Rainfall Prediction in Australia: Clusterwise Linear Regression Approach

Arshad Mahmood

This thesis is submitted in total fulfilment of the requirement
for the degree of Doctor of Philosophy

Faculty of Science and Technology
Federation University Australia
PO Box 663
University Drive, Mount Helen
Ballarat, Victoria, Australia 3353

Submitted in February 2017

Abstract

Accurate rainfall prediction is a challenging task because of the complex physical processes involved. This complexity is compounded in Australia as the climate can be highly variable. Accurate rainfall prediction is immensely beneficial for making informed policy, planning and management decisions, and can assist with the most sustainable operation of water resource systems.

Short-term prediction of rainfall is provided by meteorological services; however, the intermediate to long-term prediction of rainfall remains challenging and contains much uncertainty. Many prediction approaches have been proposed in the literature, including statistical and computational intelligence approaches. However, finding a method to model the complex physical process of rainfall, especially in Australia where the climate is highly variable, is still a major challenge.

The aims of this study are to: (a) develop an optimization based clusterwise linear regression method, (b) develop new prediction methods based on clusterwise linear regression, (c) assess the influence of geographic regions on the performance of prediction models in predicting monthly and weekly rainfall in Australia, (d) determine the combined influence of meteorological variables on rainfall prediction in Australia, and (e) carry out a comparative analysis of new and existing prediction techniques using Australian rainfall data.

In this study, rainfall data with five input meteorological variables from 24 geographically diverse weather stations in Australia, over the period January 1970 to December 2014, have been taken from the Scientific Information for Land Owners (SILO). We also consider the climate zones when selecting weather stations, because Australia experiences a variety of climates due to its size. The data was divided into training and testing periods for evaluation purposes.

In this study, optimization based clusterwise linear regression is modified and new prediction methods are developed for rainfall prediction. The proposed method is applied to predict monthly and weekly rainfall. The prediction performance of the clusterwise linear regression method was evaluated by comparing observed and predicted rainfall values using the performance measures: root mean squared error, the mean absolute error, the mean absolute scaled error and the Nash-Sutcliffe coefficient of efficiency. The proposed method is also compared with the clusterwise linear regres-

sion based on the maximum likelihood estimation, linear support vector machines for regression, support vector machines for regression with radial basis kernel function, multiple linear regression, artificial neural networks with and without hidden layer and k-nearest neighbors methods using computational results.

Initially, to determine the appropriate input variables to be used in the investigation, we assessed all combinations of meteorological variables. The results confirm that single meteorological variables alone are unable to predict rainfall accurately. The prediction performance of all selected models was improved by adding the input variables in most locations.

To assess the influence of geographic regions on the performance of prediction models and to compare the prediction performance of models, we trained models with the best combination of input variables and predicted monthly and weekly rainfall over the test periods. The results of this analysis confirm that the prediction performance of all selected models varied considerably with geographic regions for both weekly and monthly rainfall predictions. It is found that models have the lowest prediction error in the desert climate zone and highest in subtropical and tropical zones.

The results also demonstrate that the proposed algorithm is capable of finding the patterns and trends of the observations for monthly and weekly rainfall predictions in all geographic regions. In desert, tropical and subtropical climate zones, the proposed method outperform other methods in most locations for both monthly and weekly rainfall predictions. In temperate and grassland zones the prediction performance of the proposed model is better in some locations while in the remaining locations it is slightly lower than the other models.

DECLARATION

I, Arshad Mahmood, declare that the PhD thesis entitled “Rainfall Prediction in Australia: Clusterwise Linear Regression Approach” contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the Federation University Australia Library, being made available for loan and photocopying, subject to the provision of the copyright Act 1968. I also give permission for the digital version of my thesis to be made available on the web, via the University Library catalogue, the Australian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the Federation University Australia to restrict access for a period of time.

Signature..

February 2017

Arshad Mahmood

Acknowledgement

I would like to acknowledge my great appreciation to my principal supervisor associate professor Adil Bagirov for providing me with the necessary guidance throughout my PhD research. The valuable comments and suggestions provided by him were immensely helpful to me in successfully writing the journal and conference papers on which this thesis is based. It has been an honor and pleasure working with him.

Thanks to my associate supervisors Dr. Andrew Barton and Dr. Shamsul Huda for their guidance and helpful discussions. In addition to the invaluable support of all my supervisors, special thanks to associate professor Brad Mitchell for keeping me off the streets.

Thank also to the Faculty of Science and Technology and Research Services for providing financial assistance throughout my candidature.

Finally, I would like to thank my family members, friends and other staff members of the Federation University Australia who helped me in many different ways during my PhD.

Dedicated

To

My parents and my family

Table of Contents

List of Tables

List of Figures

List of Publications

Journal papers

1. Bagirov A., Mahmood A. and Barton A, *Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach*, Atmospheric Research, Vol. 188, 2017, 20-29.

2. Bagirov A. and Mahmood A., *A comparative assessment of models to predict monthly rainfall in Australa*, Water Resources Management (Submitted).

3. Bagirov A. and Mahmood A., *Prediction models based on clusterwise linear regression and applications to rainfall forecast*, Advances in Atmospheric Sciences (Submitted).

Conference Presentations

1. Mahmood A., Bagirov A. and Barton A, Performance of clusterwise linear regression model for rainfall prediction in Australia. 36th International Symposium on Forecasting, Satander, 19-22 June 2016.

2. Mahmood A. and Bagirov A., A comparative assessment of prediction models to predict monthly rainfall in Australia. Federation University Australia Conference 2016, Ballarat, 21 July 2016.

3. Mahmood A. and Bagirov A., A comparison of ARIMA model with TDNN model and k-NN mothod to forecast monthly rainfall. Young Statisticians Conference, Adelaide, Australia, 5-6 February 2015.

Chapter 1

Introduction

In this chapter we discuss the importance of water, availability of water in Australia, the influence of meteorological parameters on rainfall prediction and the importance of accurate rainfall prediction. Following that, we will present the objectives of this thesis and provide its outlines.

1.1 Water

Water is the most widely used substance on the earth. Water, which is quite limited, is critical for sustaining life. The survival of life on this planet mainly depends on water. It plays a major role in the environment, economics and social development. Humans need water for household functions, agriculture, industries, irrigation, livestock watering and as an energy resource. In short, the absence of water means no survival.

Increased urbanisation, improved living standards, expanding irrigation and industrial water use have increased the demand of water. With seemingly ever-increasing demand for water, it is becoming more important for the management of water resources to be as efficient as possible to ensure reliable water supplies. The uncertainty about future water availability due to climate change will be a significant challenge for the management of water resources [21].

Water is a renewable resource. It is constantly transferred between the ocean, the atmosphere and the land (hydrologic cycle). The hydrologic cycle begins with the evaporation of water from the ocean and rivers. The lifted water vapors condense to form clouds which then return to the surface as precipitation. After precipitation some of the water evaporates back into the atmosphere while some may penetrate the surface and become ground water. Ground water either seeps into the oceans, rivers and streams or transpires back into the atmosphere. Water resources can be categorized into surface water and ground water. Surface water is water found in streams, rivers, lakes and wetlands and ground water is water contained underground, in geological formations known as aquifers.

Globally, only 3% of the world's water is fresh, but 2.5% of this is frozen. Humans rely on 0.5% of the world's water for their needs [32]. Fresh water distribution varies significantly among, and within, countries. For example, only Brazil, Russia, China, Canada, Indonesia, U.S., India, Columbia and the Democratic Republic of Congo possess 60% of the world's available fresh water supply [32].

Australia is the driest populated continent on the earth, with 70% of its area classified as desert or semi-desert. As an island continent, water supply mainly depends on rainfall. Average annual rainfall in Australia is 465.2 mm (1961-1990), which is the lowest of all the continents (except Antarctica) [68]. 2015 was the 57th driest year on record since 1900. The average rainfall for that year was 446.65 mm, 5% below the 1961-1990 average. Figure 1.1, published by the World Business Council for Sustainable Development in 2005, indicates that water supply in Australia is the lowest per person per year [32]. Despite this, Australia has one of the highest per capita water consumption rates in the world [6]. The agriculture industry consumes the largest volume of water, making up 62% of all water consumed in 2013-14 [69].

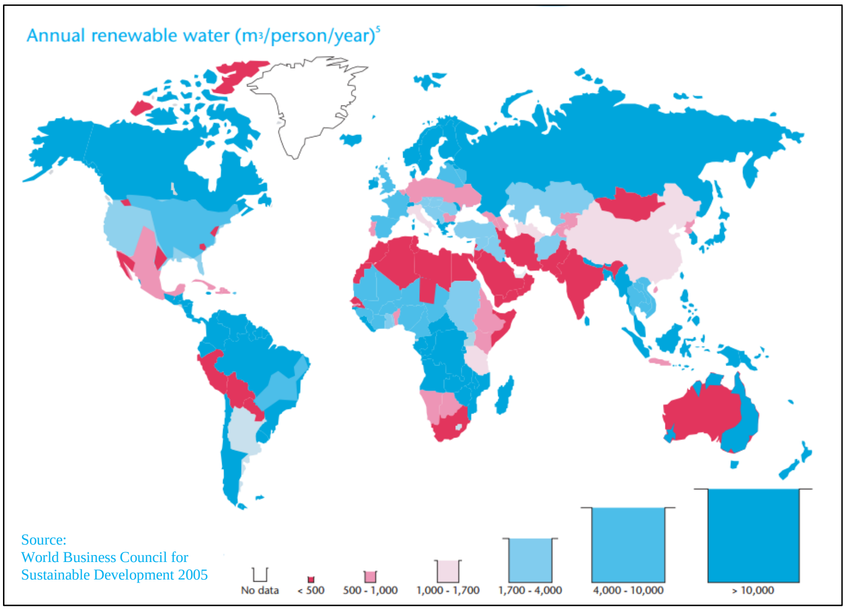

Figure 1.1: Annual renewable water per person per year.

1.2 Rainfall

Australia is the driest inhabited continent on earth with a highly variable climate over time and space. It can be characterized by highly variable rainfall patterns between regions, seasons and years, including extended periods of drought [67]. Australia is an island continent that experiences different rainfall patterns in the various regions. In Australia, there are six climate zones: equatorial, tropical, sub-tropical, temperate, grassland and desert (Figure 1.2) and two main seasonal patterns: Summer/Autumn/Winter/Spring pattern and wet/dry pattern. The Summer/Autumn/Winter/Spring pattern affects the temperate, grassland and desert zones and wet/dry pattern affects the tropical, sub-tropical and equatorial zones. Australia's climate zones range from high rainfall in the tropical regions in the north through to the driest desert region in the interior (Figure 1.3). It is common in Australia for one region to be in drought while another is in flood. Figure 1.4 displays a time series plot of the total annual rainfall in Australia from 1900 to 2014, indicating a slightly increasing trend. The increase in average annual rainfall is due to increases in rainfall across parts of north-west Australia since 1970 [63].

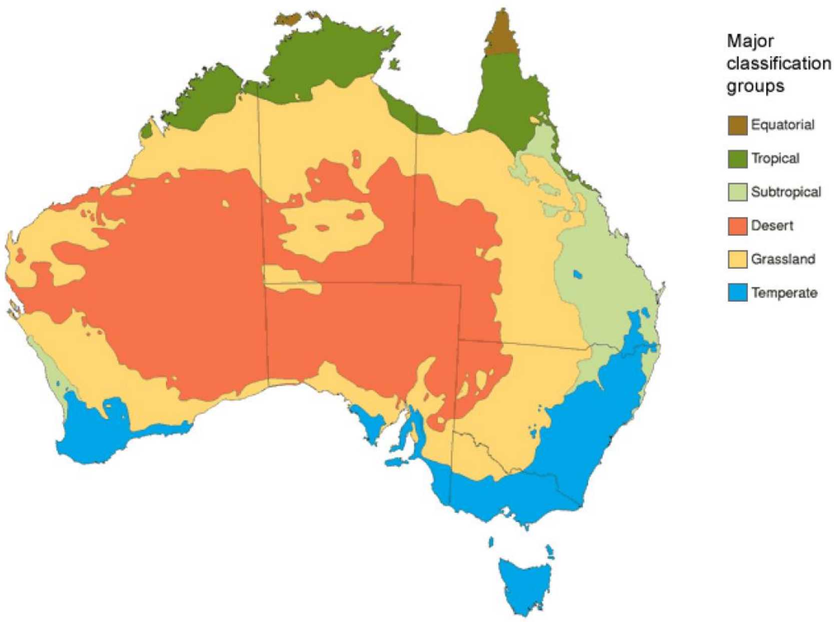

Figure 1.2: Australian major climate zones (Source: Australian Bureau of Meteorology).

The Intergovernmental Panel on Climate Change (IPCC) reported an increase in temperature, a decrease in precipitation and an increase in the number of heavy rainfall events in some regions (Figure 1.5). Since 1950, there has been a major change

Figure 1.3: Average annual Australian rainfall based on 30 years climatology (1961-1990).

in rainfall patterns with significant geographic variations [37]. The average winter rainfall has dropped 17% in south-west Australia since 1970, late Autumn and early Winter rainfall has dropped 15% since 1990, and overall the number and intensity of extreme rainfall events is projected to increase for most regions [64]. Similarly, the duration, frequency and intensity of heat waves have increased across many parts of Australia. The Australian Export Grains Innovation Center analyzed the rainfall data of more than 8000 Bureau of Meteorology weather stations across the country and concluded that the rainfall zones had shifted from between 100 to 400km since 2000, Summer rainfall has increased, and correspondingly Winter rainfall has decreased. These changes can be seen very clearly in the map released by the Australian Export Grains Innovation Center (see Figure 1.6).

These changes in rainfall and temperature pose a series of risks to water availability and water management systems. For example, increases in extreme rainfall events may increase the availability of fresh water in some areas, but at the same time it might come in the form of storms, leading to flooding and damage, and resulting in more harm than good. Rising temperatures could increase the rate of evaporation from surface water and reservoirs, leading to the loss of fresh water.

Any change in the probability of rainfall (heavy rainfall or drought) has important implications for future resource planning, management and investment. Increases in heavy rainfall events may increase the frequency of flood events and landslides, and

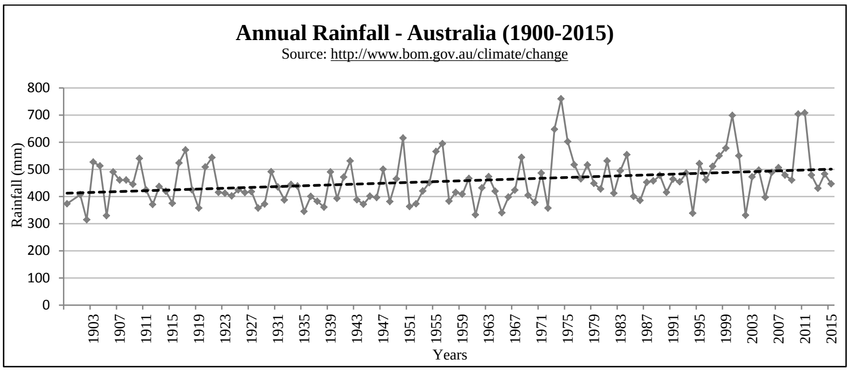

Figure 1.4: Average annual Australian rainfall from 1900 to 2015.

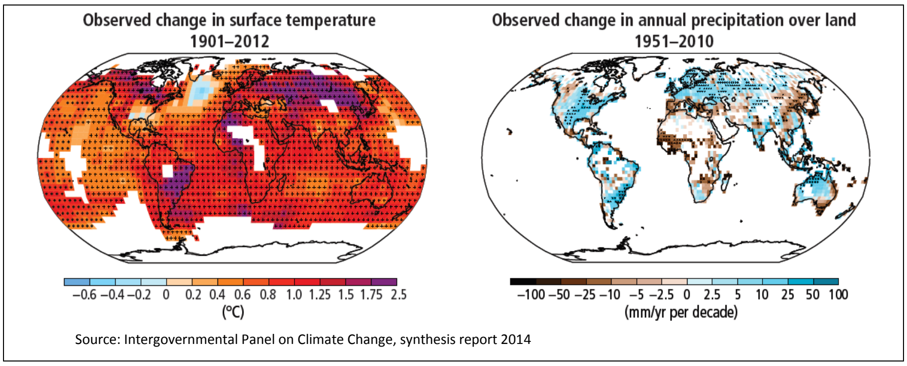

Figure 1.5: Observed changes in surface temperature and precipitation.

build up sediment in dams. Similarly, abnormally dry periods lead to drought.

A drought is a prolonged dry period when there is not enough water for the common needs of users. Australia has experienced four major droughts over the past century. The Federation Drought (1895 - 1902) affected most of the country, but particularly Queensland, Victoria and New South Wales. The Darling River in New South Wales and rivers in Queensland were almost run dry. In 1982 - 83, south-eastern Australia experienced low rainfall levels, resulting in a total loss to the Australian economy of around $7 billion. In 1991 - 95, north-eastern New South Wales and Queensland experienced their lowest rainfall levels on record, resulting in around $5 billion loss to the economy. The 2000s drought, known as Millennium drought, caused a prolonged period of dry conditions in southern Australia. All key cities including Sydney, Melbourne, Brisbane, Adelaide, Perth, Canberra and Hobart were affected. The Murray-Darling Basin was severely affected by this drought. These droughts re-

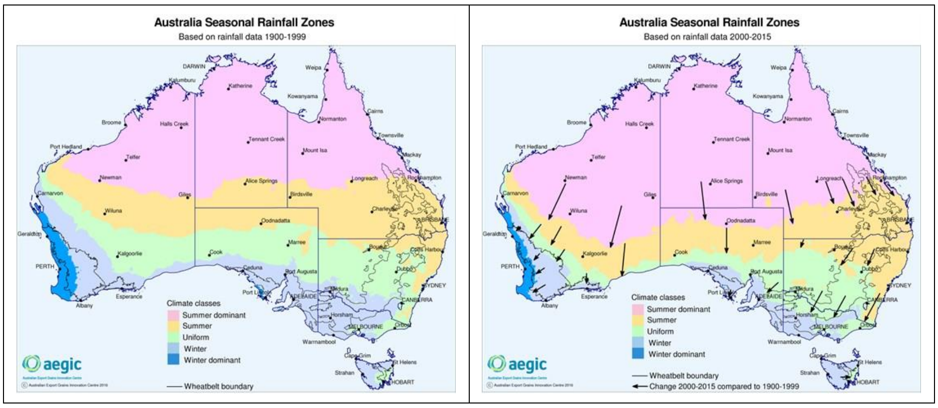

Figure 1.6: Australian seasonal rainfall zones based on rainfall data 1900-1999 (left) and 2000-2015 (right)

sulted in physical hardship, social heartbreak, animal suffering, environmental damage and financial and economic consequences.

Drought is likely to be of more concern in Australia because of population growth, industry development, long-term climate variations and high living standards. In addition, temperature is likely to be high and thus has implications for water supply owing to evaporation. Droughts had a significant influence on policy makers and were a challenge for water resource management in Australia. It is critical that appropriate water resource infrastructure and management is established to mitigate these effects.

Floods are caused by prolonged or very heavy rainfall, severe storms or tropical cyclones. In Australia, floods are usually caused by rainfall. The IPCC reported that the frequency of heavy precipitation or the proportion of total rainfall from heavy falls will increase in the 21st century over many areas of the globe [70]. The Australian Bureau of Meteorology reported in its state of the climate report that heavy rainfall events have increased since 1850 [63]. As a result, a series of flood events occurred. In the 20th century, 77 floods were recorded while in the first decade of the 21st century, six floods were recorded [78]. The consequences of floods include: loss of human life, damage to property, destruction of crops, loss of livestock and damage to infrastructure.

Australia's agricultural sector is a major consumer of water and an essential part of the national economy. The Agriculture industry depends on rainfall. Any change in the rainfall directly impacts the Australian economy. For example, financial loss from the drought in 2002-2003 was estimated at $7.36 billion AUD and $6.2 billion AUD from the 2006-2007 drought. Similarly, floods lead to financial loss too. For example, in the 2011 floods in Queensland, the damage to local government infrastructure was estimated to be $2 billion, while public infrastructure across the state sustained damage of between $5-6 billion. To counter these challenges, it is important to develop improved prediction models to predict weekly, monthly and seasonal (or longer) rainfall more accurately.

Rainfall predictions can be made for some time periods, including weekly, monthly and seasonal. To date mainly seasonal rainfall prediction models have been studied by many researchers.

In rainfall prediction, the month is used to define the start, duration and end of the rainy season. Monthly rainfall values provide more accurate an intra-year rainfall distribution than seasonal rainfall values. Monthly rainfall is an important contributing factor in agricultural and hydrological activities. Its accurate prediction can improve the quality of decision making in such activities. Similarly weekly rainfall predictions also has importance such as heavy rainfall in a short time (flood) causes significant events that affect human life. Also weekly rainfall predictions has importance in agricultural operation activities such as land preparation, crop and variety selection and irrigation.

