Author:	Mekanik, F., Imteaz, M. A., Gato-Trinidad, S. & Elmahdi, A.
Title:	Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes
Year:	2013
Journal:	Journal of Hydrology
Volume:	503
Pages:	11-21
URL:	http://doi.org/10.1016/j.jhydrol.2013.08.035

Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes

F. Mekanik[a], M. A. Imteaz[a] , S. Gato-Trinidad[a], and A.Elmahdi[b]

[a]Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Melbourne, VIC, Australia,

[b] Climate and water division, Bureau of Meteorology, Melbourne, VIC, Australia

Corresponding author: F. Mekanik

Email: fmekanik@groupwise.swin.edu.au

Swinburne University of Technology, Mail 38, FEIS, Hawthorn, VIC, Australia, 3122

Phone: (+61) (03) (92145624)

Abstract

In this study, the application of Artificial Neural Networks (ANN) and Multiple Regression analysis (MR) to forecast long-term seasonal spring rainfall in Victoria, Australia was investigated using lagged El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as potential predictors. The use of dual (combined lagged ENSO-IOD) input sets for calibrating and validating ANN and MR Models is proposed to investigate the simultaneous effect of past values of these two major climate modes on long-term spring rainfall prediction. The MR models that did not violate the limits of statistical significance and multicollinearity were selected for future spring rainfall forecast. The ANN was developed in the form of multilayer perceptron using Levenberg-Marquardt algorithm. Both MR and ANN modelling were assessed statistically using mean square error (MSE), mean absolute error (MAE), Pearson correlation (r) and Willmott index of agreement (*d*).The developed MR and ANN models were tested on out-of-sample test sets; the MR models showed very poor generalization ability for east Victoria with correlation coefficients of -0.99~ -0.90 compared to ANN with correlation coefficients of 0.42~0.93; ANN models also showed better generalization ability for central and west Victoria with correlation coefficients of 0.68~0.85 and 0.58~0.97 respectively. The ability of multiple regression models to forecast out-of-sample sets is compatible with ANN for Daylesford in central Victoria and Kaniva in west Victoria (r=0.92 and 0.67 respectively). The errors of the testing sets for ANN models are generally lower compared to multiple regression models. The statistical analysis suggest the potential of ANN over MR models for rainfall forecasting using large scale climate modes.

Keywords: rainfall, ENSO, IOD, ANN, multiple regression

1. Introduction

Rainfall is final result of complex global atmospheric phenomena and long-term prediction of rainfall remains a challenge for many years. An accurate long-term rainfall prediction is necessary for water resources management, food production and maintaining flood risks. Several large scale climate phenomena affect the occurrence of rainfall around the world; of these large scale climate modes El

Nino southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) are well known for their effect on India, North and South America and Australia. Many studies have tried to establish the relationship between these climate modes for daily, monthly and seasonal rainfall occurrence around the world (Barsugli and Sardeshmukh, 2002; Chattopadhyay et al., 2010; Hartmann et al., 2008; Lau et al., 2001; Shukla et al., 2011; Yufu et al., 2002). This study is motivated by the need to better understand the effect of these climate modes on future seasonal spring rainfall in Victoria, located at southeast Australia (Figure 1). Many past researches have tried to find a relationship between these large climate modes and southeast and east Australian rainfall (Murphy and Timbal, 2008; Verdon et al., 2004), however seasonal rainfall predictability is estimated at an upper limit of only 30% for this region. In the work of Murphy and Timbal (2008) the maximum correlation of 0.37 was obtained for spring rainfall and spring NINO4. Compared to other states of Australia, e.g. Western Australia, New South Wales and Queensland, this predictability is very low (Murphy and Timbal, 2008;Ummenhofer et al., 2008; Verdon et al., 2004).

The majority of these studies did not consider the effect of lagged climate modes on future seasonal rainfall predictions. According to Schepen et al., (2012) a strong relationship between simultaneous climate modes and rainfall does not essentially mean that there is a lagged relationship as well. Of the few studies focusing on the lagged climate –rainfall relationship one can mention Abbot and Marohasy (2012), Drosdowsky and Chambers (2001), Kirono et al., (2010), and Schepen et al., (2012). Kirono et al., (2010) considered the relationship between Australian rainfall and two months average large scale climate indicators. Abbot and Marohasy (2012) also used past values of climate indices, monthly historical rainfall data and atmospheric temperature for monthly and seasonal forecasting of rainfall in Queensland, Australia; however the climate indices they used were limited to Southern Oscillation Index (SOI), Dipole Mode Index (DMI), Pacific Decadal Oscillation (PDO) and a sea surface temperature based index of ENSO (NINO3.4) lagged by 1~2 months. Schepen et al., (2012) used a Bayesian joint probability modeling approach for seasonal rainfall prediction; however their results were to some extent different from Kirono et al., (2010).

It is already established by many researchers that the changes of sea surface temperature (SST) and sea level pressures (SLP) in Pacific and Indian Ocean which result in the occurrence of ENSO and IOD cycles have enormous effect on the pattern, intensity and the amount of rainfall around the world, but how these changes are related to predicting future rainfall is still not clear. Since these two large modes of climate both contribute to climate patterns and especially rainfall creation, thus, the objective of this study is to investigate the relationship of combined ENSO and IOD lags on Victoria's spring rainfall, as a case study. To achieve this objective two different methods have been investigated; Multiple regression analysis (MR) which is a linear technique and Artificial Neural Networks (ANN) which is a nonlinear method. Three regions in Victoria, each having three rainfall stations is chosen as the case study. Model outputs were aimed to be deterministic forecast as opposed to probabilistic forecast.

2. Methods and data

2.1 Data

Historical monthly rainfall data was obtained from the Australian Bureau of Meteorology website (BOM) (www.bom.gov.au/climate/data/).Three different regions were considered in this study: west Victoria, central Victoria and east Victoria; for each region three stations were selected (Figure.1). The stations were chosen based on their recorded length of data and having fewer missing values. Spring (September - November) rainfalls in millimeters were obtained from monthly rainfall data from January 1900 to December 2009.

