australia. Int J Climatol 12(6):625–636

Stone RC, Hammer GL, Marcussen T (1996) Prediction of global rainfall probabilities using phases of the southern oscillation index. Nature 384(6606):252–255

Tasadduq M, Rehman S, Bubshait K (2002) Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Renew Energy 25(4):545–554

Valle S, Li WH, Qin SJ (1999) Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. Ind Eng Chem Res 38(11): 4389–4401

van Ogtrop FF, Vervoort RW, Heller GZ, Stasinopoulos DM, Rigby RA (2011) Long-range forecasting of intermittent streamflow. Hydrol Earth Syst Sci 15(11):3343–3354

Verdon DC, Franks SW (2005) Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia. Water Resour Res 41(9). doi:10.1029/2004WR003845

Villar JCE et al (2009) Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). Int J Climatol 29(11):1574–1594

Wang G, Hendon HH (2007) Sensitivity of Australian rainfall to inter-El Nino variations. J Clim 20(16):4211–4226

Wang QJ, Schepen A, Robertson DE (2012) Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging. J Clim 25(16):5524–5537

Wilson LL, Lettenmaier DP, Skyllingstad E (1992) A hierarchical stochastic-model of large-scale atmospheric circulation patterns and multiple station daily precipitation. J Geophys Res-Atmos 97(D3):2791–2809

Zhao M, Hendon HH (2009) Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model. Q J R Meteorol Soc 135(639):337–352