Many algorithms have been developed by researchers to improve rainfall predictions. However, predicting rainfall is a complex process, needing continual improvement. Accurate rainfall prediction is a serious concern in many countries; especially in Australia where the climate is highly variable. Accurate rainfall prediction allows the relevant authorities to better prepare and plan. The accurate prediction of future water availability is critical for making informed policy, planning and management decisions, and will also help with the most sustainable operation of water resource systems.

1.3 Objectives of the Thesis

The research objectives of this thesis are to:

- Develop an optimization based clusterwise linear regression method. This will be done by modifying existing clusterwise linear regression methods for large scale data sets.
- Develop prediction methods based on clusterwise linear regression to improve rainfall prediction accuracy.
- Carry out a comparative assessment of linear and non-linear methods in predicting the monthly and weekly rainfall in Australia.
- Assess the influence of geographic regions on the performance of models in predicting monthly and weekly rainfall in Australia.

- Determine the combined influence of meteorological parameters on rainfall prediction in Australia.

1.4 Thesis Outline

This thesis consists of seven chapters. The current chapter, Chapter 1, is an introductory chapter. Chapter 1 provides detailed information about the importance of water, rainfall, meteorological parameters and their influence on rainfall and climate in Australia.

Chapter 2 provides an overview of relevant work, a brief explanation of existing prediction methods and justification for the further development of new methods for rainfall prediction.

In Chapter 3, we provide a brief description of clusterwise regression and clusterwise Linear Regression based on Maximum Likelihood Estimation. In this chapter, we also develop optimization based clusterwise linear regression method by modifying the existing methods. Following that, we develop prediction methods based on clusterwise regression to improve rainfall prediction accuracy.

Chapter 4 provides geographic details of the study area and descriptive statistics of the data. Furthermore, the implementation of models and statistical evaluation of the models performance are also discussed in this chapter.

Chapter 5 presents computational results for monthly rainfall predictions and a comparison of the proposed method with the clusterwise linear regression method based on maximum likelihood estimation, linear support vector machines, support vector machines with radial basis kernel function , multiple linear regression, artificial neural networks with and without a hidden layer and k-nearest neighbors methods.

Chapter 6 presents computational results for weekly rainfall predictions and a comparison of the proposed method with other selected prediction methods. Finally, findings of this research and recommendations for further research are given in Chapter 7.

Chapter 2

Literature Review

In this chapter, an overview of the relevant work and a review of existing prediction methods and their applications for rainfall predictions are given. We give justification for the further development of new methods for rainfall predictions.

2.1 Introduction

Over the past century, various algorithms have been formulated to model and predict rainfall. These models can be classified as physical models and data-driven models. Physical models are based on the physical laws to model all relevant physical processes that contribute to the rainfall process. In Australia, three physical models based on climate indices have been officially used for rainfall predictions. The first was developed in 1989 by the Australian Bureau of Meteorology (BOM). The second was developed in 1994 by the Queensland Government through the Department of Environment and Resource Management. The third was developed in 2013 by the BOM, was named Predictive Ocean Atmosphere Model for Australia (POAMA) [3]. The first two models were used to generate seasonal (three months) rainfall predictions, while the third one was used to generate seasonal and monthly rainfall predictions.

The predictions from these models give the probability of exceeding the average rainfall. The major drawbacks of these models are:

1. They do not provide any information about the level of expected deviation from the mean rainfall value within the described prediction period [3].

2. They only provide seasonal predictions which are useful for some sectors, such as agriculture, but in some areas predictions within the season are more important than seasonal predictions, e.g. the management of water infrastructure [81].

Data-driven models use historical time series data to make future predictions. These models can be classified into linear and non-linear models. They are based, in particular, on statistical and computational intelligence approaches. These models are less data demanding and a lower computational cost than physical models. Most data-driven models use sets of attribute values as input to predict the corresponding rainfall value. Each input set of attribute values may contain lagged rainfall values, or other lagged climate-related values such as temperature. Comparative studies have shown that data-driven models have better performance for rainfall prediction when compared to physical models [2, 3].

2.2 Overview of existing prediction methods

Rainfall data is classified as discrete time series as it is observed over a specific time interval. A time series can be separated into four components; trend, cyclical, seasonal and irregular variations. A trend is a long-term tendency (increasing or decreasing) in a time series. For example, an upward trend in annual rainfall in Australia (Figure 1.4). Seasonal variations are the short term fluctuations within a year during the season, for example, the temperature increased in Summer and decreased in Winter. Seasonal variation is an important factor for proper future planning. Cyclical variation is the medium-term change in the time series, which repeats in cycles. Irregular variations are unpredictable changes in the time series. When a time series contains a trend, seasonality or other systematic components, the general summary statistics can be seriously misleading.

To observe any trend over time, any regular seasonal behavior, or any changes in level and variability over time, observations are usually plotted against corresponding times. These features need to be identified to incorporate them into mathematical models. In practice, a suitable mathematical model is developed to a given time series and the corresponding parameters are estimated. The procedure of fitting a time series to a proper model is called time series analysis. The objectives of time series analysis are to describe and summarize data, develop prediction models and predict future values.

A number of data-driven prediction models were considered for rainfall prediction, among which the most popular and widely used models are artificial neural networks (ANN), autoregressive integrated moving average (ARIMA), the K-Nearest-Neighbours (k-NN), multiple linear regression (MLR) and support vector machines for regression (SVMreg) methods. A brief description of these models including their application in rainfall prediction is given below.

2.2.1 Regression Analysis

Regression analysis is the most widely used statistical method for investigating and modelling the relationship between a variable of interest (response) and a set of related predictor (explanatory) variables. Applications of regression occur in almost every field including hydrology. Regression models are used for several purposes, including data description, parameter estimation and prediction. These models can be classified into simple linear regression, multiple linear regression and nonlinear regression models. The simple linear regression model is a model to find the linear relationship between a response variable and one independent variable. The simple regression model is often written as:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where y is the response variable, β_0 is the y intercept, β_1 is the gradient or slope of the regression line, x is the regressor or predictor variable and ε is the random error component.

As the response y is a random variable, there is a probability distribution for y at each possible value for x. The slope β_1 is the change in the mean of the distribution of y made by a unit change in x. The term linear is used because the model equation is a linear function of the unknown parameters β_0 and β_1.

A regression model that involves more than one regressor variable is called multiple linear regression model. The general form of the multiple linear regression model is written as:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$$

where y is the response variable, $\beta_0, \beta_1, \ldots, \beta_p$ are regression coefficients and x_1, x_2,..., x_p are regressor or predictor variables.

A nonlinear regression model is a model to find the nonlinear relationship between a response variable and one or more independent variable. Nonlinear regression may be written as:

$$y = \frac{\alpha}{1 + \exp(\beta^t)}$$

The major assumptions of the regression model are:

- The relationship between the response variable y and the regressors is linear, at least approximately.
- The error term ε has zero mean and constant variance σ^2.
- The errors are uncorrelated.
- The errors are normally distributed.

The assumption of normality is required for hypothesis testing and constructing the confidence interval. The assumption of uncorrelated or independent errors is not appropriate for time series data, as the errors in time series data present serial correlation. Such error terms are assumed to be autocorrelated. There are several causes of autocorrelation in regression problems involving time series data, including failure to include one or more significant predictors in the regression model. For example, suppose we are predicting monthly rainfall using the monthly temperature as a predictor. The effects of autocorrelation on the ordinary least squares regression procedure are [56]:

- Regression coefficients are no longer minimum variance estimates.
- Residual mean square may seriously underestimate σ^2.
- Test of hypothesis and confidence interval based on F and t distribution are no longer appropriate.

The MLR models are mostly used for comparison with other models for rainfall prediction. In the paper [52], the author compared the ANN with the MLR model for long-term seasonal Spring rainfall prediction in Victoria, Australia using lagged El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as input variables. The paper [27] presented a comparison of the ANN with MLR for rainfall-runoff prediction in Japan. In the paper [82], the author predicted an Indian Summer monsoon rainfall with ANN and MLR models using El Nino indices as predictor variables and compared the performance of both models.

Aksoy et al. in [5] predicted precipitation in Jordan using ANN and compared the prediction performance with the MLR model. Similarly Chattopadhyay et al. in [19] compared the ANN and the MLR model performance in predicting the annual average Indian Summer monsoon rainfall. The paper [76] also compared the ANN and MLR for daily rainfall prediction in Sao Paulo, Brazil using temperature, wind, humidity, air temperature, precipitable water, relative vorticity and moisture divergence flux as input variables.

2.2.2 ARIMA Model

The most widely used time series prediction technique is the autoregressive integrated moving average, denoted by ARIMA (p, d, q), where p and q are the autoregressive and moving average orders and d is the order of differentiation, operated on the original series to handle non-stationeries. The ARIMA model is a generalization of autoregressive moving average (ARMA) model. The ARMA model is a combination of autoregressive and moving average processes.

Auto regressive (AR) process is a function of p past observations $(X_{t-1}, X_{t-2}, \ldots, X_{t-p})$ and random disturbance term process, where p indicates the number of steps into the past needed to forecast the current value. An autoregressive model of order p, abbreviated $AR(p)$, can be represented by the equation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + w_t$$

where X_t denotes the time-series and w_t indicates a white-noise process.

Moving average process is the linear function of the white noise. A moving average model of order q, abbreviated $MA_{(q)}$ can be formulated as follows:

$$X_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_q w_{t-q}$$

. General auto-regressive moving average model $ARMA(p, q)$ is defined by:

$$X_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - .... - \phi_p X_{t-p} = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_q w_{t-q}$$

Using the Box and Jenkins notation, the $ARMA(p, q)$ model can be written as:

$$\phi(B)X_t = \theta(B)W_t$$

where B is the back shift operator, X_t is the zero mean time series, w_t is a white noise, and ϕ and θ are the respectively the p^{th} and q^{th} order autoregressive and moving average components.

ARMA models are based on the assumption that the time series are approximately stationary. A stationary time series is one whose statistical properties such as mean, variance and covariance are all constant (do not change) over time. Different transformation techniques such as logarithms, square root or differencing are used for eliminating each type of feature to make the time series stationary. These transformation techniques may turn skewed data into symmetric data and make the variance constant over time. The ARMA model fitted to the transformed series is called the ARIMA (p, d, q) model and can be written as:

$$\phi(B)(1 - B)^d X_t = \theta(B)W_t$$

where d is the order of differentiation of the original data to obtain a stationary process. The ARIMA models are stationary only if $|\phi| \neq \pm 1$.

If $|\phi| < 1$ then the unique stationary solution is casual, that is the series X_t can be expressed in terms of the current and past values. If $|\phi| > 1$ then the unique stationary solution is non-casual since X_t is the function of current and future W_s, s

$\leq t$. Similarly if $|\theta| < 1$ then the unique stationary solution is invertible, which means w_t can be expressed in terms of current and past X_s, $s \leq t$.

ARIMA models have been widely used in hydrological research. In the paper [20], the authors applied ARIMA and autoregressive neural network models to predict Summer monsoon rainfall in India. The paper [30] investigated the application of ARIMA and ANN models for monthly rainfall prediction in Iran. In [57], pre-monsoon rainfall trends were assessed, and rainfall was predicted using the univariate ARIMA model. In the paper [71], the ARIMA model was applied to predict seasonal rainfall in Indonesia. Similarly, in the paper [84], the authors predicted the yearly rainfall with ARIMA and ANN models in India.

2.2.3 Artificial Neural Networks (ANN)

Neural networks were first introduced in 1943 by McCulloch and Pitts and gradually progressed with advances in calibration methodologies. ANN are widely used in diverse disciplines including science, engineering and economics. Researchers in hydrology adopted this method during the last decade and this computational tool is still in its growing stage.

Artificial neural networks are massive, parallel distributed, information-processing systems with characteristics resembling the biological neural networks of the human brain [2].

An artificial neural network consists of simple neurons, and links that process information in order to find the relationship between inputs and outputs. An artificial neural network works like a human brain. It takes input (like synapses) and then applies the activation function which combines the input into a single value and produces an output (Figure 2.1). The activation function is generally divided into two parts, the combination function and the transfer function. The combination function assigns weights to each input and combines the weighted inputs in a single value. The transfer functions produce an output. Transfer functions could be any mathematical function. A number of transfer functions were considered in the literature, among which the most common are sigmoid, hyperbolic tangent and step functions. The step function limits the output of the neuron to either 0 or 1, the sigmoid function between 0 and 1 and the hyperbolic tangent function from -1 to +1. The graphic illustration of these functions is given in Figure 2.2.

Network Architectures

The combination of two or more artificial neurons makes an artificial neural network and the way these individual neurons interconnect is called topology or architecture of the artificial neural network. The interconnection of the artificial neurons can be

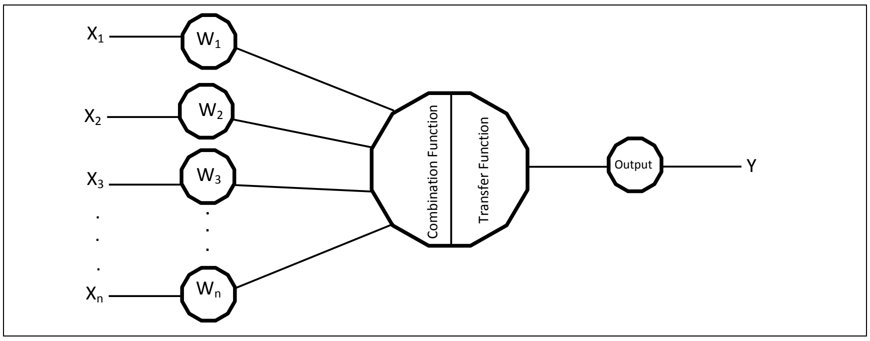

Figure 2.1: Architecture of an Artificial Neural Network.

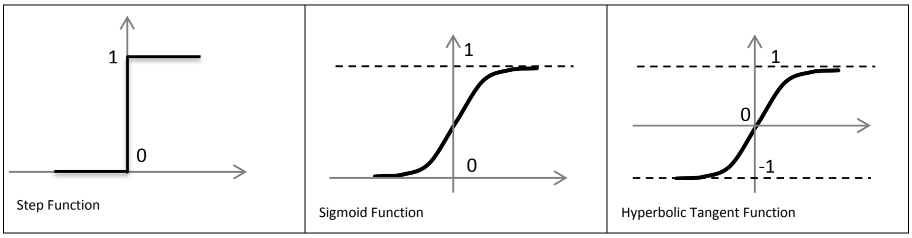

Figure 2.2: Transfer functions.

done in many ways, resulting in several architectures. The taxonomy of the neural networks is given in Figure 2.3. All possible architectures are divided into two classes; feed-forward networks and recurrent/ feedback networks. An illustration of a feed-forward network and a recurrent/feedback network architectures is given in Figure 2.4. Figure 2.4 (left side), represents the feed-forward neural network architecture, where data moves only in a forward direction from input layer which goes through the hidden layer and then to the output layer. The right side of Figure 2.4 represents recurrent/feedback architecture where hidden layers and output layers have recurrent connections. Feedback neural networks have at least one feedback loop, either from a hidden layer to an input layer or from an output layer to a hidden layer (Figure 2.5).

Feed-forward neural networks can be further divided into single layer feed forward neural networks and multi-layer feed-forward neural networks. Figure 2.6 illustrates the architecture of both single layer and multi-layer feed-forward neural networks. The single layer neural network is the simplest form of ANN. In a single layer feed-

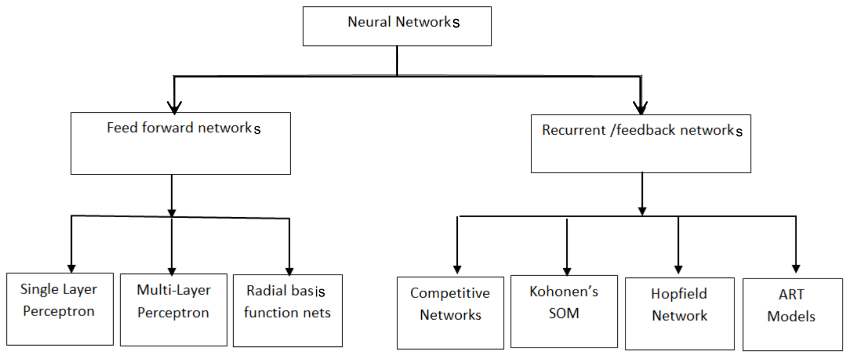

Figure 2.3: Taxonomy of Neural Networks.

forward neural network the input layer directly links to the output layer. There is no hidden layer in this neural network. On the other hand, a multi-layer feed-forward neural network has one or more hidden layers between the input layer and the output layer. These neural networks are also called multi-layer perceptron. Feed-forward neural networks are most appropriate models for modelling relationships between the response variable and predictor variables. The multi-layer feed-forward neural networks are the most widely studied and applied neural network models in practice.

In the multi-layer feed-forward neural networks, the data is divided into three sets: the training set used to update the network weights and biases; the validation set, used to guarantee the generalisation capability of the model; and the test set, used to check the generalisation. The multi-layer feed-forward neural networks have the ability to learn through training. Training a multilayer perceptron is a procedure to find the combination of weights which results in the smallest error. There are many algorithms that could be used to train a multi-layer perceptron, but the back propagation algorithm and the algorithms derived from it are the most computationally straightforward algorithms for training the multi-layer perceptron.

Applications of artificial neural networks for rainfall prediction

The artificial neural network has been suggested as an alternative method for time series modelling and predictions. ANN makes no prior assumptions like statistical techniques, and can model highly nonlinear data. These features make it more practical and accurate in modelling and predicting complex data such as rainfall data. ANN has been applied extensively in hydrology, including in rainfall prediction. Some of the most recent applications of ANN for rainfall prediction are given below:

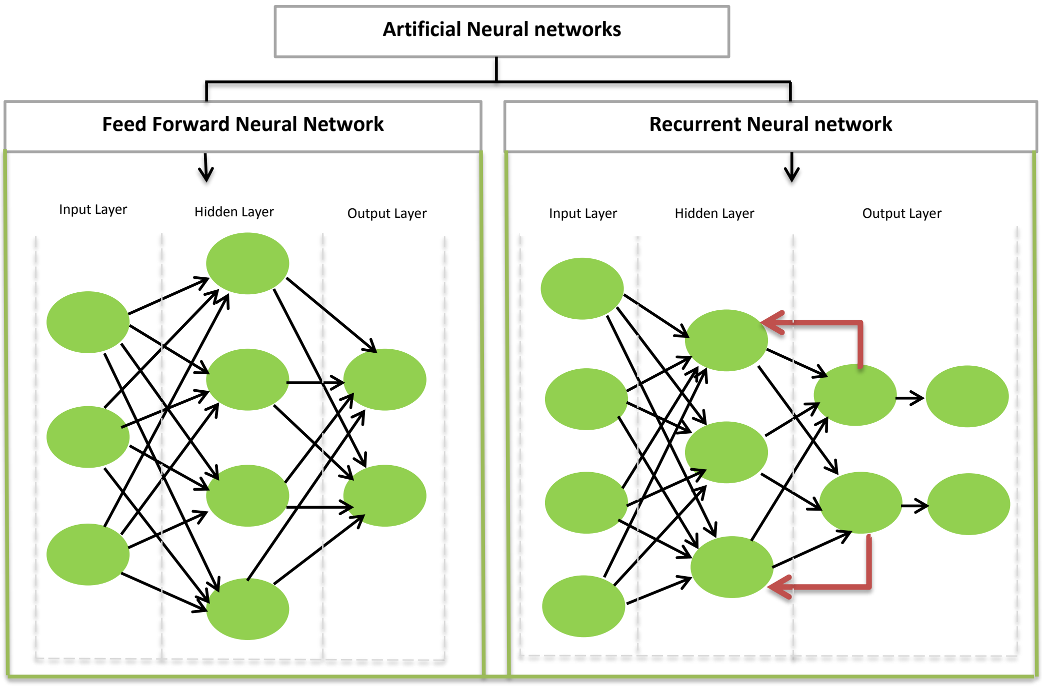

Figure 2.4: Artificial Neural Networks architecture.

Abbot et al. in [2] applied the time delay neural network (TDNN) for monthly and seasonal rainfall prediction in Queensland, Australia using monthly rainfall, atmospheric temperature, SOI, PDO, and El Nino 3.4 as input variables. The TDNN model prediction performance was compared with the POAMA model and it was concluded the TDNN model performed better when compared to the POAMA model.

Abbot et al. in [3] applied ANN and POAMA for rainfall prediction in Queensland, Australia. The input variables used were rainfall, maximum and minimum temperatures, Southern Oscillation Index (SOI), Inter-decadal Pacific Oscillation (IPO), Dipole Mode Index (DMI) and El Nino 3.4. The study found the ANN model provides more skilled predictions compared to POAMA model.

Acharya et al. in [4] developed a multi-model ensemble technique (MME) based on ANN for rainfall prediction in India. The performance of the developed model compared with the traditional multi-model ensemble techniques and it was concluded that the MME method based on ANN is better in predicting rainfall than traditional MME methods.