El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) were chosen as rainfall drivers based on the previous studies (Kirono et al., 2010; Meneghini et al., 2007; Risbey et al., 2009). ENSO is represented by two different types of indicators: the Southern Oscillation Index (SOI) which is a measure of Sea Level Pressure (SLP) anomalies between Darwin and Tahiti; and the Sea Surface Temperature (SST) anomalies in equatorial Pacific Ocean noted as Nino3 (5^oS – 5^oN, 150^o– 90^oW), Nino3.4 (5^o S – 5^oN, 170^o – 120^oW) and Nino4 (5^oS – 5^oN, 160^o – 150^oW) (Risbey et al., 2009). Nino3.4 and SOI which are the common indices in identifying El Nino/ La Nina years are used as ENSO indicators in this study. IOD is also a coupled ocean-atmosphere phenomenon in the equatorial

Indian Ocean (Saji et al., 1999). A measure of IOD is the Dipole Mode Index (DMI) which is the difference in average SST anomalies between the tropical Western Indian Ocean (10^{o}S - 10^{o}N, 5^{o}O - 70^{o}E) and the tropical Eastern Indian ocean (10^{o}S - Equator, 90^o - 110^{o}E) (Kirono et al., 2010). The climate indices data were obtained from Climate Explorer website (http://climexp.knmi.nl/).

The data were divided in to two sets, from 1900-1990 for calibration and from 1991-2006 for validation of the models. Three years 2007-2009 were selected as the out-of-sample set to evaluate the generalization ability of the developed models. The data were normalized between the range of 1 and 0 using Eq. (1).

$$\bar{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

The models were evaluated using Mean square error (MSE), Mean Absolute Error (MAE), and Pearson correlation (r) which are widely used for prediction purposes; the models were further assessed using Willmott index of agreement *(d)* (Eq. 2)

$$\mathrm{d} = 1 - \left(\frac{[\Sigma|\hat{y}_i - x_i|^2]}{[\Sigma(|\hat{y}_i - \bar{x}_i| + |x_i - \bar{x}_i|)^2]}\right) \tag{2}$$

where, $\hat{y}_i$ is the predicted value of the ith observation and x_i is the ith observation. The closer the *(d)* is to one the better the model has fitted the observations (Willmott 1982)

2.2 Multiple Regression Analysis (MR)

Multiple regression analysis (MR) is a linear statistical technique that allows for finding the best relationship between a variable (dependent, predicant) and several other variables (independent, predictor) through the least square method. Multiple regression models can be presented by the following equation:

$$Y = a + b_1X_1 + b_2X_2 + c \tag{3}$$

where, Y is the dependent variable (spring rainfall), X_1 and X_2 are first and second independent variable respectively (lagged ENSO and IOD indicators), b_1 and b_2 are model coefficients of first and second independent variable respectively, a is constant, and c is the error.

It is important to evaluate the goodness-of-fit and the statistical significance of the estimated parameters of the constructed regression models; the techniques commonly used to verify the goodness-of-fit of regression models are the hypothesis testing, R-squared and analyse of the residuals. For this purpose the F-test is used to verify the statistical significance of the overall fit and the t-test is used to evaluate the significance of the individual parameters; The latter one tests the importance of individual coefficients where the former one is used to compare different models to evaluate the model that best fits the population of the sample data (Um et al., 2011).

Verifying the multicollinearity is also an important stage in MR modeling; Multicollinearity occurs when the predictors are highly correlated which will result in dramatic change in parameter estimates in response to small changes in the data or the model. The indicators used to identify multicollinearity among predictors are tolerance (T) and variance inflation factor (VIF):

$$\text{Tolerance} = 1 - \mathrm{R}^2 \qquad \text{VIF} = \frac{1}{\text{Tolerance}} \tag{4}$$

where, R^2 is the coefficient of multiple determination :

$$\mathrm{R}^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \tag{5}$$

where, SST is the total sum of squares, SSR is the regression sum of squares and SSE is the error sum of squares. According to Lin (2008) a tolerance of less than 0.20–0.10 or a VIF greater than 5–10 indicates a multicollinearity problem.

Analysing the pattern of residuals is another method of evaluating the goodness-of-fit of the models. If any autocorrelation exists among the residuals then the models have not captured all the relationship there is between the inputs and the output; The criterion that can evaluate this is the Durbin-Watson test (DW) which tests for serial correlations between errors. The test statistics have a range of 0 to 4, according to Field (2009) values less than 1 or greater than 3 are definitely matter of concern.

2.3 Artificial Neural Networks (ANN)

Many probabilistic and deterministic modeling approaches have been used by hydrologist and climatologist in order to capture rainfall characteristics. Conceptual and physically based models require an in depth knowledge of this complex atmospheric phenomena; these models need a large amount of calibration data and they have to deal with over parameterisation effect and parameter redundancy impact (De Vos and Reintjas, 2005). Artificial Neural Networks (ANN) is a mathematical model that has the ability to find the nonlinear relationship between input and output parameters without the need to solve complex partial differential equations (Yilmaz et al., 2011). ANN has been used in many hydrological and meteorological applications; It has been used for rainfall-runoff modelling (Akhtar et al., 2009; Chiang and Chang, 2009; Chiang et al., 2004; De Vos and Rientjes, 2005; Sudheer et al., 2002; Tokar and Johnson, 1999) for streamflow forecasting (Campolo et al., 1999; Firat and Gungor, 2007; Kisi, 2007; Riad et al., 2004; Turan and Yurdusev, 2009) and for ground water modelling (Coulibaly et al., 2001; Daliakopoulos et al., 2005; Rogers and Dowla, 1994). It has also been used for many cases of rainfall forecasting (Hsu et al., 1995; Luk et al., 2001; Mekanik et al., 2011; Rami'rez et al., 2005; Toth et al., 2000).

ANN has been inspired by biological neural networks; it consists of simple neurons and connections that process information in order to find a relationship between inputs and outputs. The most common ANN architecture used by hydrologist is the Multilayer Perceptrons (MLP) which is a feedforward network that consists of three layers of neurons, the input layer, the hidden layers and the output layer (Figure 2). The number of input and output neurons is based on the number of input and output data; The input layer only serves as receiving the input data for further processing in the network. The hidden layers are a very important part in an MLP since they provide the nonlinearity between the input and output sets. More complex problems can be solved by increasing the number of hidden layers or the hidden neurons in the hidden layers. The output neuron is the desired output of the model. The process of developing an ANN model is to find a) suitable input data set, b) determine the number of hidden layers and neurons, and c) training, validating and testing the network. Mathematically, the network depicted in Figure. 2 can be expressed as follow:

$$Y_t = f_2\left[\sum_{j=1}^{J} w_j f_1\left(\sum_{i=1}^{I} w_i x_i\right)\right] \quad (6)$$

where, Y_t is the output of the network, x_i is the input to the network, w_i and w_j are the weights between neurons of the input and hidden layer and between hidden layer and output respectively; f_1 and f_2 are the activation functions for the hidden layer and output layer respectively. According to Maier and Dandy (2000) if extrapolating beyond the range of the training data is needed it is recommended to use sigmoiadal-type transfer functions in the hidden layers and linear transfer functions in the output layer. In this study f_1 is considered tansigmoid function which is a nonlinear function and f_2 is considered the linear purelin function defined as follow:

$$f_1 = \frac{2}{(1+\exp(-2x))} - 1 \quad (7)$$

$$f_2(x) = x \quad (8)$$

The ANN models were trained based on Levenberg-Marquardt algorithm; number of hidden neurons was chosen based on constructive algorithm. In ANN modeling there is always the chance of having an over fitted model. To avoid this problem in this study early stop technique is applied while training and validating the models. Through using this method, the network stops the training when the error over the validation set starts to increase while the error over training set is still decreasing; In this way the network avoids over fitting (Luk et al., 2000; Sarle, 1995).

3 Result and Discussion

3.1 Multiple Regression models

In this study the ability of El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) past values to forecast future spring rainfall was analysed using multiple regression analysis (MR). According to Lim et al., (2010) ENSO and IOD have a strong influence in the austral spring on eastern and southern Australia. Cheiw et al., (1998) also suggest that ENSO indicators can be used to some extent to forecast spring rainfall in eastern Australia; They found that the highest correlation between rainfall and climate indicators are obtained using SOI and SST values averaged over two or three months. On the other hand, Verdon et al., (2004) indicate that the influence of ENSO in Victoria (southeast Australia) appears to be weak.

In this study correlation between spring rainfall at year *n* and Dec_{n-1}-Aug_n monthly values of ENSO and IOD indicators (Nino3.4, SOI and DMI) were calculated ("*n"* being the year for which spring rainfall is being predicted); It was discovered that only the three months of June, July and August of Nino34, SOI and DMI have significant correlation with spring rainfall (Table1); This result is in accordance to the findings of Cheiw et al., (1998) and Verdon et al., (2004), substantiating that not only the highest correlations between rainfall and climate indicators are obtained up to three month lags i.e. there is no further significant relationship after lag 3; also these correlations are very weak for Victoria ($|r_{max}|$=0.30 for east Victoria, $|r_{max}|$=0.39 for central Victoria and $|r_{max}|$=0.36 for west Victoria). ENSO-IOD input sets were organized based on these months as potential predictors of spring rainfall for multiple regression analysis (Table2). F-test and t-test was conducted to evaluate the significant level of the models and the regression coefficients; among the constructed models the ones that did not violate the limits of statistical significance was selected, the models with lower error were chosen as the best model for each station. The regression coefficients, variance inflation factor (VIF), Durbin-Watson statistics (DW) and the Pearson correlation (r) of the best models are shown in Table 3. It can be seen from this Table that VIFs for the selected models are near one, i.e. there is no multicolinearity among the predictors; also, the DW statistics is showing that the residuals of the models have no autocorrelation confirming the goodness-of-fit of the models. Nino34-DMI based models proved to be statistically significant and having better forecasting ability than SOI-DMI models for Victoria, with a maximum Pearson r of 0.35 for east Victoria, 0.37 for central Victoria and 0.39 for west Victoria. It is important to note that models with two/three months combined inputs such as $Nino34_{(Jun\text{-}Jul\text{-}Aug)}$-$DMI_{(Jul\text{-}Aug)}$ with different combination of months were also developed, however these high lagged models were all found to be not statistically significant in anyway and were all discarded.

Table 4 shows the MSE, MAE and Pearson correlation (r) of the best MR models for the three regions. It can be seen from Table 4 that the errors are relatively low for all the stations.

3.2 ANN models

As discussed in Section 3.1, the strongest, statistically significant relationship between spring rainfall and climate indicators occur in the months of June, July and August (Table 1); due to the statistical limits of MR analysis the three months of June, July and August of ENSO and IOD could not be incorporated together in one single model and had to be separated in order to obtain statistically significant models. ANN is free from these assumptions and thus it is capable of taking ENSO $_{(Jun\text{-}July\text{-}Aug)}$-IOD $_{(Jun\text{-}July\text{-}Aug)}$ sets as inputs for predicting future rainfall. Two sets of Nino3.4 $_{(Jun\text{-}July\text{-}Aug)}$-DMI $_{(Jun\text{-}July\text{-}Aug)}$ and SOI $_{(Jun\text{-}July\text{-}Aug)}$-DMI $_{(Jun\text{-}July\text{-}Aug)}$ were used as inputs for developing ANN models for the three regions. Table 5 summarises the prediction skills of these models regarding MSE, MAE and Pearson correlation (r); it can be seen from Table 5 that the correlation coefficients of ANN models for east and west Victoria is significantly higher compared to the MR models (Table 4) and the errors (MAE and MSE) are generally lower. Also for central Victoria, the correlation coefficients are generally higher than the MR models; however the performance of the MR models regarding MSE and MAE is better for this region. The higher correlation coefficient of ANN models indicate that ANN is more capable of finding the pattern and trend of the observations compared to MR models.

After calibrating and validating the models, in order to evaluate the generalization ability of the developed MR and ANN models, out-of-sample tests were carried out on the years 2007-2009 (Table 6). It can be seen that MR models are showing very poor generalization ability for east Victoria, (r = -0.99, -0.90 and -0.99 for Bruthen, Buchan and Orbost respectively) compared to ANN with correlation coefficients of 0.93, 0.76 and 0. 42; ANN models also showed better generalization ability for central and west Victoria with correlation coefficients of 0.68~0.85 and 0.58~0.97 respectively compared to MR models, however the ability of MR models to forecast out-of-sample sets is compatible with ANN for Daylesford in central Victoria and Kaniva in west Victoria (r=0.92 and 0.67 respectively). Also the errors of the testing sets for ANN models are generally lower compared to multiple regression models.