Aksoy et al. in [5] predicted precipitation in arid and semi-arid regions in Jordan using a feed-forward back propagation neural network, a radial basis function neural network, a generalized regression neural network and the MLR at three petrological stations. Antecedent precipitation and the periodic component were used as input variables. The Feed-forward back propagation neural network was found to be the best model for monthly rainfall prediction as compared to the other three models.

Chattopadhyay et al. [19] predicted the annual average Indian Summer mon-

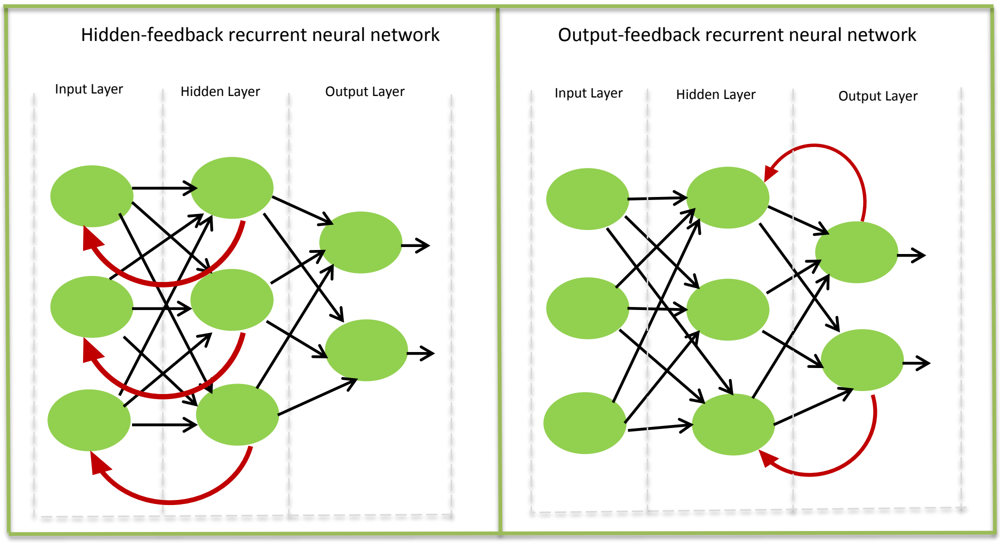

Figure 2.5: Recurrent Neural Network architectures.

soon rainfall using ANN with three different back propagation learning rules and the asymptotic regression model. The three different back propagation learning rules were momentum learning, conjugate gradient descent learning, and Levenberg Marquardt learning. The study found that ANN with conjugate gradient descent learning and Levenberg Marquardt learning perform better than the other models.

Ravinesh et al. in [24] applied ANN for predicting monthly Standardized Precipitation and Evapotranspiration Index (SPEI) in eastern Australia, using a total of 18 predictor variables. The predictor variables used in developing ANN were monthly rainfall totals, mean temperature, minimum temperature, maximum temperature, evapotranspiration, climate indices (Southern Oscillation Index, Pacific Decadal Oscillation, Southern Annular Mode and Indian Ocean Dipole) and the Sea Surface Temperatures (Nino 3.0, 3.4 and 4.0). The study found that the ANN model is a useful data-driven tool for monthly SPEI predictions.

El-Shafie et al. in [28] examined the performance of the multi-layer perceptron neural network (MLP-NN), the radial basis function neural network (RBFNN) and the input delay neural network (IDNN) for weekly and monthly rainfall prediction in Malaysia using temperature and humidity as input variables. The results reveal that the IDNN model achieved the highest accuracy level in predicting rainfall when compared to other models.

El-Shafie et al. in [27] applied the ANN and MLR models for rainfall-runoff prediction in Japan.The results showed that ANN could describe the behaviour of the rainfall-runoff relationship more accurately than the classical regression model.

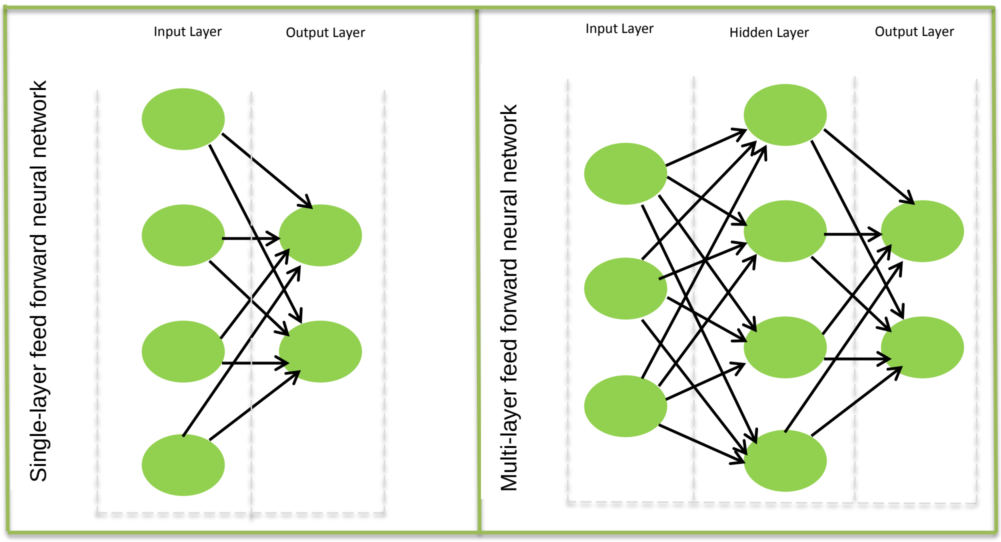

Figure 2.6: Single-layer and multi-layer feed-forward neural network architectures.

Farajzadeh et al. in [30] considered rainfall predictions with ANN and ARIMA models using only historical monthly rainfall data as input. The performance of the models was compared and it was found that there was no significant difference in the performance of either model in predicting monthly rainfall.

Karamouz et al. in [44] developed and applied the time delay recurrent neural network for long-lead seasonal rainfall prediction in three case studies in Iran. The investigators compared the prediction performance of the time delay recurrent neural network with ARMAX and found that the time delay recurrent neural network perform better than the ARMAX model in all three cases.

Mekanik et al. in [52] investigated the application of ANN and MLR models for long-term seasonal Spring rainfall prediction in Victoria, Australia using lagged El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as input variables. Both models were assessed and ANN model was found to have lower prediction errors compared to the MLR model.

Mandal et al. in [51] applied ANN and the model tree for short-term rainfall prediction in India. Three training algorithms, namely multilayer perceptron, radial basis function and time lagged recurrent networks were used in training the ANN. In comparison, researchers found that the performance of both models was the same in predicting short-term rainfall.

Ramirez et al. in [76] investigated the ANN for daily rainfall prediction in Sao Paulo, Brazil using potential temperature, wind, humidity, air temperature, precipitable water, relative vorticity and moisture divergence flux as input variables. The prediction performance of ANN was compared with the MLR model. The study suggested that ANN is more suitable for rain prediction in Sao Paulo, Brazil.

Shukla et al. in [82] predicted Indian Summer monsoon rainfall with the ANN and MLR models using Nino indices as predictors. The comparative analysis suggested that ANN model performance is better than the MLR model.

2.2.4 Support Vector Machine for regression (SVMreg)

Support Vector Machines (SVM) is a statistical learning technique, developed by Vapnik and his colleagues in 1995 [23]. Initially, SVM was developed for classification problems but was soon successfully applied in other areas, such as regression estimation and time series prediction problems. In the supervised data classification, the basic idea of SVM is to find a maximum margin hyperplane, which separates two finite point classes given in the n-dimensional space. The maximum margin hyperplane is the one which has the maximum distance to the closest data points, as illustrated in Figure 2.7. Here we will briefly describe SVMreg estimation.

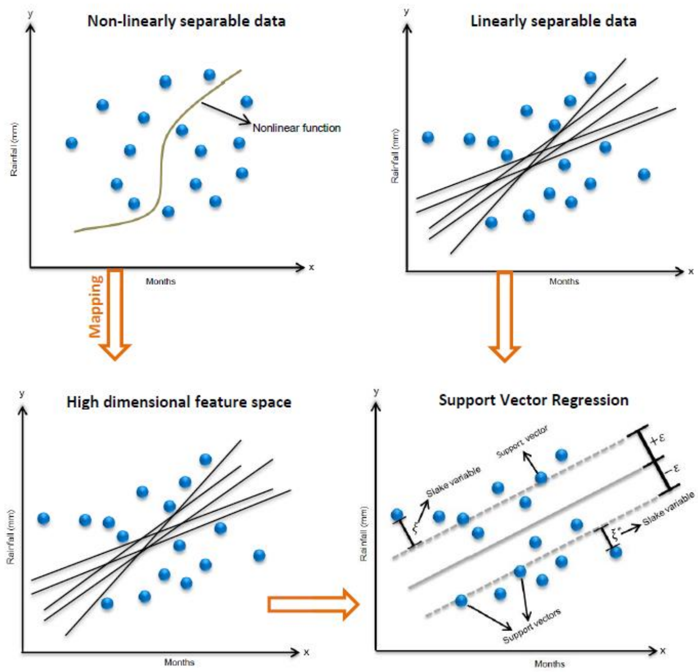

Figure 2.7: An illustration of Support Vector Machines.

Consider the training data $T = \{(x^1, y_1), (x^2, y_2), \dots, (x^k, y_k)\} \subset \mathbb{R}^n \times \mathbb{R}$, where x^i is an input vector and $y_i \in \mathbb{R}$ is a corresponding output, $i = 1, \dots, k$, k is the number of observations in the training data set. Given an $\varepsilon > 0$ the aim of SVMReg is to find a function $f(x)$ that has at most ε deviation from the targets y_i for all the training data.

In the linear SVMReg, the regression function f can be written as:

$$f(x) = w^t x + b \tag{2.2.1}$$

where w is the weight vector, b is the bias of the regression function and t stands for the transpose of a vector. w and b are estimated by minimizing the following risk function:

$$\text{minimize} \;\; R = \frac{1}{2} w^t w + C \sum_{i=1}^{k} (\xi_i + \xi_i^*), \tag{2.2.2}$$

subject to

$$\begin{cases} y_i - [w^t x^i + b] & \leq \varepsilon + \xi_i \\ [w^t x^i + b] - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^*, & \geq 0. \end{cases} \tag{2.2.3}$$

The first part ($w^t w$) of the risk function regularizes weight sizes and penalizes large weights, $C > 0$ is a penalty parameter which determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated. This corresponds to dealing with a so-called ε-insensitive loss function $|\xi|_\varepsilon$ defined as:

$$|\xi|_\varepsilon = \begin{cases} 0, & \text{if} \;\; |\xi| \leq \varepsilon \\ |\xi| - \varepsilon, & \text{otherwise.} \end{cases}$$

Here ξ_i and ξ_i^* are slack variables introduced to deal with infeasibility.

Problems (2.2.2)-(2.2.3) are optimization problems which are usually solved using Lagrange multipliers. The linear SVMreg model can be extended to a non-linear SVMreg model by applying kernel functions. The kernel function can be defined as:

$$K(x^i, x^j) = \phi(x^i)^t \phi(x^j)$$

where ϕ is the nonlinear transformation of the input space. In particular, ϕ can be a normal distribution function which also is known as radial basis function (RBF). The details of the SVMreg model can be found in [22, 55].

There have been limited studies on the application of the SVMReg model for rainfall prediction. In [49, 50] this model was combined with the multi-objective genetic algorithm to predict hourly rainfall in Taiwan with a lead time of 1 to 6 hours. The meteorological variables: air pressure, air temperature, wind velocity, wind direction and sunshine duration were used as inputs. In the paper [59], the authors applied the SVMreg to predict extreme rainfall events in India with a lead time of 6 to 48 hours.

The SVMReg models were applied for warm season thermodynamically driven rainfall prediction in [53]. The discrete wavelet transform and SVMreg methods were combined in [45] to predict one-day-ahead precipitation in Turkey. In [48], the authors compared the performance of the SVMReg model and the back propagation neural networks for typhoon rainfall prediction in Taiwan and found that SVMReg performed much better than the back propagation neural networks.

2.2.5 k-Nearest Neighbour Method

The k-Nearest Neighbours method (k-NN), is a non-parametric statistical pattern recognition procedure, extended to time series prediction in [92]. For a time series prediction, this algorithm identifies the most similar past sequences in the training data set to the one being predicted and combines their output values to predict the next value of the target sequence.

Next, we briefly describe the k-NN for regression. Consider a finite time series $y_t, t = 1, 2, \ldots, m$ without explanatory variables. In the first step the series is transformed into equal length feature vectors of d observations: $y_t^d = (y_t, y_{t-1}, \ldots, y_{t-(d-1)})$. Here $d < m$ is a predetermined integer called an embedding dimension. In the next step, either a set of $m-d$ overlapping vectors with $t = (d, d+1, \ldots, m-d)$ or a set of m/d non-overlapping vectors with $t = (d, 2d, \ldots, m-d)$ is defined. These vectors are called d-histories. As a result the d-dimensional space is considered to be the phase space of the time series.

In the third step, the similar repetitive patterns in the time series are identified. This can be done by calculating either the distance or the correlation between the last observed vector $y_m^d = (y_m, y_{m-1}, \ldots, y_{m-(d-1)})$ and all d-histories. In k-NN applications the Euclidean distance is predominantly used.

In the fourth step of the k-NN, the calculated distances are ranked and the k vectors having the lowest distance from the target feature vector are selected. Consecutive observations in the selected feature vectors are then combined to form a prediction, often using a simple arithmetic mean with equal weights. In the case of using correlations, k vectors having the highest correlation with the feature vector are combined to form a prediction. The predicted value is simply based on the k most similar values in the neighbourhood. No theoretical or analytical assumptions are required between the inputs and the outputs.

The k-NN for univariate time series can be easily extended to multivariate time series by extending the creation of vectors for each input variable and applying the distance function to a multivariate case. Owing to simplicity and easy implementation, the k-NN method is very attractive to forecasters. It is considered one of the top ten most influential data mining algorithms in research [91]. An illustration of the k-NN method where n=7 and $k = 3$ are shown by [92] and is given in Figure 2.8.

In [34] the author applied the k-NN model to predict daily mean discharge in a mountain basin in northeastern Italy, and compared the prediction performance with the Autoregressive Exogenous model. Shamseldin et al. in [79] applied the nearest neighbour method as the nearest neighbour linear perturbation model (NNLPM) for river flow prediction and compared the results with the simple linear model and the linear perturbation model. Toth et al. in [87] compared the prediction performance of k-NN, ANN and autoregressive moving average models for short-term rainfall prediction in Italy with a leading time from 1 to 6 hours. Brath et al. in [17] investigated the prediction performance of ARMA, ARIMA, artificial neural networks and nearest neighbour methods for rainfall and discharge predictions. In [29] Eskandarinia et al. applied k-NN for daily flow forecasting in Iran and compared the prediction performance with the ANN model.

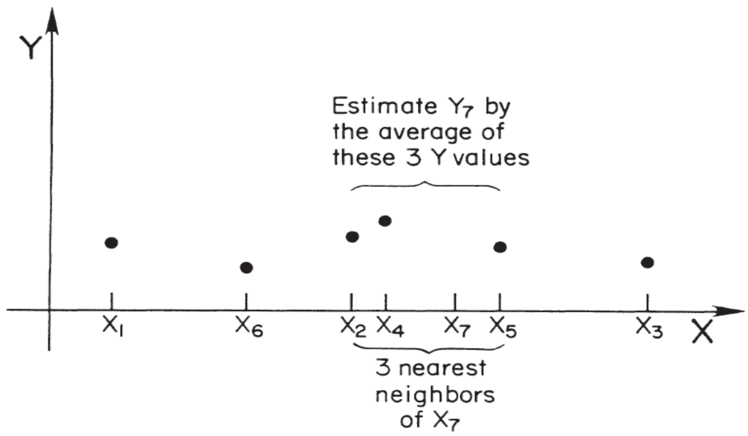

Figure 2.8: An illustration of k-NN method.

2.3 Concluding remarks

The importance of accurate prediction is highlighted where the hydrologic regime is extremely variable from year to year, or where there is an increasing demand for water due to population growth, irrigation or industry needs. In Australia, almost every industry relies on water, particularly agriculture [65] and water supply is totally dependent on rainfall and snow [66].

The accurate prediction of rainfall is very important for food production planning, water supply planning, reservoir operations and other water resource management and analysis activities. Accurate prediction is one of the major challenges facing meteorologists. Researchers tried a number of prediction methods to predict rainfall; some are more accurate than others [10].

Rainfall in Australia is highly variable in both space and time and dynamics are inherently non-linear in nature. High rainfall variability highlights the need for a deep analysis of rainfall data for a comprehensive understanding of the reality of variability and accurate rainfall prediction. Accurate rainfall prediction is critical, but difficult because of many variables interacting with rainfall including, but not limited to: all three surrounding oceans, temperature, vapour pressure, evaporation and solar radiation.

Clusterwise linear regression has the capability to model rainfall dynamics in Australia with several interrelated parameters and help improve water resources planning and management. This method identify clusters of data that have some common characteristics and fits the regression functions within each cluster. Predictions can be computed from each cluster using prediction methods.

Several methods have been used for rainfall prediction. However, rainfall prediction is a complex process, needing continual improvement. . Research in this area is vital. There is always a need to improve the existing models and develop new methods for accurate rainfall prediction.

Chapter 3

Clusterwise Linear Regression

In this chapter, we provide a brief description of the clusterwise linear regression method based on maximum likelihood estimation (CR-EM). Then we present nonsmooth nonconvex optimization formulation of the clusterwise linear regression problem and an incremental algorithm for solving it. Following that, we develop prediction methods for clusterwise linear regression.

3.1 Introduction

In many applications, the presence of heterogeneous groups, nonlinear relationships or time series necessitate the use of two or more regression functions to best summarise the underlying structure of the data. The clusterwise regression is one such technique, which can be used to discover trends within data when more than one trend is likely to exist. The process of rainfall is very complex as it is highly nonlinear, time varying, spatially distributed and not easily describable by simple models.

Many algorithms ranging from linear to nonlinear have been formulated to model this complex nonlinear process but no one is adequate for accurate rainfall prediction. The Clusterwise linear regression is one of the nonlinear prediction methods, which has the capability of processing highly nonlinear data such as rainfall data. The Clusterwise linear regression is a hybrid of clustering and regression methods. The aim of the clusterwise regression is to find simultaneously an optimal partition of data in k clusters and fit the regression functions to each cluster to minimise the overall fit.

Clustering is a data mining technique, consisting of finding subsets of similar points in a data set on the basis of similarities. Several clustering algorithms exist in the literature. These methods can be categorized into: hierarchical, partitioning (e.g. k-means method), density-based (e.g. expectation maximization algorithm), model-based (e.g. neural networks (self-organizing map)), grid-based methods (e.g. STING

(STatistical INformation Grid approach)) and soft computing methods (e.g. fuzzy clustering). Regression analysis consists of fitting a function (often linear) to the data to discover how one or more variables vary as a function of another.

Different approaches were developed to solve the clusterwise linear regression problem including those based on data mining [85, 86], statistical [26, 33, 35] and optimization [14, 15, 16] techniques. The parameters that need to be estimated in these models are; number of clusters, regression coefficients for each cluster and variance of residuals within each cluster. In [85, 86] Spath proposed methods to solve CLR using generalizations of classical clustering algorithm k-means with a criterion based on the minimization of the squared residuals. Garcia et al. in [36] extended the TCLUST methodology to perform robust clusterwise linear regression. TCLUST methodology is a statistical clustering technique based on the modification of a trimmed k-means clustering algorithm. DeSarbo and Cron (1988) in their paper [26] introduced conditional mixture, maximum likelihood methodology together with the Expectation Maximization (EM) algorithm for performing clusterwise linear regression. Henning [40, 41] developed three models: the mixture model with either fixed or random regressors, and the fixed partition model with fixed regressors using the maximum likelihood estimation approach together with the EM Algorithm. Preda and Saporta [73] applied CLR to investigate the stock exchange data and proposed the partial least squares approach, given in [90] to estimate the regression coefficients for each cluster.

Existing clusterwise linear regression algorithms suffer from the same drawbacks as their clustering counterparts: they are very sensitive to the choice of an initial solution; and they may lead to sub-optimal solutions [93]. Furthermore, most of these algorithms assume the number of clusters to be known a priori. Most of the algorithms try to separate data into subsets of observations and use one regression function for each subset.

There have been several attempts to simultaneously find all regression functions to approximate a data set and to estimate the number of subsets. The paper [25] presents a methodology which simultaneously clusters observations into a preset number of groups and estimates the coefficients of the corresponding regression functions. Then a simulated annealing-based methodology is described to accommodate overlapping or non-overlapping clustering. In the paper [46], the authors show that estimation of the clusterwise regression model is equivalent to solving a nonlinear mixed integer programming problem.

An information-based criterion for determining the number of clusters in the clusterwise regression problem is proposed in [80]. It is shown that, under a probabilistically structured population, the proposed criterion selects the true number of regression hyperplanes with probability one among all class-growing sequences of classifications, when the number of observations from the population increases to infinity.