Figures 3 to 5 show comparisons between multiple regression and ANN models for the 9 stations. In general regression models are showing an underestimation of the actual observations compared to ANN models. While Pearson correlation shows how well the models are following the trend of the actual observations, Willmott index of agreement "*d*" shows how well the models are fitting the observations. This index is tabulated in Table 7 for multiple regression and ANN models; the closer the value of "*d*" is to one the better is the model accuracy. It can be seen from this table that ANN models are having higher "*d*" values compared to MR models.

4. Conclusion

This study focused on investigating the use of combined lagged El Nino Southern Oscillation (ENSO) and Indian Ocean dipole (IOD) as potential predictors of spring rainfall. Multiple regression (MR) and Artificial Neural Network (ANN) approach was used for this purpose. Three regions (east, centre and west) of Victoria were chosen as case study each having three rainfall stations. Nino3.4 and Southern Oscillation Index (SOI) were used as ENSO indicators and Dipole Mode Index (DMI) was chosen as IOD indicator.

The Pearson correlation coefficients of past values of the climate indices with spring rainfalls for the 9 stations were calculated; It was discovered that only the three months of June, July and August of Nino34, SOI and DMI have significant correlation with spring rainfall and these correlations are very weak. Nino3.4-DMI and SOI-DMI input sets were organized based on these months as potential predictors of spring rainfall for MR analysis. Among the several developed models the ones that did not violate the limits of statistical significance and multicollinearity and had lower model error were used for prediction purposes.

ANN modelling was also conducted for the 9 stations of Victoria using the combined lagged Nino3.4-DMI and SOI-DMI. Multilayer Perceptron (MLP) architecture was chosen for this purpose due to its wide use in hydrologic modellings. The models were trained based on Levenberg-Marquardt algorithm. ANN models showed higher correlation compared to MR models indicating that ANN is more capable of finding the pattern and trend of the observations compared to MR models. Also, ANN models generally showed lower errors and are more reliable for prediction purposes.

After calibrating and validating the models they were tested on out-of-sample sets. It was found that generalization ability of MR models for east Victoria is very poor compared to the other two regions and also compared to ANN. ANN was able to perform out of sample test with correlation coefficient of 0.42~0.93 for east Victoria, 0.68~0.85 for central Victoria and 0.58~0.97 for west Victoria. Multiple regression models were compatible with ANN in two stations of Daylesford and Kaniva in central and west Victoria with correlation coefficient of 0.92 and 0.67 respectively. Although the effect of ENSO and IOD in Victoria is quite weak, however with the use of combined lagged ENSO-IOD sets in nonlinear ANN and linear multiple regression analysis, long term rainfall forecast can be achieved.

5. 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.

Akhtar, M., Corzo, G., Van Andel, S., & Jonoski, A. (2009). River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin. *Hydrology and Earth System Sciences, 13*(9), 1607.

Barsugli, J. J., & Sardeshmukh, P. D. (2002). Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. *Journal of Climate, 15*(23), 3427-3442.

Campolo, M., Andreussi, P., & Soldati, A. (1999). River flood forecasting with a neural network model. *Water Resources Research, 35*(4), 1191-1197.

Chattopadhyay, G., Chattopadhyay, S., & Jain, R. (2010). Multivariate forecast of winter monsoon rainfall in India using SST anomaly as a predictor: Neurocomputing and statistical approaches. *Comptes Rendus Geoscience, 342*(10), 755-765.

Chiang, Y. M., & Chang, F. J. (2009). Integrating hydrometeorological information for rainfall-runoff modelling by artificial neural networks. *Hydrological Processes, 23*(11), 1650-1659.

Chiang, Y. M., Chang, L. C., & Chang, F. J. (2004). Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modeling. *Journal of Hydrology, 290*(3), 297-311.

Chiew, F., Piechota, T. C., Dracup, J., & McMahon, T. (1998). El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology, 204*(1), 138-149.

Coulibaly, P., Anctil, F., Aravena, R., & Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations. *Water Resources Research, 37*(4), 885-896.

Daliakopoulos, I. N., Coulibaly, P., & Tsanis, I. K. (2005). Groundwater level forecasting using artificial neural networks. *Journal of Hydrology, 309*(1), 229-240.

De Vos, N., & Rientjes, T. (2005). Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation. *Hydrology and Earth System Sciences Discussions Discussions, 2*(1), 365-415.

Drosdowsky, W., & Chmabers, L. E. (2001). Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *Journal of Climate, 14*(7), 1677-1687.

Field, A. P. (2009). *Discovering statistics using SPSS*: SAGE publications Ltd.

Firat, M., & Güngör, M. (2007). River flow estimation using adaptive neuro fuzzy inference system. *Mathematics and Computers in Simulation, 75*(3-4), 87-96.

Hartmann, H., Becker, S., & King, L. (2008). Predicting summer rainfall in the Yangtze River basin with neural networks. *International Journal of Climatology, 28*(7), 925-936.

Hsu, K., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research, 31*(10), 2517-2530.

Kirono, D. G. C., Chiew, F. H. S., & Kent, D. M. (2010). Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological Processes, 24*(10), 1237-1247.

Kişi, Ö. (2007). Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering, 12*, 532.

Lau, K., & Weng, H. (2001). Coherent modes of global SST and summer rainfall over China: An assessment of the regional impacts of the 1997-98 El Nino. *Journal of Climate, 14*(6), 1294-1308.

Lim, E. P., Hendon, H. H., Anderson, D. L. T., Charles, A., & Alves, O. (2011). Dynamical, statistical-dynamical, and multimodel ensemble forecasts of Australian spring season rainfall. *Monthly Weather Review, 139*(3), 958-975.

Lin, F. J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality & Quantity, 42*(3), 417-426.

Luk, K., Ball, J., & Sharma, A. (2000). A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology, 227*(1), 56-65.

Luk, K. C., Ball, J., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer modelling, 33*(6), 683-693.

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and software, 15*(1), 101-124.

Mekanik, F., Lee, T. S., & Imteaz, M. A. (2011). Rainfall modeling using Artificial Neural Network for a mountainous region in west Iran. *19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty, MODSIM2011,* 3518-3524.

Meneghini, B., Simmonds, I., & Smith, I. N. (2007). Association between Australian rainfall and the Southern Annular Mode. *International Journal of Climatology, 27*(1), 109-121.

Murphy, B. F., & Timbal, B. (2008). A review of recent climate variability and climate change in southeastern Australia. *International Journal of Climatology, 28*(7), 859-879.

Ramírez, M. C. P., Velho, H. F. C., & Ferreira, N. J. (2005). Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *Journal of Hydrology, 301*(1), 146-162.

Riad, S., Mania, J., Bouchaou, L., & Najjar, Y. (2004). Predicting catchment flow in a semi-arid region via an artificial neural network technique. *Hydrological Processes, 18*(13), 2387-2393.

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.

Rogers, L. L., & Dowla, F. U. (1994). Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resources Research, 30*(2), 457-481.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian ocean. *Nature, 401*(6751), 360-363.

Sarle, W. S. (1995). *Stopped training and other remedies for overfitting.* Paper presented at the Proceedings of the 27th Symposium on the Interface.

Schepen, A., Wang, Q. J., & Robertson, D. (2012). Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *Journal of Climate, 25*(4), 1230-1246.

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*(1-2), 99-109.

Sudheer, K., Gosain, A., & Ramasastri, K. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes, 16*(6), 1325-1330.

Tokar, A. S., & Johnson, P. A. (1999). Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering, 4*(3), 232-239.

Toth, E., Brath, A., & Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology, 239*(1), 132-147.

Turan, M. E., & Yurdusev, M. A. (2009). River flow estimation from upstream flow records by artificial intelligence methods. *Journal of Hydrology, 369*(1-2), 71-77.

Um, M. J., Yun, H., Jeong, C. S., & Heo, J. H. (2011). Factor analysis and multiple regression between topography and precipitation on Jeju Island, Korea. *Journal of Hydrology, 410*(3-4), 189-203.

Ummenhofer, C. C., Sen Gupta, A., Pook, M. J., & England, M. H. (2008). Anomalous rainfall over southwest Western Australia forced by Indian Ocean sea surface temperatures. *Journal of Climate, 21*(19), 5113-5134.

Verdon, D. C., Wyatt, A. M., Kiem, A. S., & Franks, S. W. (2004). Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resources Research, 40*(10), W10201.

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society, 63*, 1309-1369.

Yilmaz, A. G., Imteaz, M. A., & Jenkins, G. (2011). Catchment flow estimation using Artificial Neural Networks in the mountainous Euphrates Basin. *Journal of Hydrology, 410*(1-2), 134-140.

Yufu, G., Yan, Z., & Jia, W. (2002). Numerical simulation of the relationships between the 1998 Yangtze River valley floods and SST anomalies. *Advances in Atmospheric Sciences, 19*(3), 391-404.

Table

Table1. Pearson correlation (r) of lagged climate indices and spring rainfall

Region	Station	Lagged climate indices								
		Nino34$_{(Jun)}$	Nino34$_{(Jul)}$	Nino34$_{(Aug)}$	SOI$_{(Jun)}$	SOI$_{(Jul)}$	SOI$_{(Aug)}$	DMI$_{(Jun)}$	DMI$_{(Jul)}$	DMI$_{(Aug)}$
East	Bruthen	-0.20[b]	-0.25[a]	-0.28[a]	---	---	---	-0.25[a]	---	---
	Buchan	-0.22[b]	-0.26[a]	-0.24[b]	-0.20[b]	---	---	-0.30[a]	---	---
	Orbost	---	-0.24[b]	-0.26[a]	---	---	---	-0.29[a]	-0.21[b]	---
Centre	Malmsbury	-0.22[b]	-0.22[b]	-0.29[a]	---	0.32[a]	0.30[a]	---	-0.30[a]	-0.31[a]
	Daylesford	-0.30[a]	-0.28[a]	-0.33[a]	0.20[b]	0.37[a]	0.34[a]	---	-0.29[a]	-0.28[a]
	Heathcote	-0.30[a]	-0.30[a]	-0.38[a]	---	0.36[a]	0.39[a]	---	-0.25[a]	-0.28[a]
West	Horsham	-0.22[b]	-0.23[b]	-0.31[a]	----	0.26[a]	0.25[a]	---	-0.28[a]	-0.31[a]
	Kaniva	-0.32[a]	-0.32[a]	-0.36[a]	0.23[b]	0.33[a]	0.31[a]	---	-0.30[a]	-0.31[a]
	Rainbow	-0.31[a]	-0.31[a]	-0.36[a]	0.20[b]	0.33[a]	0.33[a]	---	-0.25[a]	-0.26[a]

a: correlation is significant at the 0.01% level

b: correlation is significant at the 0.05% level

Table

Table2. Multiple regression model sets developed for each station

	Nino3.4-DMI	SOI-DMI
Bruthen	Jun-Jun, Jul-Jun, Aug-Jun	--------
Buchan	Jun-Jun, Jul-Jun, Aug-Jun	Jun-Jun
Orbost	Jul-Jun, Jul-Jul, Aug-Jun, Aug-Jul,	--------
Malmsbury	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug	Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug
Daylesford	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug
Heathcote	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug	Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug
Horsham	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug	Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug
Kaniva	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug
Rainbow	Jun-Jul, Jun-Aug, Jul-Jul, Jul-Aug, Aug-Jul, Aug-Aug	Jun-Jul, Jun-Aug, Jul-Jul, Jun-Aug, Aug-Jul, Aug-Aug