The paper [74] studies the problem of estimating the number of clusters in the context of logistic regression clustering. The classification likelihood approach is employed to tackle this problem. A model-selection based criterion for selecting the number of logistic curves is proposed, and its asymptotic property is also considered.

In this study, a method based on nonsmooth, nonconvex optimization formulation and an incremental approach is proposed for solving the clusterwise linear regression. This method starts with one regression function and summaries the underlying structure of the data by dynamically adding one hyperplane at each iteration. A special procedure is introduced to generate good starting points for solving global optimization problems at each iteration of the incremental algorithm. Such an approach allows one to find a high quality solution to the problem when a data set is sufficiently dense.

Several incremental algorithms have been proposed to solve the sum of squares clustering problems. The global k-means algorithm and its variations [7, 47] are based on constructing the clusters incrementally, starting from finding the center for the whole data set and then adding a cluster at a time and refining the new set of clusters by applying k-means.

We develop a similar scheme in order to solve the CLR problem by using a nonsmooth, non-convex optimization formulation from [14] and an incremental algorithm from [16]. In the proposed method we used affine functions as representatives of clusters instead of classical centers.

3.2 Clusterwise linear regression based on maximum likelihood estimation (CR(EM))

Clusterwise regression based on maximum likelihood methodology also known as finite mixture models for regression problems [61] and finite mixtures of linear regressions [31] in the literature. Finite mixture models are a class of probability distribution, initially developed by Newcomb (1886) [60] and Pearson (1894) [72]. Later in 1972, Quandt and Richard in [75] extended the mixture models for the modeling of regression data. Finite mixture models are convex combinations of two or more probability density functions.

The illustration of the overall density and components densities with the EM algorithm using rainfall data of Wiluna, Western Australia is given in Figure 3.1. In the figure dotted line shows the overall density, and red and green are the two component densities.

In mixture model approach, it is assumed that observations arise from k distinct random processes. Each of these processes is modelled by the specific density function. Let x be a random variable and $f(x, \theta_k)$ be a probability density function for each

$k = 1, .., K$. Then the random variable x is said to arise from a finite mixture model if it has a density function of the form:

$$\begin{aligned} h(x, \Theta) &= \sum_{k=1}^{K} \pi_k f(x, \theta_k), \\ \pi_k \geq 0, &\sum_{k=1}^{K} \pi_k = 1. \end{aligned} \tag{3.2.1}$$

Let assume y_i is distributed as a finite mixture of conditional normal densities, then a finite mixture regression model of K components can be defined as:

$$\begin{aligned} h(y|x, \Theta) &= \sum_{k=1}^{K} \pi_k f(y|x, \theta_k), \\ \pi_k \geq 0, &\sum_{k=1}^{K} \pi_k = 1 \end{aligned} \tag{3.2.2}$$

where $f(y|x, \theta_k)$ is the probability density function of the k-th component, y is a response variable with conditional density h, x is a vector of independent variables, θ_k is the component specific parameter vector for the density function, π_k is the mixing proportion also known as prior probability of component k and Θ is the vector of all parameters.

Initially, parameters of the mixture models have been estimated using the method of moments [72]. Later in 1974 Hosmer [42] suggested maximum likelihood estimation technique to estimate the mixture model parameters. The likelihood of finite mixtures can be maximised using optimization procedures such as the Newton-Raphson method or by using the Expectation-Maximization(EM) algorithm [89]. The *EM* Algorithm is most popular and widely employed in the literature because of computational attractiveness and easy to program.

The *EM* algorithm was first considered in 1977 by Arther Dempster and Donald Rubin [1]. It seeks to find the maximum likelihood estimates (MLE) iteratively applying the following two steps, **E step** (for expectation) and **M step** (for maximization). **Expectation step (E step):** Estimate the expected value of the complete data log likelihood function $\sum_{i=1}^{N} log(\sum_{k=1}^{K} \pi_k f(y|x, \theta_k))$, with respect to the unknown data y given the observed data x and the current parameter estimates.

Let $\theta^{(r)}$ be the current parameters estimate at the rth iteration. On the next iteration the EL algorithm calculate the following function:

$$Q(\theta, \theta^{(r)}) = \sum_{i=1}^{N} \sum_{k=1}^{K} w_{i,k}^{(r)} \pi_k f(y|x, \theta_k), \tag{3.2.3}$$

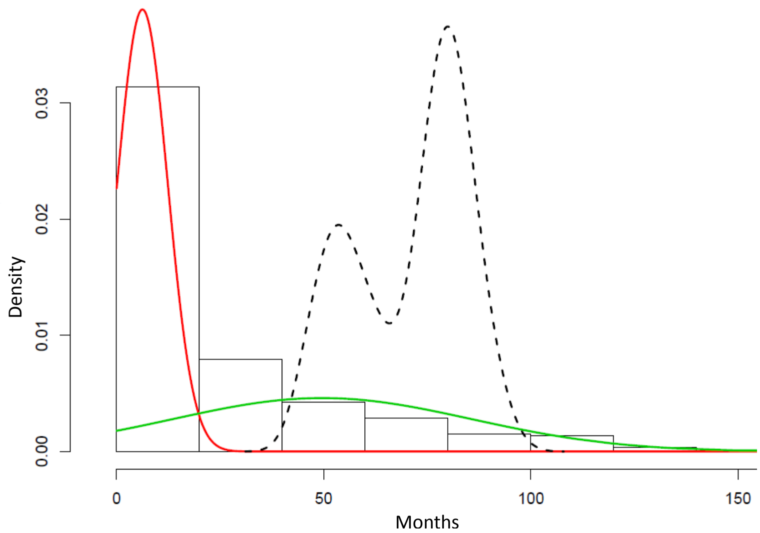

Figure 3.1: Fitted overall density(dotted) and two component densities (red and green) with EM algorithm, using the rainfall data of Wiluna, Western Australia.

where

$$w_{i,k}^{(r)} = \frac{\pi_k f(y|x, \theta_k)}{\sum_{k=1}^{K} \pi_k f(y_i|x_i, \theta_k)}$$

is the posterior probability that the ith observation belongs to the kth component of the mixture after the rth iteration.

Maximization step (M step): maximize the expectation of log-likelihood for each component separately using the posterior probabilities as weights. In the M step the $Q(\theta, \theta^{(r)})$ is maximized with respect to θ and the $(r+1)th$ iteration of the EM algorithm is defined as:

$$\theta^{(m+1)} = \underset{\theta \in \Theta}{argmax}\ Q(\theta, \theta^{(r)})$$

The E and the M steps are repeated until a convergence criterion is met under pre-specified criteria or reached to a maximum number of iterations.

3.3 Clusterwise linear regression based on nonsmooth optimization (CLR(Opt))

In this section first we present the clusterwise regression problem and then review the *Spath* algorithm and nonsmooth optimization approach.

3.3.1 Problem statement

Given a dataset $A = \{(a^i, b_i) \in \mathbb{R}^n \times \mathbb{R} : \ i = 1, \ldots, l\}$, the aim of the CLR is to find simultaneously an optimal partition of data in k clusters and regression coefficients within clusters in order to minimize the overall fit. Let $A^j, \ j = 1, \ldots, k$ be clusters such that

$$A^j \neq \emptyset, \ A^j \bigcap A^t = \emptyset, \ j, t = 1, \ldots, k, \ t \neq j \text{ and } A = \bigcup_{j=1}^{k} A^j.$$

Let $\{x^j, y_j\}$ be linear regression coefficients computed using only data points from the cluster $A^j, \ j = 1, \ldots, k$. Then for the given data point $(a, b) \in A$ and coefficients $\{x^j, y_j\}$ the square regression error $E_{ab}(x^j, y_j)$ is:

$$E_{ab}(x^j, y_j) = \big(\langle x^j, a\rangle + y_j - b\big)^2 .$$

We associate a data point with the cluster whose regression error at this point is smallest. Then the overall fit function is [14, 15, 16]:

$$f_k(x, y) = \sum_{i=1}^{l} \min_{j=1,\ldots,k} E_{ab}(x^j, y_j), \tag{3.3.1}$$

where $x = (x^1, \ldots, x^k) \in \mathbb{R}^{nk}$ and $y = (y_1, \ldots, y_k) \in \mathbb{R}^k$. The function f_k is called *the k-th clusterwise linear regression function or the k-th overall fit function.* For $k = 1$ the function f_k is convex and for $k > 1$ it is nonsmooth nonconvex piecewise quadratic function.

The k-clusterwise linear regression problem is formulated as follows:

$$\text{minimize} \ \ f_k(x, y) \ \ \text{subject to} \ \ x \in \mathbb{R}^{nk}, \ y \in \mathbb{R}^k. \tag{3.3.2}$$

It should be noted that the number of clusters k is not always known a priori and this number should be specified before solving Problem (3.3.2). There are various algorithms exist in the literature for solving the problem (3.3.2) [14, 15, 16, 18, 25, 85].

3.3.2 Späth algorithm for clusterwise linear regression

The Späth algorithm [85] for solving Problem (3.3.2) is based on the well known k-means algorithm. The algorithm was described for $p = 2$, however, in our description below we will present it for $p \geq 1$ in general.

The main and most time-consuming step in Algorithm 1 is Step 2 where one solves the linear regression problems to find regression coefficients. This is a convex optimization problem. It is a quadratic programming problem when $p = 2$. More

Algorithm 1 Späth algorithm

1: (Initialization) Select mutually disjoint clusters $A^1, \cdots, A^k$ such that $\bigcup_{j=1}^k A^j = A$.

2: For $j = 1, \cdots, k$, solve the following linear regression problem:

$$\text{minimize} \;\; \varphi(x^j, y_j) = \sum_{(a,b)\in A^j} E_{ab}(x^j, y_j) \;\; \text{subject to} \;\; x^j \in \mathbb{R}^n, \;\; y_j \in \mathbb{R}. \quad (3.3.3)$$

and obtain regression coefficients $(x^j, y_j), \; j = 1, \ldots, k$.

3: For $j = 1, \cdots, k$, recompute the cluster A^j such that a point $(a, b) \in A$ belongs to A^j if

$$E_{ab}(x^j, y_j) = \min_{l=1,\ldots,k} E_{ab}(x^l, y_l)$$

and go to Step 2 until no more data points change their clusters.

specifically, it is a classical least squares regression problem.

Algorithm 1 converges to local minimizers of Problem (3.3.2) whereas only global or near global solutions provide meaningful clusters. As the number of clusters k and the number of data points m increase, the number of local solutions to the Clusterwise linear regression increases drastically. The quality of the solution obtained by the algorithm depends on the initial set of clusters. For a moderate number of clusters in small data sets, a multi-restarting strategy can be applied. However, the success of this strategy deteriorates as the number of clusters or the size of the data set increase.

We propose to use an incremental approach to solving Problem (3.3.2). This approach iteratively adds one linear function at a time and uses the current clusters to construct a good starting cluster distribution for the next iteration. Incremental algorithms are increasingly popular in data mining and in particular to solve clustering problems [8, 12, 10, 47]. For example, the modified global k-means algorithm has been shown to be very efficient for solving clustering problems (see, [8, 9]).

3.3.3 Nonsmooth nonconvex optimization approach for solving the clusterwise linear regression problems

The brief description of the CLR problem and an algorithm for its solution is given below. The detailed description can be found in [14, 15, 16]. In what follows, we denote by $\mathbb{R}^n$ an n-dimensional Euclidean space with an inner product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ and associated norm $\|x\| = \langle x, x\rangle^{1/2}, x, y \in \mathbb{R}^n$.

The problem (3.3.2) is nonsmooth and nonconvex when $k \geq 2$. The discrete gradient method introduced in [11] is applied to solve it. It is an efficient method for solving nonsmooth optimization problems. This method does not require the calculation of subgradients of the objective function. The problem (3.3.2) has a special structure called piecewise partial separability. Therefore the version of the discrete gradient method given in [13] is applied to solve it. This method is a local search method and can find only local solutions to Problem (3.3.2). However, this problem is nonconvex optimization problem which means that it may have many local solutions and only global or near global solutions are of interest. Such solutions provide the best approximation of data with least number of linear functions.

Conventional global optimization techniques are not always efficient for solving such problems due to their size. Since we use the local search algorithm for solving Problem (3.3.2), it is important to develop a special procedure to generate initial solutions to computing global or near global solutions. To develop such a procedure, an incremental approach is proposed in [16]. In this approach linear functions are computed incrementally starting from one linear function and adding one linear function at each iteration of the incremental algorithm. The use of the incremental approach allows one to design an efficient algorithm for generating initial solutions. This is done by introducing the so-called auxiliary CLR problem.

Given the solution $(x^1, y_1, \cdots, x^{k-1}, y_{k-1})$ to the $(k-1)$-CLR problem (3.3.2) we define the regression error of the data point $(a, b) \in A$ at the $(k-1)$-th iteration by

$$r^{ab}_{k-1} = \min_{j=1\ldots k-1} E_{ab}(x^j, y_j)$$

and introduce the following function

$$\bar{f}_k(u, v) = \sum_{(a,b)\in A} \min\{r^{ab}_{k-1}, E_{ab}(u, v)\}, \;\; u \in \mathbb{R}^n, \;\; v \in \mathbb{R}. \tag{3.3.4}$$

The function $\bar{f}_k$ is called *the k-th auxiliary clusterwise linear regression function.* This function is nonsmooth and in general, nonconvex. Clearly,

$$\bar{f}_k(u, v) = f_k(x_1, y^1, \cdots, x_{k-1}, y^{k-1}, u, v), \forall u \in \mathbb{R}^n, v \in \mathbb{R}.$$

Furthermore

$$\max_{(u,v)\in\mathbb{R}^{n+1}} \bar{f}_k(u, v) = \sum_{(a,b)\in A} r^{ab}_{k-1} = \min_{x\in\mathbb{R}^{(k-1)n}, y\in\mathbb{R}^{k-1}} f_{k-1}(x, y). \tag{3.3.5}$$

The problem:

$$\text{minimize } \bar{f}_k(u, v) \text{ subject to } u \in \mathbb{R}^n, v \in \mathbb{R} \tag{3.3.6}$$

is called *the k-th auxiliary clusterwise linear regression problem.* This problem has $n+1$ variables and unlike Problem (3.3.2) the number of variables does not depend on the number of linear regression functions.

3.3.4 Computation of initial solutions

Given the solution $(x^1, y_1, \cdots, x^{k-1}, y_{k-1})$ to the $(k-1)$-CLR problem (3.3.2) consider the following set of hyperplanes:

$$C_k = \left\{(u,v) \in \mathbb{R}^{n+1} : \ E_{ab}(u,v) > r^{ab}_{k-1} \ \forall (a,b) \in A\right\}.$$

The set C_k contains all hyperplanes which do not attract any point from the set A. It is clear that over this set the function $\bar{f}_k$ is constant and reaches its global maximum value (3.3.5). Therefore any hyperplane from the set C_k is a stationary point for the function $\bar{f}_k$. This means that if we choose a starting point in this set then most local methods will be unable to escape it and will not decrease the value of both the auxiliary and overall fit functions. Therefore it is crucial to select starting points in the complementary set $\overline{C}_k$ of the closure of C_k:

$$\overline{C}_k = \left\{(u,v) \in \mathbb{R}^{n+1} : \ \exists \ (a,b) \in A \text{ such that } E_{ab}(u,v) < r^{ab}_{k-1}\right\}.$$

Clearly, $\overline{C}_k$ is the set of hyperplanes which attract at least one data point from the set A. Any hyperplane from this set will decrease the value of the auxiliary CLR function. Our aim is to find hyperplanes which provide significant decrease of the value of this function. Next we describe how such hyperplanes can be found.

Let $\bar{A}_0$ be a set of all data points $(a,b) \in A$ which do not lie on any of hyperplanes $(x_1, y^1), \ldots, (x_{k-1}, y^{k-1})$. If $\bar{A}_0 = \emptyset$ then the hyperplanes $(x_1, y^1), \cdots, (x_{k-1}, y^{k-1})$ perfectly approximate the set A. Therefore we assume that $\bar{A}_0 \neq \emptyset$. Take any $(a,b) \in \bar{A}_0$. Assume that this point belongs to the cluster determined by the linear regression coefficients $\{x^j, y_j\}$ where $j \in \{1, \ldots, k-1\}$. Define another hyperplane (x^{ab}, y_{ab}) parallel to the hyperplane (x^j, y_j) passing through the point (a,b). Then $x^{ab} = x^j$ and $y_{ab} = b - \langle x^j, a\rangle$. It is clear that $(x^{ab}, y_{ab}) \in \overline{C}_k$ for all $(a,b) \in \bar{A}_0$. The value $\tilde{f}_{k-1}$ of the function f_{k-1} over A with hyperplanes $(x_1, y^1, \cdots, x_{k-1}, y^{k-1})$ is:

$$\tilde{f}_{k-1} = \sum_{(a,b)\in A} r^{ab}_{k-1}$$

and the value $\tilde{f}_k$ of the function f_k over A with hyperplanes $(x_1, y^1, \cdots, x_{k-1}, y^{k-1})$ and (x^{ab}, y_{ab}) is:

$$\tilde{f}_k = \bar{f}_k(x^{ab}, y_{ab}) = \sum_{(c,d)\in A} \min\{r^{cd}_{k-1}, E_{cd}(x^{ab}, y_{ab})\}.$$

The difference between these two values is:

$$d(x^{ab}, y_{ab}) = \tilde{f}_{k-1} - \tilde{f}_k = \sum_{(c,d)\in A} \max\{0, r_{k-1}^{cd} - E_{cd}(x^{ab}, y_{ab})\}.$$

$d(x^{ab}, y_{ab}) > 0$ for any data point $(a, b) \in \bar{A}_0$. Let $\gamma_1 \in [0, 1]$ be a given number. Let

$$\bar{d}_1 = \max\{d(x^{ab}, y_{ab}) : \ (a, b) \in \bar{A}_0\} \tag{3.3.7}$$

and the set

$$\bar{A}_1 = \{(a, b) \in A : d(x^{ab}, y_{ab}) \geq \gamma_1 \bar{d}_1\}. \tag{3.3.8}$$

This set contains all the solutions providing decrease above a threshold $\gamma_1 \bar{d}_1$. For $\gamma_1 = 0$ the set $\bar{A}_1 = \bar{A}_0$ and for $\gamma_1 = 1$ the set $\bar{A}_1$ contains only data points providing largest decrease $\bar{d}_1$ of the k-th CLR function.

For each $(a, b) \in \bar{A}_1$ compute the set B_{ab} as follows:

$$B_{ab} = \left\{(c, d) \in A : \quad E_{cd}(x^{ab}, y_{ab}) < r_{k-1}^{cd}\right\}. \tag{3.3.9}$$

The set B_{ab} contains all points from the set A attracted by the clusterwise linear regression function (x^{ab}, y_{ab}). We compute $(\bar{x}^{ab}, \bar{y}_{ab})$ as a linear regression function approximating the set B_{ab}. This additional step to update the clusterwise linear regression function (x^{ab}, y_{ab}) allows one to improve an initial solution determined by the cluster B_{ab}.

Now we can define the following set of hyperplanes:

$$\bar{A}_2 = \left\{(u, v) : \ u \in \mathbb{R}^n, \ v \in \mathbb{R} \text{ and } \exists (a, b) \in \bar{A}_1 \text{ s.t. } u = \bar{x}^{ab}, v = \bar{y}_{ab}\right\}. \tag{3.3.10}$$

The set $\bar{A}_2$ contains all hyperplanes computed using points $(a, b) \in \bar{A}_1$. Next we compute the value $\hat{f}_k(u, v)$ of the overall fit function f_k over A with hyperplanes $(x_1, y^1, \cdots, x_{k-1}, y^{k-1})$ and $(u, v) = (\bar{x}^{ab}, \bar{y}_{ab})$:

$$\hat{f}_k(u, v) = \bar{f}_k(\bar{x}^{ab}, \bar{y}_{ab}) = \sum_{(c,d)\in A} \min\{r_{k-1}^{cd}, E_{cd}(\bar{x}^{ab}, \bar{y}_{ab})\}$$

and the following number:

$$\hat{f}_{k,min} = \min\left\{\hat{f}_k(u, v) : \ (u, v) \in \bar{A}_2\right\} \geq 0. \tag{3.3.11}$$

Let $\gamma_2 \in [1, \infty[$ be a given number. Define the following set

$$\bar{A}_3 = \{(u, v) \in \bar{A}_2 : \hat{f}(u, v) \leq \gamma_2 \hat{f}_{k,min}\}. \tag{3.3.12}$$

All hyperplanes from the set $\bar{A}_3$ are considered as an initial solution to solve Problem (3.3.6). The number $\gamma_2 \hat{f}_{k,min}$ is defined as a threshold and if the value of the auxiliary CLR function at $(u, v) \in \bar{A}_2$ is greater than this threshold this hyperplane is not considered as a "promising" to be an initial solution to minimize the auxiliary CLR function, since the value of this function at this initial solution is significantly larger than its best value. If $\gamma_2 = 1$ then hyperplanes from $\bar{A}_2$ with the lowest value of the auxiliary CLR function are chosen and if γ_2 is sufficiently large then $\bar{A}_3 = \bar{A}_2$.

Thus, an algorithm for finding good initial solutions for solving Problem (3.3.6) can be summarized as follows:

Algorithm 2 An algorithm for finding initial solutions to solve Problem (3.3.6).

1: (Initialization) Select the numbers $\gamma_1 \in [0, 1]$ and $\gamma_2 \in [1, \infty[$.
2: Determine the set $\bar{A}_0$ and compute the number $\bar{d}_1$ using (3.3.7).
3: Compute the set $\bar{A}_1$ using (3.3.8).
4: For each $(a, b) \in \bar{A}_1$ compute the set B_{ab} using (3.3.9), update the clusterwise regression functions (x^{ab}, y_{ab}) and compute the set $\bar{A}_2$ applying (3.3.10).
5: Compute the number $\hat{f}_{k,min}$ using (3.3.11) and the set $\bar{A}_3$ using (3.3.12). Any hyperplane $(u, v) \in \bar{A}_3$ is an initial solution to solve Problem (3.3.6).

Next we describe an algorithm for solving the auxiliary CLR problem (3.3.6). For a given hyperplane (u, v) define the following set:

$$\mathcal{B}(u, v) \quad = \quad \{(a, b) \in A : \ E_{ab}(u, v) < r_{k-1}^{ab}\},$$

The set $\mathcal{B}(u, v)$ contains all points from the set A which are attracted by the linear regression function (u, v). It is obvious that $\mathcal{B}(u, v) \neq \emptyset$ for all $(u, v) \in \bar{A}_3$.

Algorithm 3 An algorithm for solving Problem (3.3.6).

1: (Initialization) Select numbers $\gamma_1 \in [0, 1]$, $\gamma_2 \in [1, \infty[$ and apply Algorithm 2 to compute the set $\bar{A}_3$.
2: Select the initial linear regression function $(u^0, v_0) \in \bar{A}_3$, compute the set $\mathcal{B}(u^0, v_0)$ and set $l := 0$.
3: Solve the following linear regression problem:

$$\text{minimize} \ \ \varphi(u, v) = \sum_{(a,b) \in \mathcal{B}(u^l, v_l)} E_{ab}(u, v) \text{ subject to } u \in \mathbb{R}^n, \ v \in \mathbb{R} \qquad (3.3.13)$$

and obtain regression coefficients $(\tilde{u}^l, \tilde{v}_l)$.
4: Compute the set $\mathcal{B}(\tilde{u}^l, \tilde{v}_l)$.
5: (Stopping criterion) If $\mathcal{B}(\tilde{u}^l, \tilde{v}_l) = \mathcal{B}(u^l, v_l)$, then set $(\bar{u}, \bar{v}) := (u^l, v_l)$ and stop. $(\bar{u}, \bar{v})$ is a solution to Problem (3.3.6).
6: Otherwise set $u^{l+1} := \tilde{u}^l, v_{l+1} := \tilde{v}_l$, $\mathcal{B}(u^{l+1}, v_{l+1}) := \mathcal{B}(\tilde{u}^l, \tilde{v}_l)$, $l := l + 1$ and go to Step 3.

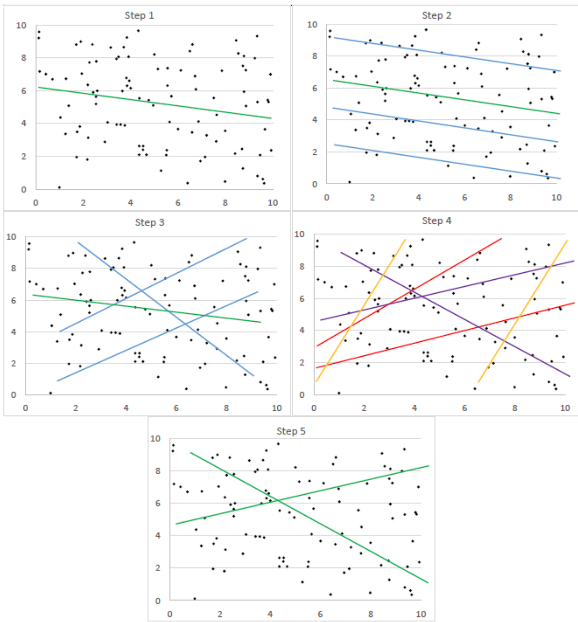

Figure 3.2: Illustration of the incremental CLR algorithm.

3.3.5 Incremental algorithm

Now we are ready to design an incremental algorithm for solving Problem (3.3.2). We use the solution $(\bar{u}, \bar{v})$ found by Algorithm 3 to generate an initial solution to this problem. Let $\gamma_3 \in [1, \infty)$ be a given number. Take any $(u, v) \in \bar{A}_3$ as an initial solution and apply Algorithm 3 starting from this solution. As a result we get a regression function $(\bar{u}, \bar{v})$ which is the local minimizer of the auxiliary CLR problem (3.3.6). We denote by $\bar{A}_4$ the set of all solutions obtained by Algorithm 3 starting from some $(u, v) \in \bar{A}_3$. Define the number

$$\bar{f}_{k,min} = \min \left\{ \bar{f}_k(\bar{u}, \bar{v}) : \ (\bar{u}, \bar{v}) \in \bar{A}_4 \right\}$$

and the set

$$\bar{A}_5 = \left\{ (\bar{u}, \bar{v}) \in \bar{A}_4 : \ \bar{f}_k(\bar{u}, \bar{v}) \le \gamma_3 \bar{f}_{k,min} \right\}.$$

Notice that if $\gamma_3 = 1$ then only best local minimizers of Problem (3.3.6) found by Algorithm 3 are chosen. If γ_3 is sufficiently large then all local minimizers from the set $\bar{A}_4$ are chosen and $\bar{A}_5 = \bar{A}_4$. Summarizing we can design an incremental algorithm as follows:

Algorithm 4 An incremental algorithm for solving Problem (3.3.2)

1: (Initialization) Select numbers $\gamma_1 \in [0,1],\ \gamma_2 \in [1,\infty)$ and $\gamma_3 \in [1,\infty)$. Compute the linear regression function $(x^1, y_1) \in \mathbb{R}^n \times \mathbb{R}$ of the whole set A. Set $l := 1$.

2: (Computation of the next linear regression function). Set $l := l+1$. Let $(x^1, y_1, \cdots, x^{l-1}, y_{l-1})$ be the solution to the $(l-1)$-CLR problem. Apply Algorithm 3 starting from each $(u,v) \in \bar{A}_3$ to find a set of solutions $\bar{A}_5$ to the l-th auxiliary CLR problem (3.3.6).

3: (Refinement of all linear regression functions). For each $(\bar{u},\bar{v}) \in \bar{A}_5$ select $(x^1, y_1, \cdots, x^{l-1}, y_{l-1}, \bar{u}, \bar{v})$ as an initial solution and apply the discrete gradient method to solve the l-CLR and compute a set $\bar{A}_6$ of solutions $(\bar{x}^1, \bar{y}_1, \cdots, \bar{x}^l, \bar{y}_l)$ to this problem.

4: (Computation of the solution) Choose any

$$(\hat{x}^1, \hat{y}_1, \cdots, \hat{x}^l, \hat{y}_l) \in \operatorname{Argmin} \left\{ f_l(\bar{x}^1, \bar{y}_1, \cdots, \bar{x}^l, \bar{y}_l) :\ (\bar{x}^1, \bar{y}_1, \cdots, \bar{x}^l, \bar{y}_l) \in \bar{A}_6 \right\}$$

as a solution to the l-CLR problem. Set $x^j := \hat{x}^j, y_j := \hat{y}_j,\ j = 1, \ldots, l$.

5: (Stopping criterion) If $l = k$, then stop. Otherwise go to Step 2.

It is easy to notice that Algorithm 4 finds solutions to all l-CLR problems where $l = 1, \ldots, k$. The graphical illustration of Algorithm 4 is given in Figure 3.2, where $k = 2$. In Step 1, one linear function (green line) is computed to approximate the whole data. Initial solutions (blue lines) in Step 2 are defined as parallel lines to the line computed in Step 1 and passing through at least one data point. These lines are used to find the set of solutions to the auxiliary CLR in Step 3 (blue lines). In Step 4, the solution to the 1-CLR problem (green line) is combined with solutions to the auxiliary CLR to construct the set of initial solutions to the 2-CLR problem. Solutions (pairs of red, yellow and pink lines) to the 2-CLR problem are found starting from each of these initial solutions. Finally, in Step 5, the best pair of lines (the pair of green lines) is chosen as a solution to the 2-CLR problem.

3.4 Prediction methods based on clusterwise linear regression

The main hypothesis of clusterwise regression is that the observations come from more than one cluster with unknown proportion. The optimal number of clusters with common characteristics are determined using the above proposed algorithm and fit the linear regression functions within each cluster. To obtain prediction, the following methods explored.

Let $T = \{(a^i, b_i)\}, i = 1, \ldots, m$ be a training data. Applying the CLR method we compute k clusters and corresponding regression coefficients $(x^j, y_j), j = 1, \ldots, k$. For a new observation $a \in \mathbb{R}^n$, the prediction value u_a can be compute as:

1. **Using the largest cluster.** In this prediction method, we select the largest j-th cluster and compute the prediction value u_a for new observation a using the corresponding regression coefficients (x^j, y_j) of the j-th cluster:

$$u_a = \langle x^j, a \rangle + y_j, \ \ j = 1, \dots, k.$$

2. **Using weights.** In this prediction method, we assign weights to each cluster. The weight w_j of the j-th cluster is computed as $w_j = l_j/l$, where l_j is the number of points in the j-th cluster and l is the total number of points in the training set. Compute predictions for new observation a from each cluster using the corresponding regression coefficients (x^j, y_j).

$$z_j = \langle x^j, a \rangle + y_j, \ \ j = 1, \dots, k.$$

   Then calculate the final prediction value u_a for a as:

$$u_a = \sum_{j=1}^{k} w_j z_j.$$

3. **Using neighbours.** In this prediction method, we select any positive integer p and compute p closest points to the new observation a in the input space from the training set. Denote the set of these p points as: $P = \{(a^1, b_1), \dots, (a^p, b_p)\}$. Next calculate the cluster distribution of training points from the set P. This means that we get the following numbers: $n_1, \dots, n_k$ where $n_i - s$ are the number of points from the set P belonging to the i-th cluster and $n_i \geq 0, i = 1, \dots, k$.

$$\sum_{i=1}^{k} n_i = p$$

   In next step calculate weights of each cluster as follows:

$$w_j = \frac{n_j}{p}.$$

   Then compute predictions $z_j = \langle x^j, a \rangle + y_j$ for new observation a from each cluster and calculate the final prediction value u_a as:

$$u_a = \sum_{j=1}^{k} w_j z_j.$$

4. **Using distance.** In this prediction method, we calculate the distance between

the new observation a and the clusters in the input space from the training set.

$$d_j = \|a - c_j\|,$$

where c_j is the centred value of the j-th cluster and $j = 1, \ldots, k$.

Then select p closest clusters to the new observation a based on the above calculated distance and assign weights as follow:

$$w_i = \frac{d_i^{-h}}{\sum_{i=1}^{p} d_i^{-h}},$$

where $i = 1, \ldots, p$ and $h = 2$.

In the next step compute predictions for new observation a from each selected p clusters using the corresponding regression coefficients (x^j, y_j).

$$z_i = \langle x^i, a \rangle + y_i, \;\; i = 1, \ldots, p.$$

Then calculate the final prediction value u_a for a as a distance weighted average:

$$u_a = \sum_{i=1}^{p} w_i z_i.$$

3.5 Summary of chapter

To solve a clusterwise linear regression problem, many approaches exist in the literature including statistical, data mining and optimization techniques. These existing techniques are very sensitive to the choice of an initial solution, and may lead to sub-optimal solutions. To solve this problem, we proposed a new method based on non-smooth, non-convex optimization by modifying the existing ones. In the proposed method we used an affine function to represent clusters instead of classical centers.This method allows one to find a global, or near global, solution to the problem. We also explored new prediction methods based on clusterwise regression to improve the prediction accuracy.

Chapter 4

Data, implementation and evaluation of models

In this chapter, we provide a description of rainfall data and meteorological variables, correlation between rainfall and other meteorological variables, climate zones and seasonal patterns in Australia, geographic details and descriptive statistics of rainfall data. Following that, we discuss the implementation of models and the evaluation of each model's performance in predicting rainfall. We also provide information about the performance measures used for the purpose of evaluation.

4.1 Rainfall and input meteorological variables

Historical daily and monthly rainfall data have been taken from the Scientific Information for Land Owners (SILO) available at [83]. SILO is an enhanced climate database hosted by the Queensland Government Department of Science, Information Technology, Innovation and the Arts. The data is reliable and quality checked. In the development of our prediction models, we used the data of six meteorological parameters from 24 geographically diverse weather stations around Australia for the period January 1970 through to December 2014. From each state/territory (Victoria, New South Wales, Queensland, South Australia, Western Australia and Northern Territory), we selected four weather stations. We also considered the climate zones when choosing these weather stations because Australia experiences a variety of climates due to its size.

In Australia, there are six major climate zones:(temperate, grassland, desert, tropical, subtropical and equatorial) and two main seasonal patterns (Summer/Autumn/Winter/Spring and wet/dry) [38]. We do not consider the equatorial zone as its area is insignificant. The temperate, grassland and desert climate zones are affected by the seasonal pattern Summer/Autumn/Winter/Spring and tropical, subtropical and equatorial zones by the wet/dry seasonal pattern. The Summer season in Australia is from December to February, Autumn from March to May, Winter from June to August and Spring from September to November. The wet season, also called the monsoon season, starts in November and finishes in March and the dry seasons starts in April and finishes in October. Figure 4.1 illustrates both seasonal patterns using the rainfall data from climate zones.

Using the Bureau of Meteorology climate classification, we selected two locations from the tropical zone, two from the subtropical, five from the desert, seven from the temperate and eight from the grassland zones. The Bureau of Meteorology classification is a modified Köppen climate classification. The Köppen classification system is the most widely used classification system with several modifications from 1900 to the present. The Bureau of Meteorology classification is derived from 0.025×0.025 degree resolution mean rainfall, mean maximum temperature and mean minimum temperature gridded data. All means are based on a standard 30-year climatology (from 1961 to 1990) [62].

For monthly rainfall predictions, a total 540 records from Jan 1970 to Dec 2014 were used. Of them, 360 records for the 30 years period from Jan 1970 to Dec 1999 were used to train the models; the remaining 180 records over a 15 year period, from Jan 2000 to Dec 2014, were used to test the models.

For weekly rainfall predictions, a total of 2356 records from Jan 1970 to Feb 2015 were used. Of these, 1834 over a 35 year period, from Jan 1970 to Feb 2005, were used to train the models; the remaining 522 records over a ten year period from Feb 2005 to Feb 2015, were used to test the data.

The average monthly rainfall varies across these sites from 15.07 mm to 125.87 mm. The geographic details of these weather stations are given in Table 4.1 and a location map based on a modified Köppen classification system is given in Figure 4.2.

The set of meteorological parameters used in this study are: (1) monthly rainfall; (2) maximum temperature (TMax); (3) minimum temperature (TMin); (4) evaporation (Evap); (5) vapor pressure (VP); and (6) solar radiation (Rad). All these parameters are important components of the hydrologic cycle. For example, evaporation is the process by which liquid water is converted to water vapor via the transfer of water molecules to the atmosphere. Solar radiation is the energy source used to change liquid water into water vapor. The vapor pressure is the pressure exerted by water vapor molecules in the air. Temperature is a measure of the ability of the atmosphere and water to receive and transfer heat from other bodies. Temperature directly affects evaporation. These meteorological parameters are interdependent and influence precipitation.

Correlations between rainfall and five input meteorological variables for each weather station including significant test results are presented in Table 4.2. The P-value higher

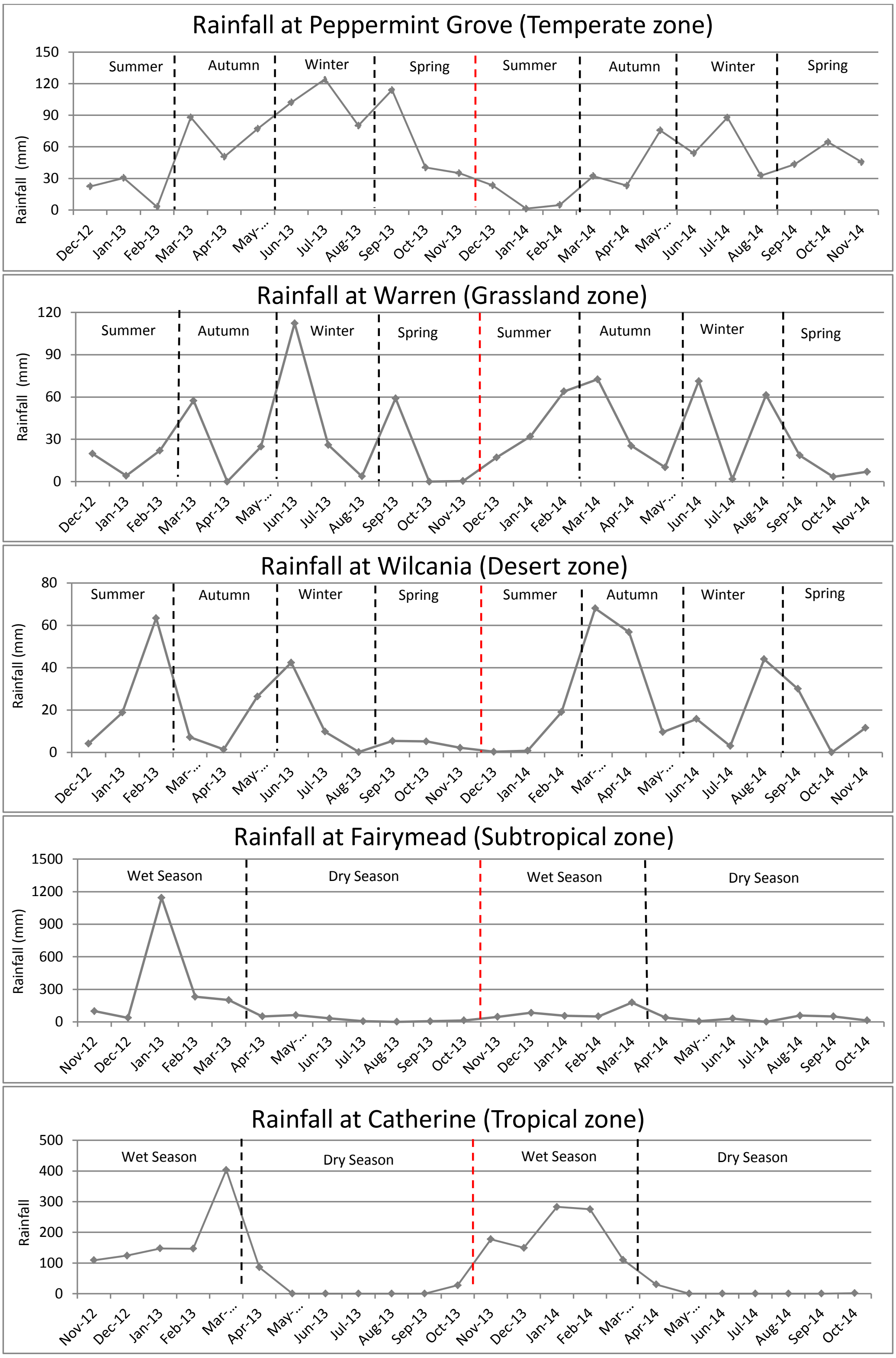

Figure 4.1: Seasonal patterns in climate zones in Australia.