Table

Table3. Summary of the best regression models

Region	Station	Models	Coefficient										R	VIF	DW
			Const.	Nino34$_{(Jun)}$	Nino34$_{(Jul)}$	Nino34$_{(Aug)}$	SOI$_{(Jun)}$	SOI$_{(Jul)}$	SOI$_{(Aug)}$	DMI$_{(Jun)}$	DMI$_{(Jul)}$	DMI$_{(Aug)}$			
East	Bruthen	Ni34$_{(Jul)}$-DMI$_{(Jun)}$	0.65	---	-0.24	---	---	---	---	-0.24	---	---	0.32	1.10	1.90
	Buchan	Ni34$_{(Jul)}$-DMI$_{(Jun)}$	0.51	---	-0.17	---	---	---	---	-0.23	---	---	0.35	1.10	2.10
	Orbost	Ni34$_{(Aug)}$-DMI$_{(Jun)}$	0.56	---	---	-0.20	---	---	---	-0.27	---	---	0.35	1.10	2.00
Centre	Malmsbury	Ni34$_{(Aug)}$-DMI$_{(Jul)}$	0.55	---	---	-0.20	---	---	---	---	-0.22	---	0.36	1.12	1.90
	Daylesford	Ni34$_{(Jun)}$-DMI$_{(Jul)}$	0.62	-0.25	---	---	---	---	---	---	-0.29	---	0.37	1.10	1.81
	Heathcote	Ni34$_{(Jun)}$-DMI$_{(Aug)}$	0.60	-0.29	---	---	---	---	---	---	---	-0.24	0.37	1.10	1.80
West	Horsham	Ni34$_{(Aug)}$-DMI$_{(Jul)}$	0.55	---	---	-0.24	---	---	---	---	-0.20	---	0.36	1.12	2.00
	Kaniva	Ni34$_{(Jun)}$-DMI$_{(Jul)}$	0.67	-0.32	---	---	---	---	---	---	-0.27	---	0.39	1.10	2.00
	Rainbow	Ni34$_{(Aug)}$-DMI$_{(Jun)}$	0.56	-0.29	---	---	---	---	---	---	---	-0.20	0.36	1.10	2.25

Table

Table 4. Performance of the regression models

Region	Station	R	MSE	MAE
East	Bruthen	0.32	0.048	0.171
	Buchan	0.35	0.026	0.171
	Orbost	0.35	0.038	0.157
Centre	Malmsbury	0.36	0.030	0.140
	Daylesford	0.37	0.039	0.155
	Heathcote	0.37	0.035	0.153
West	Horsham	0.36	0.033	0.149
	Kaniva	0.39	0.041	0.163
	Rainbow	0.36	0.031	0.142

Table

Table 5. Performance of ANN models

Region	Station	Model	R	MSE	MAE
East	Bruthen	Ni34-DMI	0.75	0.023	0.120
	Buchan	Ni34-DMI	0.65	0.028	0.154
	Orbost	SOI-DMI	0.64	0.034	0.145
Centre	Malmsbury	Ni34-DMI	0.54	0.034	0.130
	Daylesford	Ni34-DMI	0.36	0.039	0.168
	Heathcote	SOI-DMI	0.52	0.044	0.158
West	Horsham	Ni34-DMI	0.64	0.023	0.193
	Kaniva	SOI-DMI	0.56	0.042	0.158
	Rainbow	SOI-DMI	0.53	0.023	0.115

Table

Table 6. Performance of ANN and multiple regression models for the out-of-sample test set

Region	Station	ANN			Regression		
		R	MSE	MAE	R	MSE	MAE
East	Bruthen	0.93	0.018	0.120	-0.99	0.016	0.085
	Buchan	0.76	0.008	0.080	-0.90	0.023	0.180
	Orbost	0.42	0.015	0.107	-0.99	0.024	0.150
Centre	Malmsbury	0.68	0.007	0.080	0.48	0.013	0.100
	Daylesford	0.85	0.033	0.164	0.92	0.043	0.205
	Heathcote	0.71	0.018	0.125	-0.50	0.026	0.158
West	Horsham	0.80	0.009	0.080	0.25	0.030	0.149
	Kaniva	0.97	0.013	0.110	0.67	0.051	0.163
	Rainbow	0.58	0.017	0.128	-0.74	0.029	0.142

Table

Table 7. Index of agreement (d) for the out-of-sample test set

Station	Regression	ANN
Bruthen	0.26	0.47
Buchan	0.30	0.78
Orbost	0.00	0.62
Malmsbury	0.50	0.56
Daylesford	0.40	0.68
Heathcote	0.43	0.54
Horsham	0.41	0.89
Kaniva	0.44	0.82
Rainbow	0.33	0.41