Station name	Station Id	Classification	Latitude	Longitude	Elev.	Min.	Max.	Mean
Victoria								
Annuello	76000	Grassland	-34.85	142.78	52	0.00	261.40	27.20
Dookie	81013	Temperate	-36.37	145.70	185	0.00	227.60	47.23
Orbost	84030	Temperate	-37.69	148.46	41	1.90	425.00	72.06
Cape Otway	90015	Temperate	-38.86	143.51	82	1.60	218.20	77.80
New South Wales								
Warren	51034	Grassland	-31.50	147.69	192	0.00	258.80	41.49
Yamba	58012	Subtropical	-29.43	153.36	27	0.10	629.40	125.87
Moss Vale	68045	Temperate	-34.54	150.38	675	0.40	527.00	75.60
Wilcannia	46043	Desert	-31.56	143.37	75	0.00	252.30	24.31
Queensland								
Palmerville	28004	Tropical	-16.00	144.08	203	0.00	813.20	89.69
Richmond	30045	Grassland	-20.73	143.14	211	0.00	664.20	42.46
Boulia	38003	Desert	-22.91	139.90	161	0.00	464.90	22.74
Fairymead	39037	Subtropical	-24.79	152.36	5	0.00	1143.40	88.22
Northern Territories								
Katherine	14902	Tropical	-14.46	132.26	106	0.00	937.70	91.43
Newery	14820	Grassland	-16.05	129.26	101	0.00	810.00	69.17
Henbury	15552	Desert	-24.55	133.25	432	0.00	376.10	22.68
Alexandria	15088	Grassland	-19.06	136.71	274	0.00	565.40	37.76
South Australia								
Marree	17031	Desert	-29.65	138.06	50	0.00	203.30	15.07
Blinman	17014	Grassland	-31.09	138.68	615	0.00	294.20	28.86
Koppio	18043	Temperate	-34.41	135.82	173	0.00	204.00	42.92
Port Elliot	23742	Temperate	-35.53	138.69	10	0.00	152.60	41.14
Western Australia								
Ningaloo	5020	Grassland	-22.70	113.67	10	0.00	293.60	20.85
Wiluna	13012	Desert	-26.59	120.23	521	0.00	271.60	24.29
Dowerin	10042	Grassland	-31.19	117.03	273	0.00	171.20	28.52
Peppermint Grove	9594	Temperate	-34.44	119.36	60	0.00	308.60	58.75

Table 4.1: Geographic details, elevation (m.), minimum, maximum and average rainfall of weather stations

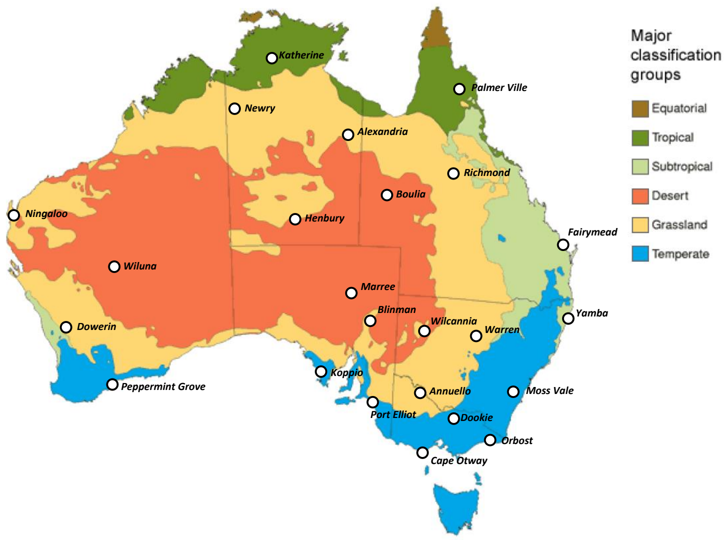

Figure 4.2: Location map of the study area (based on a modified Köppen classification system).

than 0.05 refer to statistically not significant, less than 0.05 statistically significant and less than 0.001 statistically highly significant.

One can see that, there is a highly significant correlation between rainfall and meteorological variables in most climate zones (see p-values in blue color in Table 4.2.In the desert climate zone, correlations between rainfall and evaporation, solar radiation and maximum temperature is not significant. These results justifies the use of meteorological variables for rainfall prediction.

Station name	TMax	TMin	Evap.	VP	Rad.
Temperate					
Dookie	-0.23(4.5e-08)	-0.04(0.31)	-0.26(5.7e-10)	0.12(0.004)	-0.25(4.7e-9)
Orbost	-0.17(4.5e-08)	0.01(0.88)	-0.11(0.01)	0.02(0.57)	-0.12(0.004)
Cape Otway	-0.17(2.2e-16)	0.01(2.2e-16)	-0.11(2.2e-16)	0.02(2.2e-16)	-0.12(2.2e-16)
Moss Vale	-0.02(0.69)	0.16(0.0002)	-0.09(0.03)	0.22(1.7e-7)	-0.15(0.0007)
Koppio	-0.67(2.2e-16)	-0.54(2.2e-16)	-0.63(2.2e-16)	-0.40(2.2e-16)	-0.632(2.2e-16)
Port Elliot	-0.64(2.2e-16)	-0.54(2.2e-16)	-0.61(2.2e-16)	-0.42(2.2e-16)	-0.58(2.2e-16)
Peppermint Grove	-0.62(2.2e-16)	-0.48(2.2e-16)	-0.60(2.2e-16)	-0.39(2.2e-16)	-0.57(2.2e-16)
Grassland					
Annuello	-0.09(0.03)	0.07(0.11)	-0.12(0.004)	0.32(1.4e-14)	-0.11(0.01)
Warren	0.04(0.39)	0.22(1.7e-7)	-0.03(0.54)	0.45(2.2e-16)	-0.03(0.4)
Richmond	0.20(3.4e-16)	0.47(2.2e-16)	-0.05(0.26)	0.66(2.2e-16)	-0.12(0.005)
Newry	0.20(2.2e-16)	0.56(2.2e-16)	-0.17(5.9e-5)	0.72(2.2e-16)	-0.38(2.2e-16)
Alexandra	0.20(4.2e-6)	0.45(2.2e-16)	-0.08(0.07)	0.67(2.2e-16)	-0.22(3.8e-7)
Blinman	-0.15(3.9e-4)	-0.02(0.66)	-0.18(2.4e-5)	0.38(2.2e-16)	-0.19(8.8e-6)
Ningaloo	-0.23(1.1e-7)	0.00(0.92)	-0.35(2.2e-16)	0.22(1.8e-7)	-0.43(2.2e-16)
Dowerin	-0.52(2.2e-16)	-0.37(2.2e-16)	-0.51(2.2e-16)	-0.10(0.01)	-0.55(2.2e-16)
Desert					
Wilcannia	0.00(0.96)	0.16(1.4e-4)	-0.05(0.29)	0.52(2.2e-16)	-0.05(0.28)
Boulia	0.18(2.9e-5)	0.33(4.6e-15)	0.03(0.50)	0.61(2.2e-16)	-0.06(0.15)
Henbury	0.09(0.03)	0.25(2.9e-9)	0.01(0.77)	0.59(2.2e-16)	-0.06(0.20)
Marree	0.07(0.11)	0.21(9.2e-7)	0.03(0.44)	0.49(2.2e-16)	0.00(0.98)
Wiluna	0.08(0.07)	0.23(3.9e-8)	-0.01(0.89)	0.51(2.2e-16)	-0.12(4.5e-3)
Tropical and Subtropical					
Yamba	0.17(1.1e-4)	0.29(9.6e-12)	-0.14(9.1e-4)	0.35(2.2e-16)	-0.21(4.8e-7)
Fairymead	0.34(4.9e-16)	0.44(2.2e-16)	0.14(1.4e-3)	0.50(2.2e-16)	0.05(0.27)
Palmerville	0.05(0.24)	0.63(2.2e-16)	-0.26(1.7e-9)	0.75(2.2e-10)	-0.27(2.6e-10)
Katherine	0.13(1.8e-3)	0.55(2.2e-16)	-0.39(2.2e-16)	0.70(2.2e-16)	-0.57(2.2e-16)

Table 4.2: Correlation between monthly rainfall and input meteorological parameters

4.2 Implementation of models

In this section, we discuss the implementation of the CLR method as well as seven other prediction methods used for comparison: the clusterwise regression method based on EM algorithm (CR-EM), MLR, SVMreg and artificial neural networks with and without the hidden layer.

The Clusterwise linear regression, based on nonsmooth nonconvex optimization, was implemented in Fortran 95 and compiled using *g95* compiler.

Statistical package R-Version 3.2.2 is used to implement the CR-EM, MLR, ANN, and SVMreg models. R is a language and environment for statistical computing and graphics including time series analysis, clustering, classification, linear and nonlinear modeling and statistical tests (see: https://www.r-project.org/).

We used the R package *FlexMix* for CR-EM [39], *nnet* for ANN [88], *kknn* for k-NN [77] and *e1071* for SVMreg [54]. The package *FlexMix* implements a general framework for fitting finite mixtures of regression models in the R statistical computing environment using the EM algorithm. The *nnet* is a package used to implement feed-forward neural networks with a single hidden layer, and for multinomial log-linear models. We implemented neural networks with and without hidden layer. The package *kknn* is used to perform k-nearest neighbor classification and regression. In implementing the k-NN method, the most important step is the selection of the number of neighbours. We evaluated different values ranging from 1 to 12 and select the model with minimum root mean squared error value. The package *e1071* is a vital package in R language, used to implement SVMreg with linear, polynomial, radial basis function and sigmoid kernels. We fitted SVMreg with linear and RBF kernels.

All models were developed for each weather station using training data consisting of 360 records and evaluated by using test data consisting of 180 records. Numerical experiments were carried out on a PC with Processor Intel(R) Core(TM) i5-3437U CPU 1.90 GHz.

4.3 Statistical evaluation of model performance

Prediction performance of all models was evaluated by comparing observed and predicted rainfall using four measures of forecast accuracy calculated from the test datasets: the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Mean Absolute Scaled Error (MASE) and the Nash-Sutcliffe Efficiency (CE).

The RMSE, also called the root mean square deviation, is a measure of the difference between predicted values by a model and the observed values being modeled. In calculating the RMSE, the first step is to calculate the individual squared errors. Errors are the difference between the actual and predicted values. The next step is to average the squared errors, which yields the mean square error (MSE). The third and final step is to take the square root of the MSE.

The MAE, also referred to as the mean absolute deviation, is simply the average of the absolute errors. It is easy to understand and compute. The RMSE and MAE measures are widely used because of their theoretical relevance in statistical modeling. Both RMSE and MAE are scaled dependent measures. The use of absolute values or squared values prevents negative and positive errors from offsetting each other. It is well-known that the MAE is less sensitive to outliers than the RMSE [43]. The small values of the RMSE and the MAE indicates small deviations of the predictions from actual observations.

The MASE was proposed by Hyndman and Koehler in 2006 [43]. The MASE is a scaled error based on in-sample MAE from the naive forecast method. The $MASE < 1$ value indicates predictions are better than the mean of the observed data and $MASE > 1$ indicates predictions are worse than the mean of the observed data. The MASE is not applicable in the case where all observations are equal. It is an alternative to model performance measures based on percentage errors such as Mean Absolute Percentage Error (MAPE).

The Nash-Sutcliffe Efficiency (CE) was proposed by Nash and Sutcliffe in 1970 [58]. It is a normalized statistic that determines the relative magnitude of the residual variance and data variance. It is defined as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period under investigation. The CE can range from $-\infty$ to 1. An efficiency $CE = 1$ means a perfect prediction. An efficiency of 0 indicates that the model predictions are as accurate as the mean of the observed data and an efficiency $-\infty < CE < 0$ occurs when the observed mean is a better predictor than the model predictions.

Definitions of these measures follow. Assume that $Y_1, \ldots, Y_m,\ m \geq 1$ are observed values for some parameter Y, and $F_1, \ldots, F_m$ are their predicted values.

1. The RMSE is defined as

$$RMSE = \left(\frac{1}{m} \sum_{i=1}^{m} (F_i - Y_i)^2 \right)^{1/2};$$

2. The MAE is:

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |F_i - Y_i|;$$

3. The MASE is [43]:

$$MASE = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{|Y_i - F_i|}{\frac{1}{m-1} \sum_{i=2}^{m} |Y_i - Y_{i-1}|} \right)$$

4. The CE is defined as [58]:

$$CE = 1 - \left[\frac{\sum_{i=1}^{m} (Y_i - F_i)^2}{\sum_{i=1}^{m} (Y_i - \overline{Y})^2} \right].$$

$\overline{Y}$ is the mean of observed values.

4.4 Summary of chapter

In this chapter, we provided detailed information of study data, study map, geographic details and the correlation between rainfall and meteorological variables. Historical daily and monthly rainfall data with five input meteorological variables over the period of January 1970 to December 2014 from 24 geographically diverse weather stations in Australia were taken from SILO. Climate zones and seasonal patterns were discussed in detail.

We briefly discussed the implementation of models. All models were implemented in the statistical package R-Version 3.2.2, except the CLR(Opt) model. The CLR(Opt) was implemented in Fortran 95 and compiled using the g95 compiler. The performance of the models were evaluated using the statistical measures RMSE, MAE, MASE and CE.

Chapter 5

Monthly rainfall predictions

In this chapter, first we present the monthly rainfall prediction results of the CLR(Opt), CR(EM), MLR, SVMreg(RBF), SVMreg(Linear), ANN(0), ANN(1) and K-NN models within each classification zone then we assess the CLR(Opt) model performance comparing to other models. All models were developed using all combinations of input variables. There is total of fifteen combinations. The combinations and their notations are given in Table 5.1. The models were trained using data from Jan 1970 to Dec 1999 and tested using data from Jan 2000 to Dec 2014 with each combination of input variables in all 24 locations. In all cases, negative predicted values were adjusted to zero rainfall before performance measures calculated. The best combination of input variables for each model was determined using test data. The prediction performance of models with each combination was evaluated by comparing observed and predicted rainfall. The performance measure RMSE used as a primary measure to determine the best and worst combination of input variables for each model at each location. Then we provide a comparative assessment of models within each classification zone in predicting monthly rainfall. We also compare the predictions zone-wise to assess the influence of geographic regions on the performance of models.

The structure of the chapter is as follows. We present the monthly rainfall prediction results for temperate classification zone in Section 5.1; for grassland in Section 5.2; for desert in Section 5.3; and for tropical and subtropical in Section 5.4. We also provide the comparative assessment results of prediction models within each classification zone at the end of each section. Section 5.5 concludes the chapter.

5.1 Monthly rainfall predictions in temperate zone

In this section, first we present the monthly rainfall prediction results for each model in predicting monthly rainfall with best and worst combinations of input variables in temperate classification zone. Then we summarize the performance of all models

No.	Combination	No	Combination	No	Combination
C1	TMax, TMin	C6	TMax, TMin, VP	C11	TMax, TMin, Evap, VP
C2	Evap	C7	TMax, TMin, Rad	C12	TMax, TMin, Evap, Rad
C3	VP	C8	Evap, VP	C13	TMax, TMin, VP, Rad
C4	Rad	C9	Evap, Rad	C14	Evap, VP, Rad
C5	TMax, TMin, Evap	C10	VP, Rad	C15	TMax, TMin, Evap, VP, Rad

Table 5.1: Combinations of input parameters

with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models using computational results and time series plots.

Table 5.2, summarizes the prediction performance of the CLR(Opt) model with best and worst combinations of input variables in temperate classification zone. These results show that the model provides best predictions with the combination of input variables TMax, TMin and Rad in three out of seven locations (Mossvale, Port Elliot and Peppermint Grove); with TMax, TMin, VP and Rad in two locations (Koppio and Orbost); with TMax, TMin, Evap and Rad in the remaining two locations (Dookie and Cape Otway). The CLR(Opt) model provides the worst predictions with input variable VP in four locations (Koppio, Port Elliot, Orbost and Peppermint Grove); with Rad in two locations (Dookie and Cape Otway) and with Evap in one location (Mossvale).

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Rad	45.00	32.38	0.60	0.36
Koppio	TMax, TMin, VP, Rad	22.40	16.86	0.67	0.54
Port Elliot	TMax, TMin, Rad	19.37	14.66	0.67	0.53
Dookie	TMax, TMin, Evap, Rad	25.91	18.08	0.62	0.38
Orbost	TMax, TMin, VP, Rad	35.27	26.00	0.57	0.38
Cape Otway	TMax, TMin, Evap, Rad	33.41	25.59	0.71	0.36
Peppermint Grove	TMax, TMin, Rad	35.56	23.90	0.64	0.39
Performance measures for worst input combinations					
Mossvale	Evap	57.95	44.15	0.81	-0.06
Koppio	VP	30.31	23.32	0.93	0.15
Port Elliot	VP	24.08	18.85	0.86	0.27
Dookie	Rad	33.25	25.19	0.87	-0.02
Orbost	VP	45.33	34.36	0.75	-0.02
Cape Otway	Rad	35.97	27.93	0.78	0.25
Peppermint Grove	VP	42.28	29.63	0.79	0.14

Table 5.2: The CLR(Opt) model prediction performance with best and worst combination of input variables in temperate zone.

The performance measure RMSE for the CLR(Opt) model in predicting monthly rainfall ranges from 19.37 to 45.00, MAE from 14.66 to 32.38, MASE from 0.57 to 0.71 and CE from 0.36 to 0.54. The performance measures RMSE and MAE indicate the model provides best predictions in Port Elliot; MASE in Orbost and CE in Koppio while the performance measures RMSE and MAE indicate the model provides worst predictions in Mossvale and MASE and CE indicate in Cape Otway (see Figure 5.1). The graphical display of the model predictions and the observed rainfall is given in Figure 5.2. The figure shows that the model predictions follow the series patterns at all locations.

Table 5.3, summarizes the CR(EM) model performance with best and worst combinations of input variables in temperate classification zone. These results show that the model provides best predictions with the combination of input variables TMax and TMin at Koppio; with TMax, TMin, and Evap at Port Elliot; with TMax, TMin, Evap and VP at Orbost; with TMax, TMin, Evap and Rad at Dookie and Cape Otway; and with full set of input variables at Mossvale. The model produced worst predictions with the input variable VP in all locations except Mossvale. In Mossvale the model provides worst predictions with input variable Evap.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	46.14	33.37	0.62	0.33
Koppio	TMax, TMin	22.99	16.95	0.68	0.51
Port Elliot	TMax, TMin, Evap	19.86	15.05	0.69	0.51
Dookie	TMax, TMin, Evap, Rad	25.99	17.55	0.61	0.38
Orbost	TMax, TMin, Evap, VP	36.65	27.49	0.60	0.33
Cape Otway	TMax, TMin, Evap, Rad	33.40	25.58	0.71	0.36
Peppermint Grove	TMax, TMin, Evap, VP, Rad	36.49	24.59	0.66	0.36
Performance measures for worst input combinations					
Mossvale	Evap	58.67	44.15	0.81	0.00
Koppio	VP	30.72	24.19	0.97	0.13
Port Elliot	VP	24.74	19.25	0.88	0.23
Dookie	VP	33.37	25.78	0.89	-0.03
Orbost	VP	45.44	34.37	0.75	-0.02
Cape Otway	VP	36.50	28.64	0.80	0.23
Peppermint Grove	VP	42.61	30.28	0.81	0.13

Table 5.3: The CR(EM) model prediction performance with best and worst combination of input variables in temperate classification zone.

The performance measure RMSE for the CR(EM) model ranges from 19.86 to 46.14; MAE from 15.05 to 33.37; MASE from 0.60 to 0.71 and CE from 0.33 to 0.51. Three performance measures RMSE and MAE indicate the CR(EM) model provides best predictions in Port Elliot; MASE indicates in Orbost; and CE in Koppio while RMSE and CE indicate worst predictions in Mossvale; MASE indicates in Cape Otway; and CE indicates at Mossvale (see Figure 5.1). Time series plot of the observed rainfall and model predictions for each location in temperate zone given in Figure 5.3, shows that the model predictions follow the series patterns.

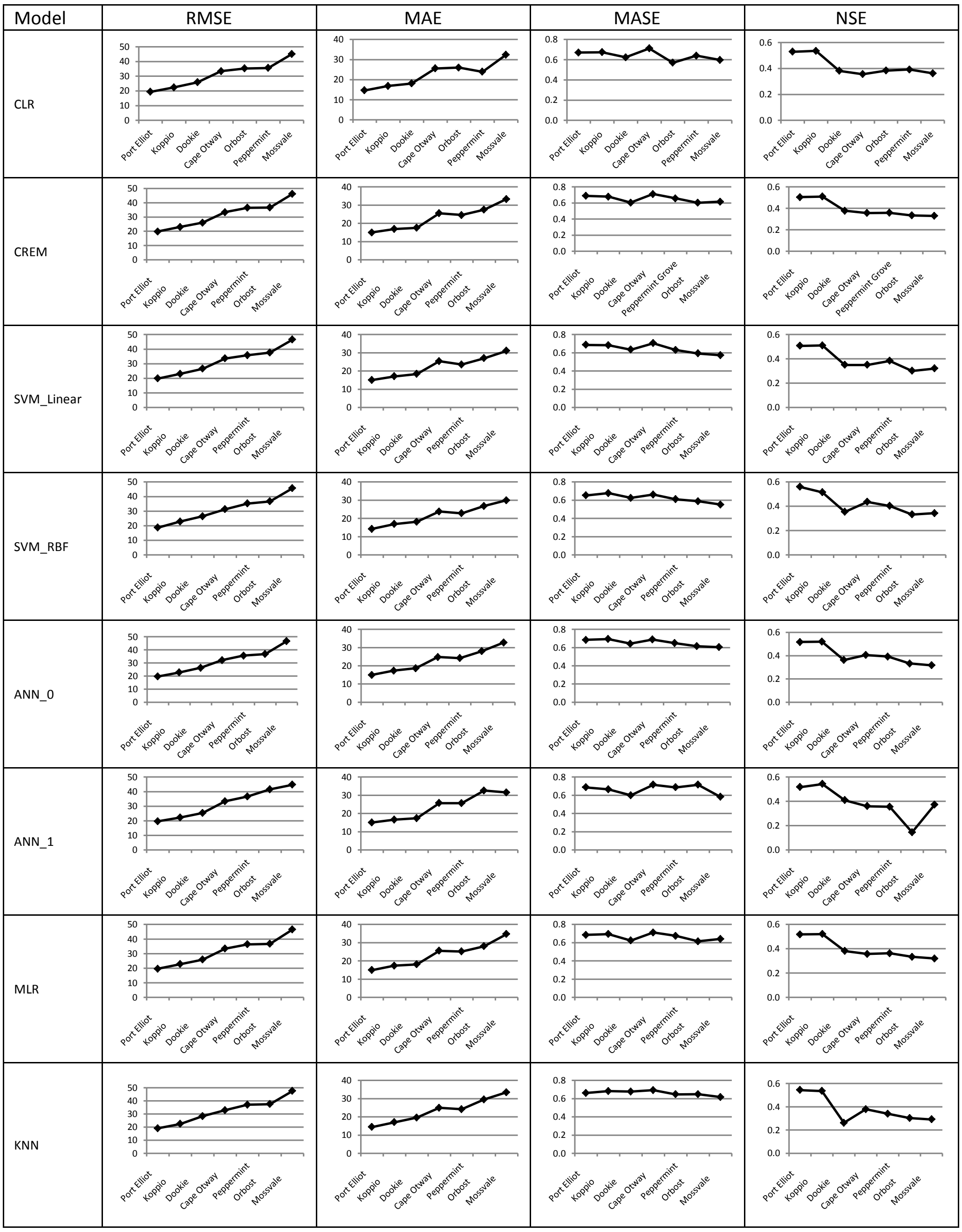

Figure 5.1: The prediction performance of models in temperate zone.

Table 5.4, summarizes the prediction performance of the SVM(Linear) model with best and worst combinations of input variables for locations in temperate classification zone. These results show that the model provides best predictions at Mossvale and Orbost with the combination of input variables TMax, TMin, Evap and Rad; at Port Elliot and Cape Otway with TMax, TMin, Evap and Rad; at Dookie with TMax,

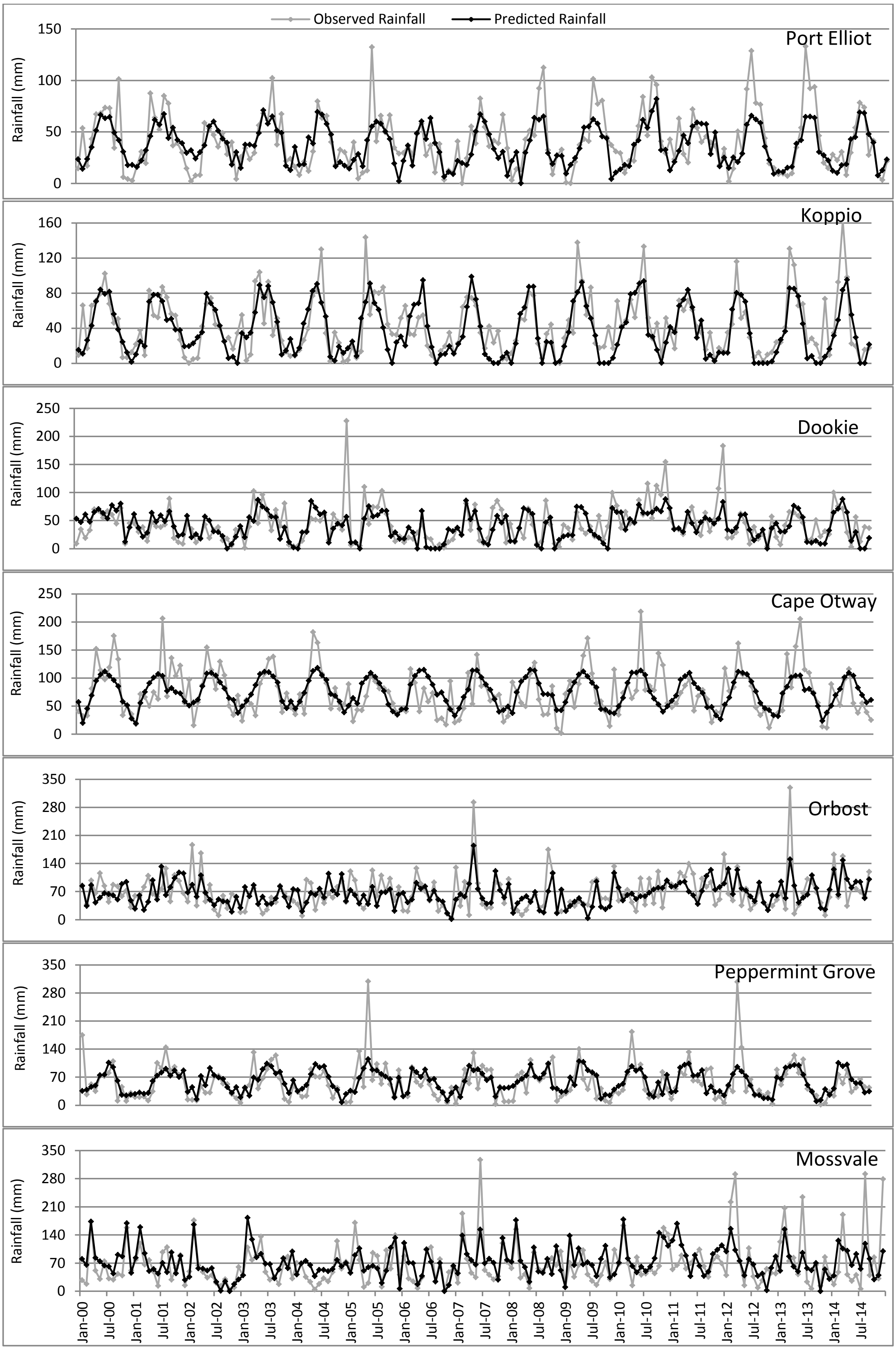

Figure 5.2: Observed rainfall vs. CLR(Opt) model predictions in temperate zone.

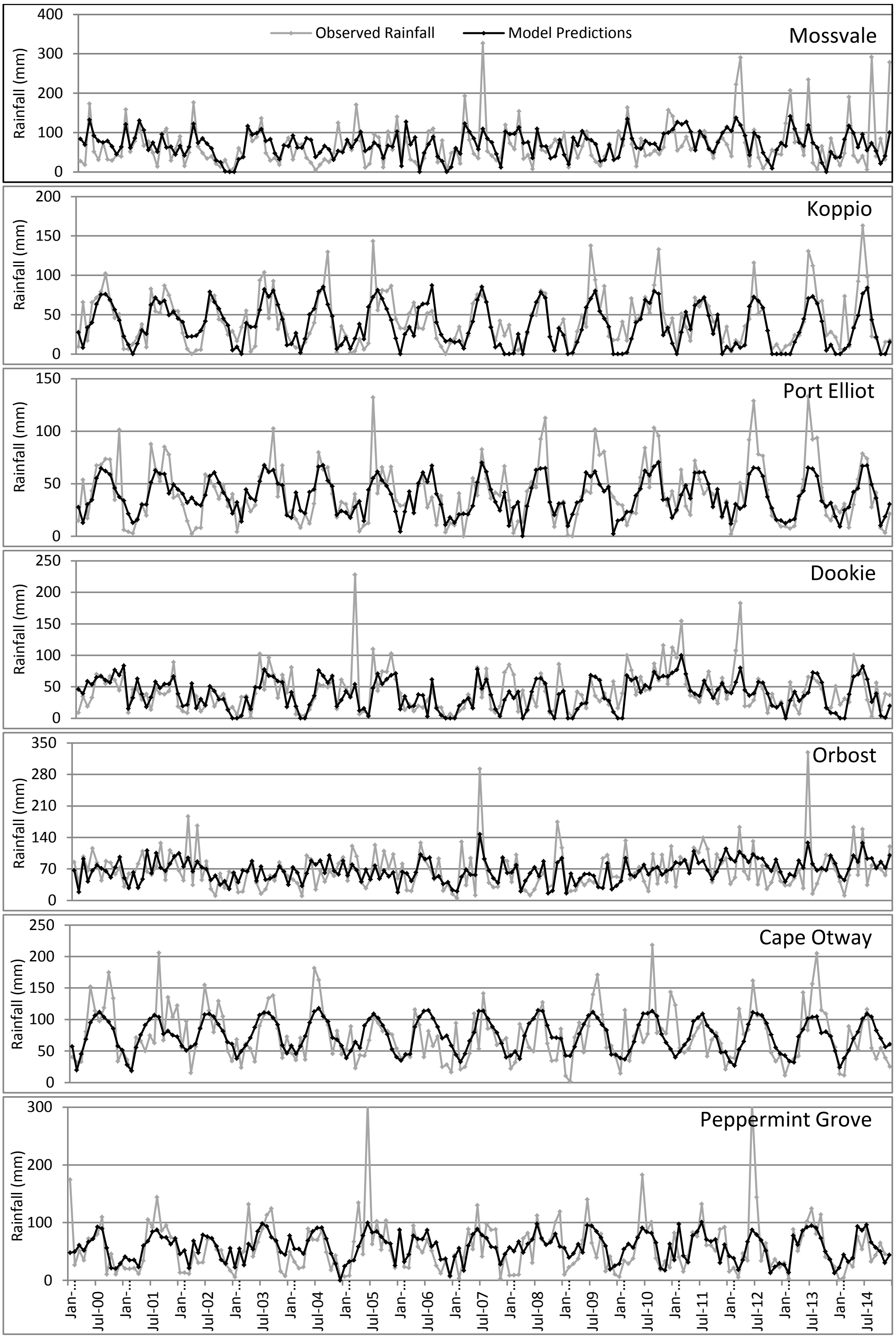

Figure 5.3: Observed rainfall vs. CR(EM) model predictions in temperate zone.

TMin, VP and Rad; and at Koppio and Peppermint Grove with full set of input variables. The model provides worst predictions with the input variable VP in all locations of temperate zone except Mossvale. At Mossvale, the model provides worst predictions with the input variable Evap.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	46.48	31.03	0.57	0.32
Koppio	TMax, TMin, Evap, VP, Rad	23.01	17.08	0.68	0.51
Port Elliot	TMax, TMin, Evap, VP, Rad	19.82	14.99	0.69	0.51
Dookie	TMax, TMin, Evap, VP, Rad	26.57	18.37	0.63	0.35
Orbost	TMax, TMin, Evap, VP, Rad	37.57	26.97	0.59	0.30
Cape Otway	TMax, TMin, Evap, VP, Rad	33.58	25.33	0.70	0.35
Peppermint Grove	TMax, TMin, Evap, VP, Rad	35.80	23.52	0.63	0.38
Performance measures for worst input combinations					
Mossvale	Evap	57.01	37.83	0.70	-0.02
Koppio	VP	31.52	23.64	0.95	0.08
Port Elliot	VP	24.86	18.60	0.85	0.22
Dookie	VP	33.61	23.96	0.83	-0.04
Orbost	VP	45.70	32.13	0.70	-0.03
Cape Otway	VP	36.18	27.99	0.78	0.25
Peppermint Grove	VP	43.16	28.92	0.77	0.10

Table 5.4: The SVM(Linear) model prediction performance with best and worst combination of input variables in temperate zone.

The performance measure RMSE for the SVM(Linear) model ranges from 19.82 to 46.48, MAE from 14.99 to 33.37, MASE from 0.60 to 0.71 and CE from 0.33 to 0.51. The performance measures RMSE, MAE and CE indicate that the model provides best predictions at Port Elliot and MASE indicates at Mossvale while the performance measures RMSE and MAE indicates worst predictions at Mossvale; MASE indicates at Cape Otway and CE indicates at Orbost (Figure 5.1). The graphical illustration of model predictions and actual observations given in Figure 5.4, shows that model follows the series patterns at all locations.

Table 5.5 summarizes the SVM(RBF) model performance with best and worst combinations of input variables for locations in temperate classification zone. Results presented in this table show that the model predicted rainfall with higher accuracy at Peppermint Grove with the combination of input variables TMax, TMin and VP; at Koppio and Dookie with TMax, TMin and VP; at Port Elliot and Cape Otway with TMax, TMin and Evap; at Mossvale with TMax, TMin, Evap and VP; and at Orbost with full set of input variables. The model provides worst predictions at four out of seven locations (Koppio, Port Elliot, Orbost and Peppermint Grove) with the input variable VP; at Cape Otway with Rad; at Mossvale with Evap; and at the location Dookie with the predictors Evap and Rad.

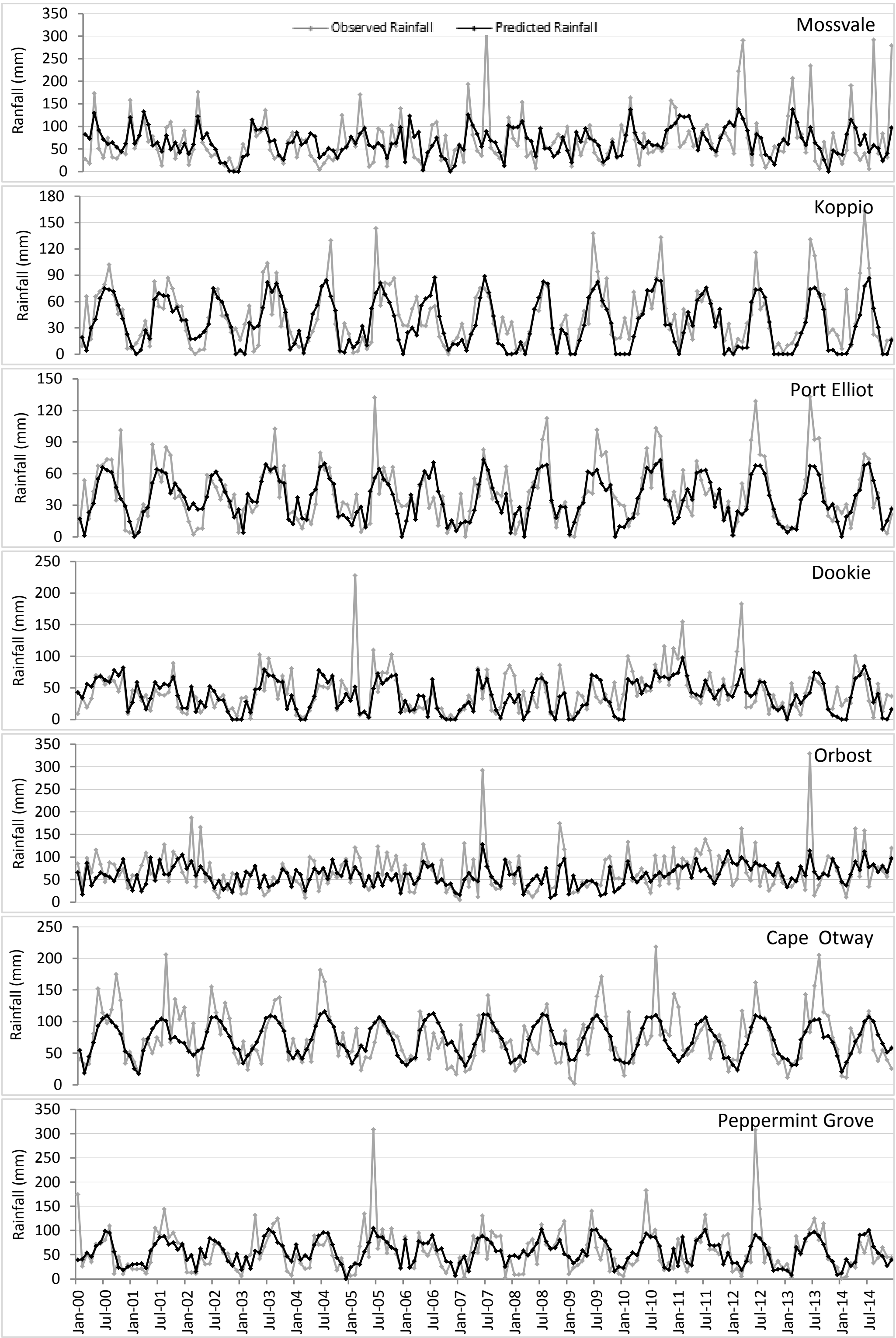

Figure 5.4: Observed rainfall vs. SVM(Linear) model predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP	45.68	29.92	0.55	0.34
Koppio	TMax, TMin, VP, Rad	22.89	16.93	0.68	0.51
Port Elliot	TMax, TMin, Evap, Rad	18.71	14.25	0.65	0.56
Dookie	TMax, TMin, VP, Rad	26.48	18.13	0.63	0.35
Orbost	TMax, TMin, Evap, VP, Rad	36.71	26.83	0.59	0.33
Cape Otway	TMax, TMin, Evap, Rad	31.27	23.81	0.66	0.44
Peppermint Grove	TMax, TMin, VP	35.22	22.78	0.61	0.40
Performance measures for worst input combinations					
Mossvale	Evap	56.62	37.19	0.69	-0.01
Koppio	VP	31.75	23.37	0.94	0.07
Port Elliot	VP	24.30	18.90	0.87	0.26
Dookie	Evap, Rad	33.14	23.83	0.82	-0.01
Orbost	VP	45.66	32.12	0.70	-0.03
Cape Otway	Rad	36.11	27.55	0.77	0.25
Peppermint Grove	VP	43.11	28.85	0.77	0.11

Table 5.5: The SVM(RBF) model prediction performance with best and worst combination of input variables in temperate zone.

In the temperate zone, the performance measure RMSE for the SVM(RBF) model ranges from 18.71 to 45.68, MAE from 14.25 to 29.92, MASE from 0.55 to 0.68 and CE from 0.33 to 0.56. The performance measures RMSE, MAE and CE indicates that the model provides best predictions at Port Elliot and MASE indicates at Mossvale while the performance measures RMSE, and MAE indicate worst predictions at Mossvale and CE at Orbost (Figure 5.1). The graphical display of observed rainfall and model predictions given in Figure 5.5 show that predictions follow the series patterns.

Table 5.6 summarizes the prediction performance of the MLR model with best and worst combinations of input variables for locations in temperate classification zone. These results show that the model provides best predictions at Port Elliot with input variables TMax, TMin and Evap; at Koppio with TMax, TMin, Evap and VP; at Dookie and Cape Otway with TMax, TMin, Evap and Rad; and at Mossvale, Orbost and Peppermint Grove with full set of input variables. The model provides worst predictions with the input variable VP in five out of seven locations (Koppio, Port Elliot, Orbost, CapeOtway and Peppermint Grove); with input variable Rad at Dookie; and with Evap at Mossvale.

The performance measure RMSE for the MLR model ranges from 19.60 to 46.51; MAE from 14.96 to 34.72; MASE from 0.61 to 0.71; and CE from 0.32 to 0.52. The performance measures RMSE, MAE and CE indicates that the MLR model provides best predictions at Port Elliot and worst predictions at Mossvale while MASE indicates best at Dookie and worst at Cape Otway (Figure 5.1). The graphical display of observed rainfall and the MLR Linear model predictions is given in Figure 5.6.

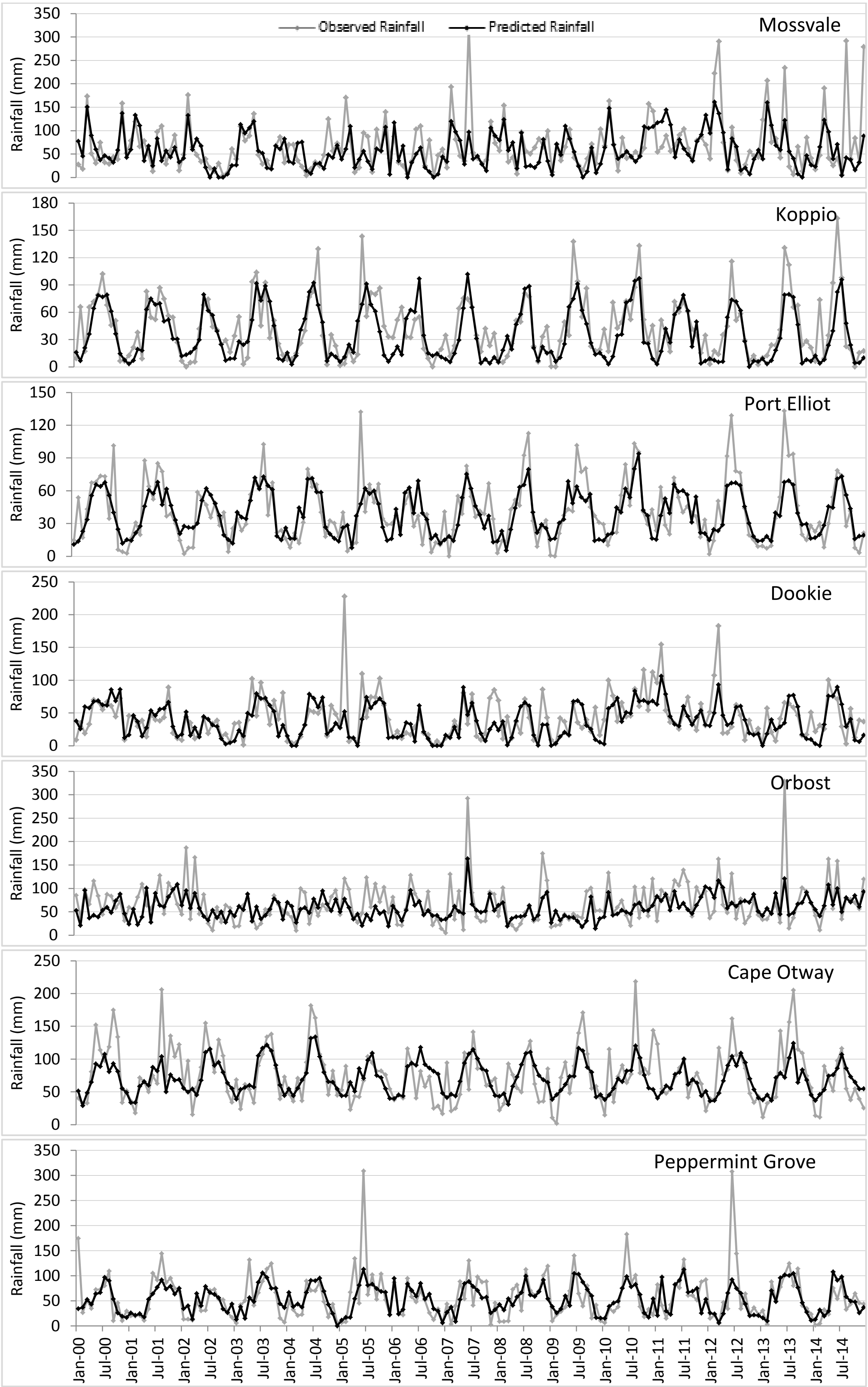

Figure 5.5: Observed rainfall vs. SVM(RBF) model predictions in temperate zone.