Figure

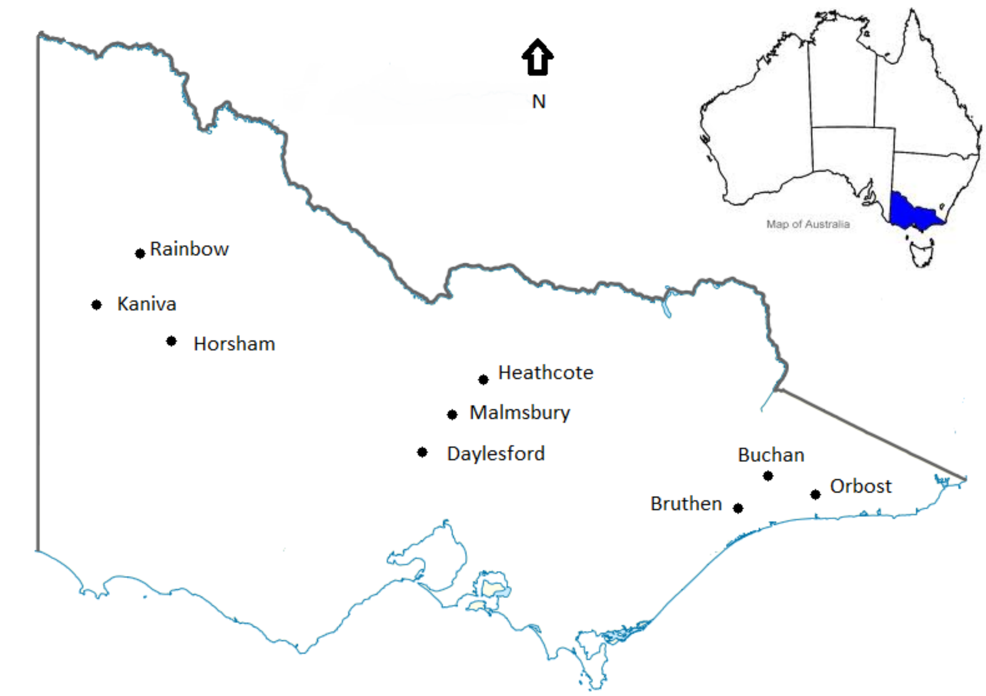

Figure 1. Map of the study area

Figure

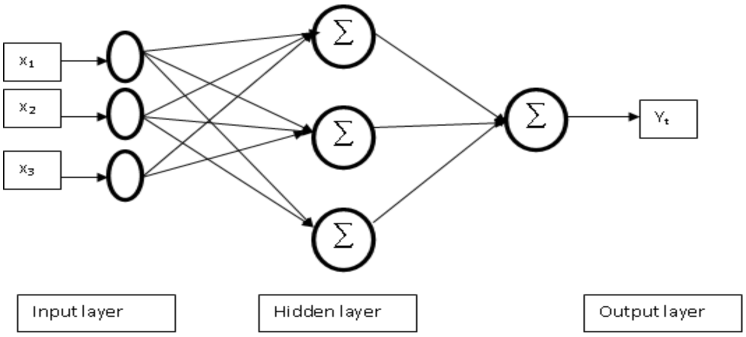

Figure 2. A typical ANN architecture

Figure

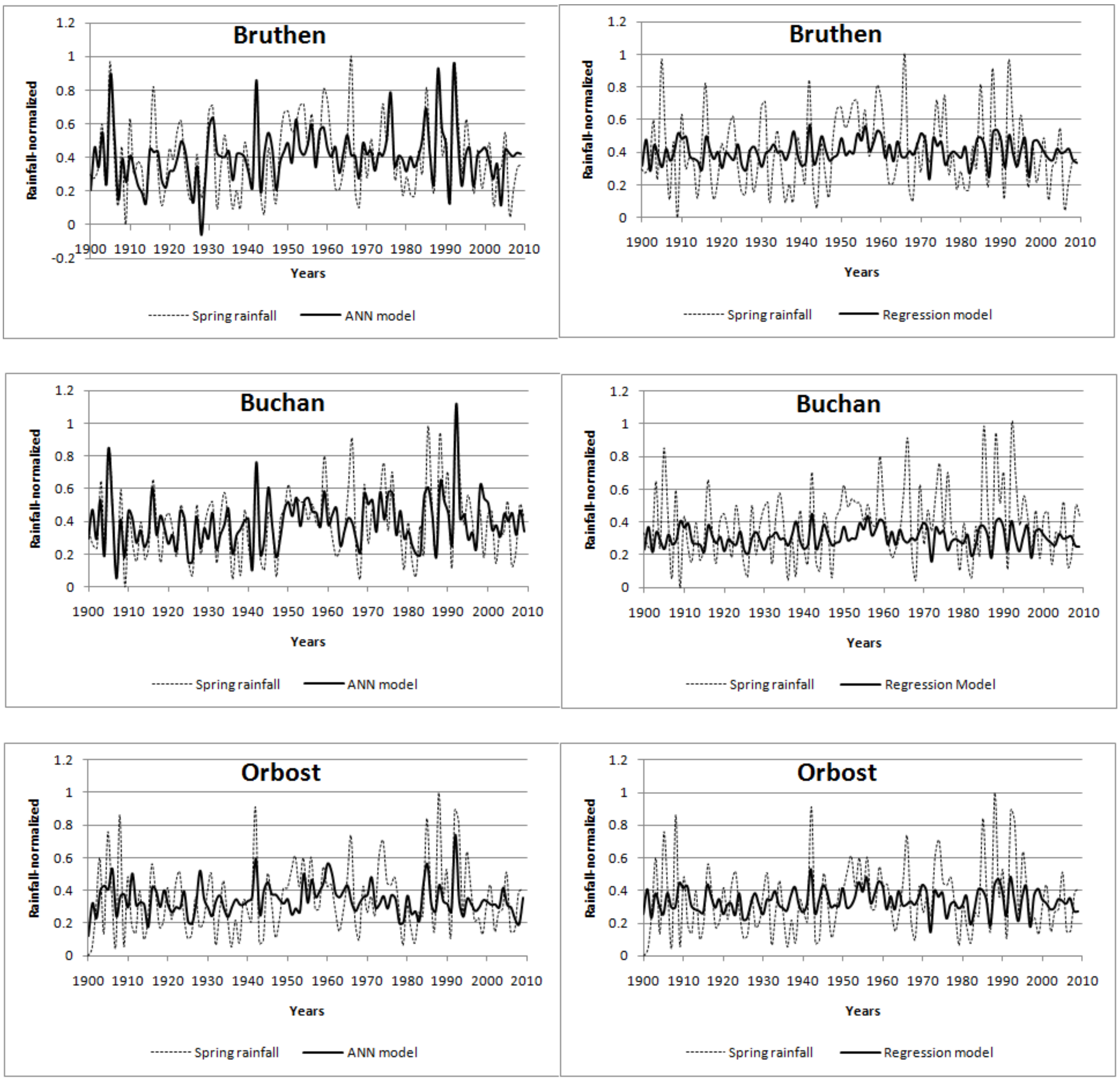

Figure 3. Comparing ANN modelling with MR modelling for east Victoria

Figure

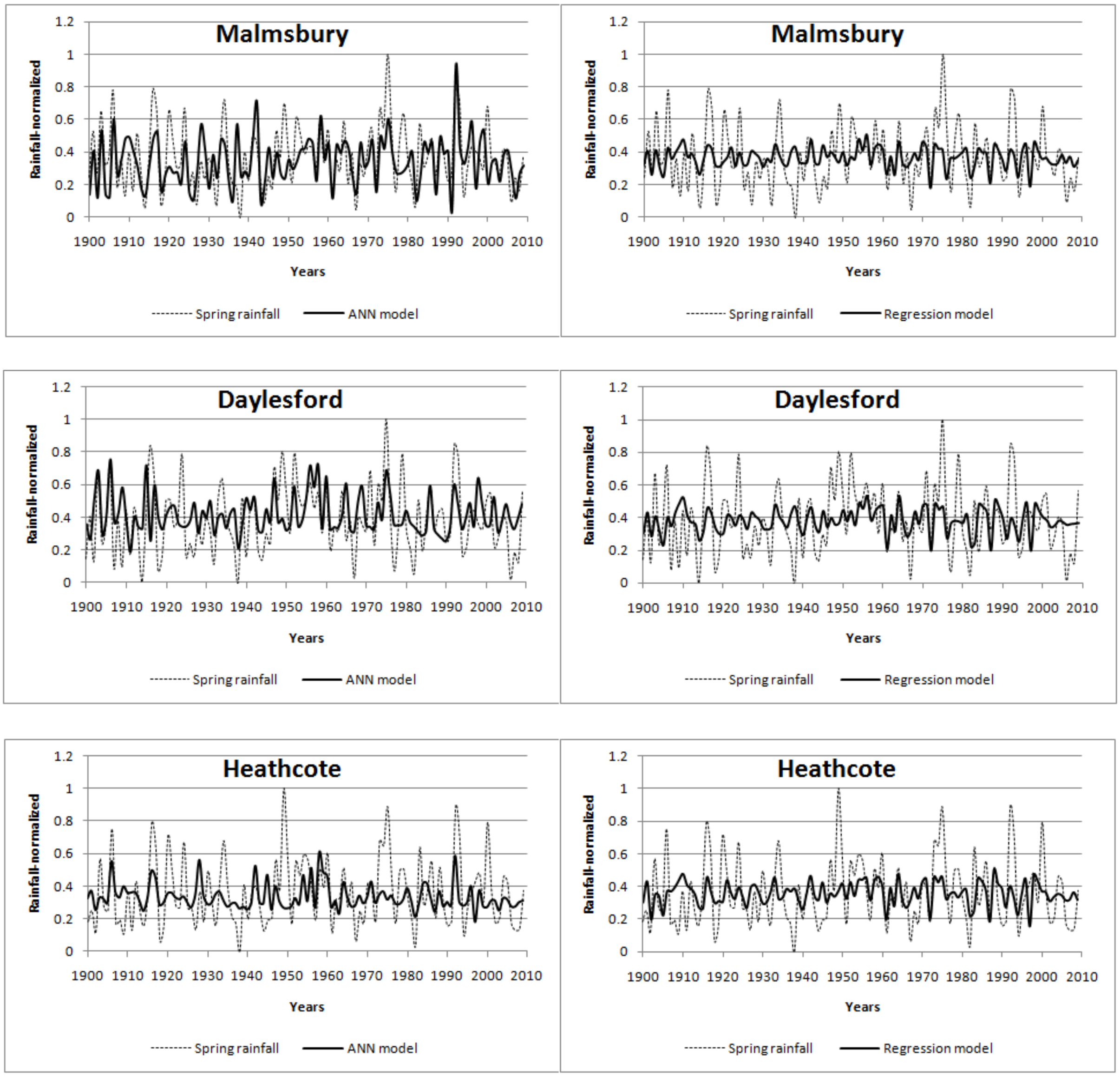

Figure 4. Comparing ANN modelling with MR modelling for central Victoria

Figure

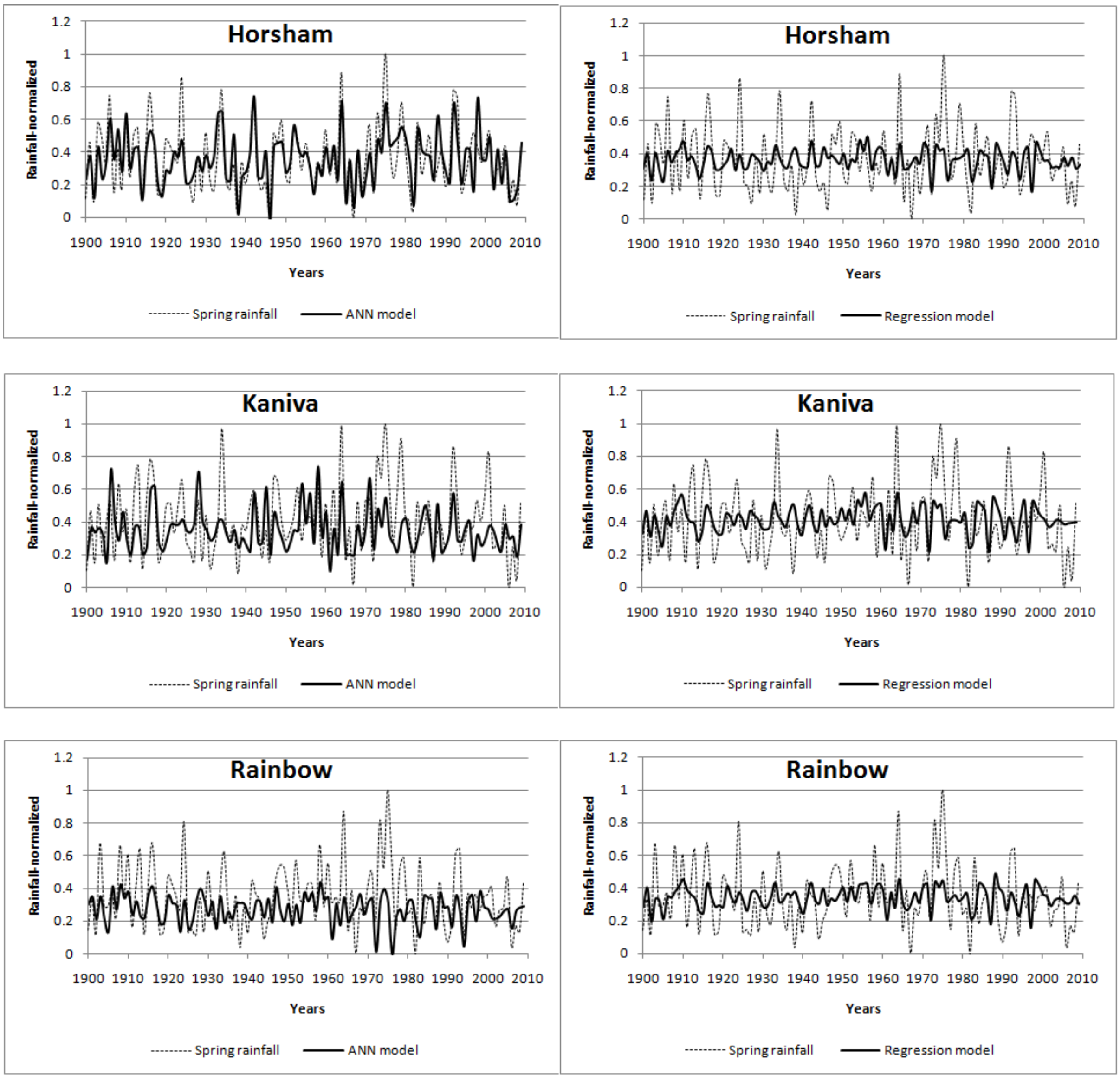

Figure 5. Comparing ANN modelling with MR modelling for west Victoria