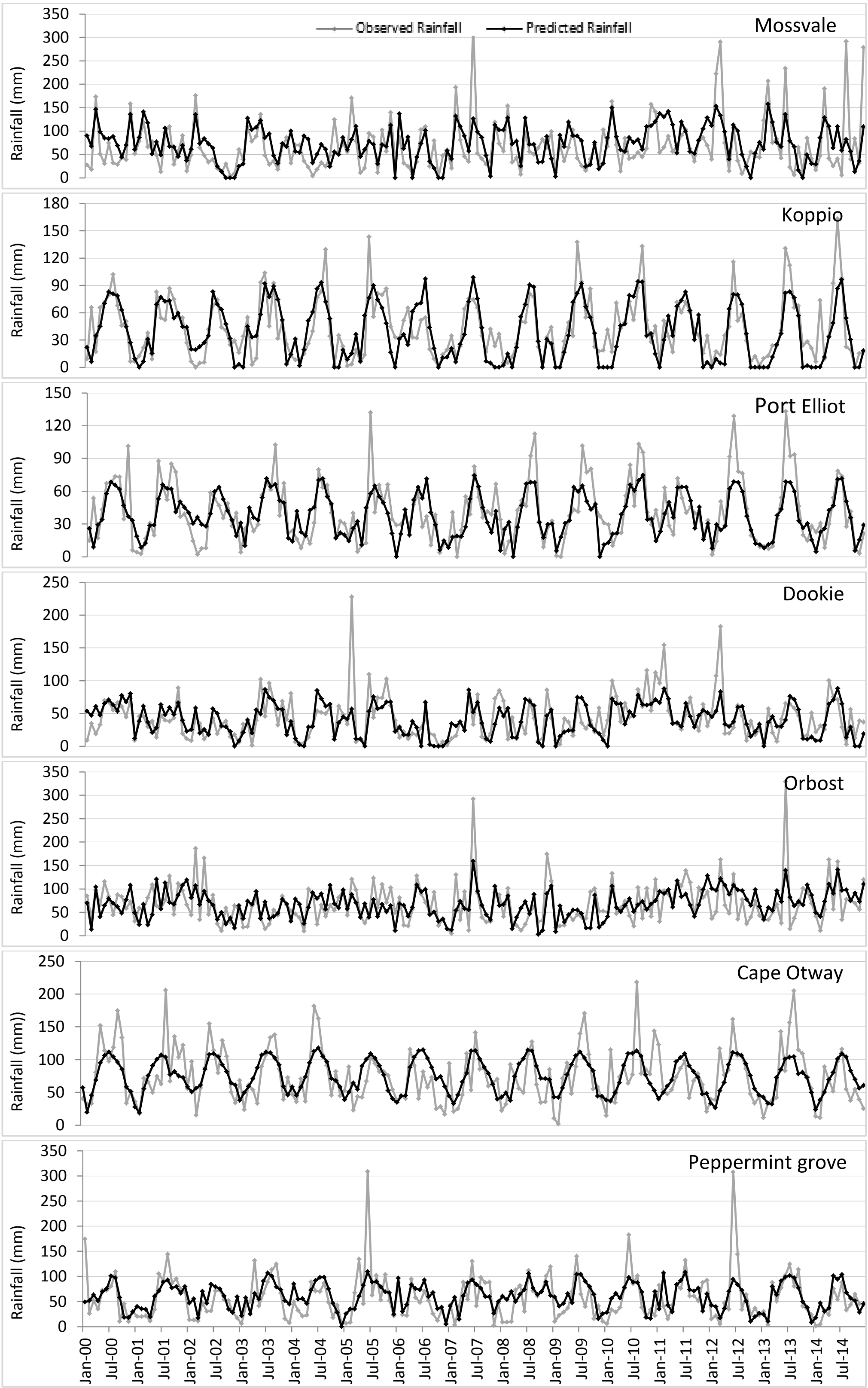

Figure 5.6: Observed rainfall vs. MLR model predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	46.51	34.72	0.64	0.32
Koppio	TMax, TMin, Evap, VP	22.75	17.37	0.69	0.52
Port Elliot	TMax, TMin, Evap	19.60	14.96	0.68	0.52
Dookie	TMax, TMin, Evap, Rad	25.91	18.08	0.62	0.38
Orbost	TMax, TMin, Evap, VP, Rad	36.68	28.03	0.61	0.33
Cape Otway	TMax, TMin, Evap, Rad	33.41	25.59	0.71	0.36
Peppermint Grove	TMax, TMin, Evap, VP, Rad	36.41	25.20	0.68	0.36
Performance measures for worst input combinations					
Mossvale	Evap	58.57	44.22	0.82	-0.08
Koppio	VP	30.89	24.32	0.97	0.12
Port Elliot	VP	24.70	19.24	0.88	0.23
Dookie	Rad	33.52	25.35	0.87	-0.03
Orbost	VP	45.50	34.40	0.75	-0.03
Cape Otway	VP	36.46	28.58	0.79	0.23
Peppermint Grove	VP	42.70	30.37	0.81	0.12

Table 5.6: The MLR model prediction performance with best and worst combination of input variables in temperate zone.

Table 5.7 summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables for locations in temperate classification zone. Results presented in this table show that the model predictions are best at Port Elliot and Peppermint Grove with input variables TMax, TMin and Evap; at Koppio and Orbost with TMax, TMin, Evap and VP; at Dookie and Orbost with TMax, TMin, Evap, Rad; and at Mossvale with full set of five input variables. The model produced worst predictions in all locations of the temperate zone with the input variable VP except Dookie and Mossvale. At Dookie, the model provides the worst predictions with predictors Evap and Rad; and at Mossvale with Evap and VP.

The performance measure RMSE for the ANN(0) model ranges from 19.60 to 46.56, MAE from 14.96 to 32.82, MASE from 0.61 to 0.69 and CE from 0.32 to 0.52. The performance measures RMSE, MAE and CE indicates that the model provides best predictions in Port Elliot and worst predictions in Mossvale while MASE indicates best predictions at Mossvale and worst at Cape Otway (see Figure 5.1). The time series plot for the observed rainfall and the model predictions are given in Figure 5.7.

Table 5.8, summarizes the ANN(1) model performance in predicting rainfall with best and worst combinations of input variables. Results presented in this table show that the model predicted rainfall with the lowest error with input variables TMax and TMin in three locations (Koppio, Cape Otway and Peppermint Grove); with Rad at Orbost; with TMax, TMin, Evap and VP at Port Elliot; with TMax, TMin, Evap and Rad at Dookie; and with full set of variables at Mossvale . The model provides worst predictions with the input variable Evap in three out of seven locations (Koppio, Port

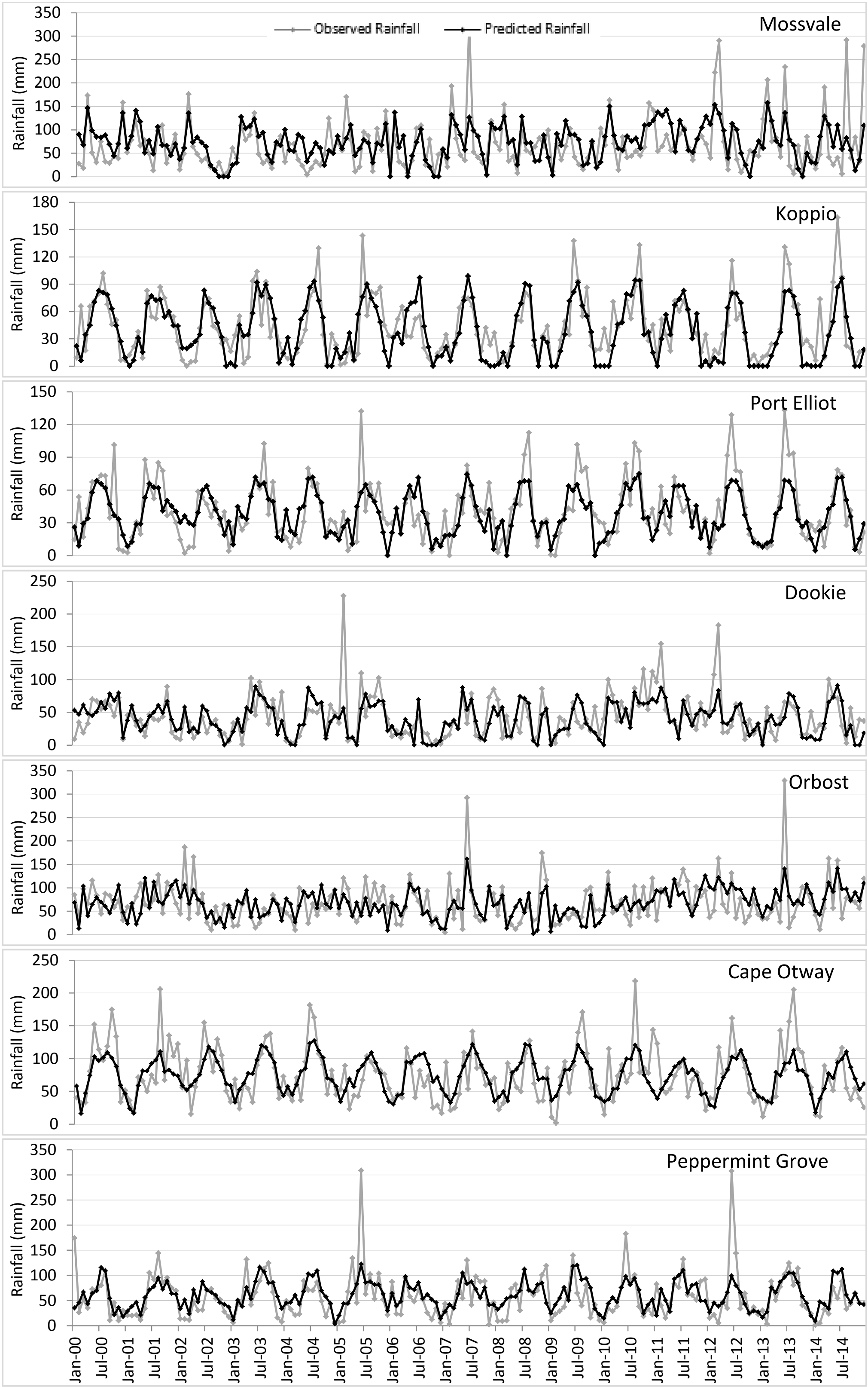

Figure 5.7: Observed rainfall vs. ANN(0) model predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	46.56	32.82	0.61	0.32
Koppio	TMax, TMin, Evap, VP	22.75	17.37	0.69	0.52
Port Elliot	TMax, TMin, Evap	19.60	14.96	0.68	0.52
Dookie	TMax, TMin, Evap, Rad	26.30	18.64	0.64	0.36
Orbost	TMax, TMin, Evap, VP	36.70	28.05	0.62	0.33
Cape Otway	TMax, TMin, Evap, Rad	32.10	24.80	0.69	0.41
Peppermint Grove	TMax, TMin, Evap	35.55	24.25	0.65	0.39
Performance measures for worst input combinations					
Mossvale	Evap, VP	60.00	44.65	0.82	-0.13
Koppio	VP	30.89	24.32	0.97	0.12
Port Elliot	VP	24.70	19.24	0.88	0.23
Dookie	Evap, Rad	33.35	25.24	0.87	-0.02
Orbost	VP	45.50	34.40	0.75	-0.03
Cape Otway	VP	36.46	28.58	0.79	0.23
Peppermint Grove	VP	42.70	30.37	0.81	0.12

Table 5.7: The ANN(0) model prediction performance with best and worst combinations of input variables in temperate zone.

Elliot and Peppermint Grove); with VP at Cape Otway; with Rad at Mossvale; and with Evap and Rad at Dookie and Orbost.

The performance measure RMSE for the ANN(1) model ranges from 19.62 to 44.69, MAE from 15.01 to 32.59, MASE from 0.58 to 0.72 and CE from 0.15 to 0.54. The performance measures RMSE and MAE indicates that the model provides best predictions at Port Elliot; CE indicates at Mossvale; and CE at Koppio while RMSE indicates worst predictions at Mossvale and MAE, MASE and CE indicates at Orbost (see Figure 5.1). The time series plot of the observed rainfall and the ANN(1) model predictions is given in Figure 5.8 shows that model predictions follow the series pattern well except in the location Orbost. At Orbost, the model just provides the average rainfall value.

Table 5.9 summarizes the prediction performance of the k-NN model with the best and worst combinations of input variables for locations in the temperate zone. The k-NN model provides the best predictions with input variables TMax and TMin at Dookie; with TMax, TMin and VP at Orbost; with TMax, TMin, Evap and Rad at Cape Otway; and with TMax, TMin, VP and Rad in the remaining four locations (Mossvale, Koppio, Port Elliot and Peppermint Grove). The model provides worst predictions with the input variable VP in four out of seven locations (Koppio, Port Elliot, Orbost and Peppermint Grove) and with Rad in three locations (Mossvale, Dookie and Cape Otway).

In the temperate zone, the performance measure RMSE for the k-NN model ranges from 19.08 to 47.48; MAE from 14.45 to 33.47; MASE from 0.62 to 0.69 and CE from

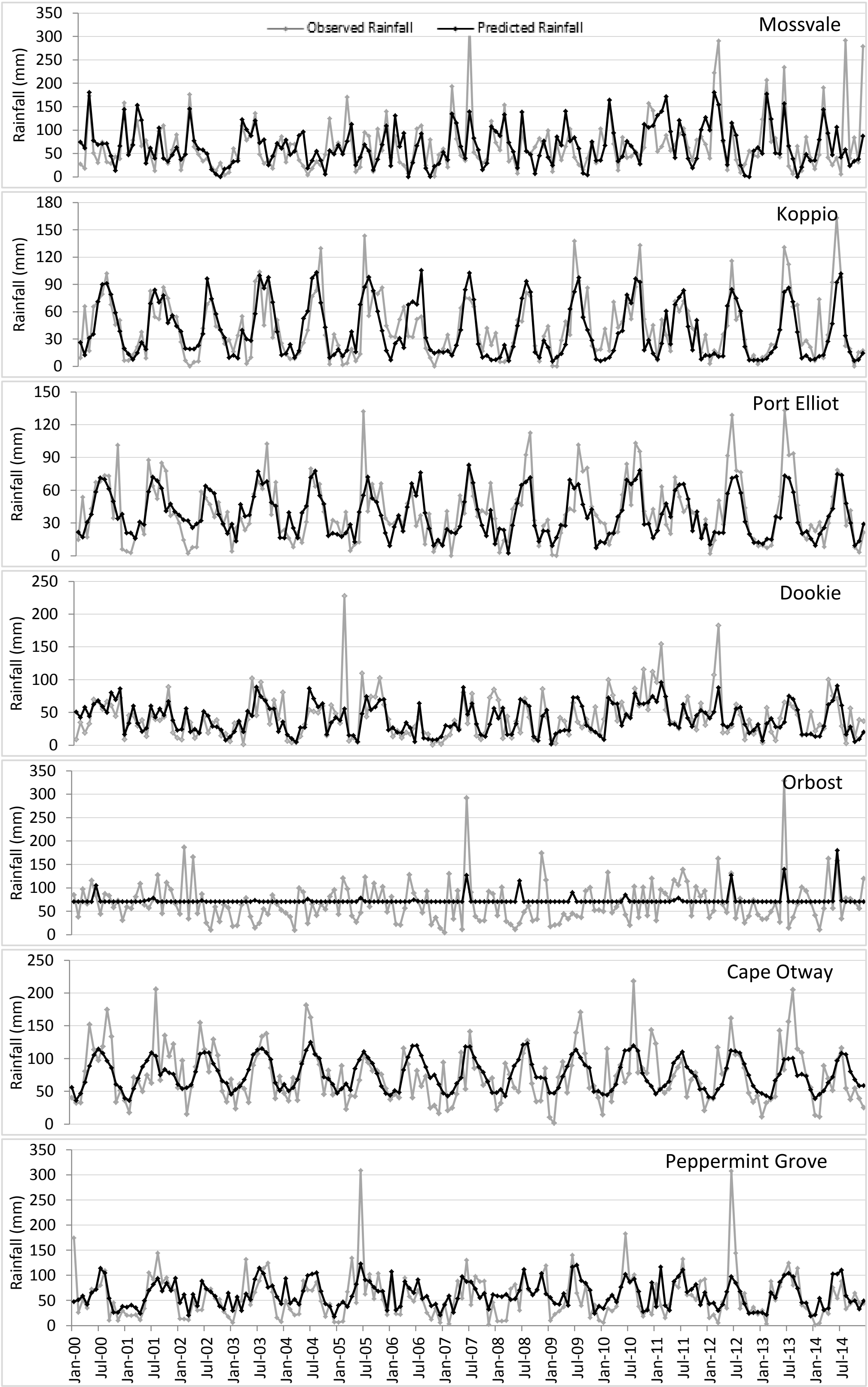

Figure 5.8: Observed rainfall vs. ANN(1) model predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	44.69	31.52	0.58	0.37
Koppio	TMax, TMin	22.24	16.61	0.66	0.54
Port Elliot	TMax, TMin, Evap, VP	19.62	15.01	0.69	0.52
Dookie	TMax, TMin, Evap, Rad	25.34	17.36	0.60	0.41
Orbost	Rad	41.54	32.59	0.72	0.15
Cape Otway	TMax, TMin	33.32	25.72	0.72	0.36
Peppermint Grove	TMax, TMin	36.62	25.62	0.69	0.35
Performance measures for worst input combinations					
Mossvale	Rad	58.41	43.79	0.81	-0.07
Koppio	Evap	32.87	26.37	1.06	0.00
Port Elliot	Evap	28.23	22.19	1.02	0.00
Dookie	TMax, TMin, Evap	33.63	25.98	0.90	-0.04
Orbost	Evap, Rad	46.89	34.90	0.77	-0.09
Cape Otway	Evap, Rad	41.87	33.43	0.93	-0.01
Peppermint Grove	Evap, Rad	45.67	33.90	0.91	0.00

Table 5.8: The ANN(1) model prediction performance with best and worst combinations of input variables in temperate zone.

0.29 to 0.54. The performance measures RMSE, MAE and CE indicate that the k-NN model provides best predictions at Port Elliot and MASE at Mossvale while the performance measures RMSE, and MAE indicate worst predictions at Mossvale; MASE at Cape Otway; and CE at Dookie (see Figure 5.1). The graphical display of observed rainfall and the k-NN model predictions is given in Figure 5.9.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, VP, Rad	47.48	33.47	0.62	0.29
Koppio	TMax, TMin, VP, Rad	22.39	17.05	0.68	0.54
Port Elliot	TMax, TMin, VP, Rad	19.08	14.45	0.66	0.54
Dookie	TMax, TMin	28.32	19.63	0.68	0.26
Orbost	TMax, TMin, VP	37.53	29.56	0.65	0.30
Cape Otway	TMax, TMin, Evap, Rad	32.84	24.95	0.69	0.38
Peppermint Grove	TMax, TMin, VP, Rad	37.07	24.19	0.65	0.34
Performance measures for worst input combinations					
Mossvale	Rad	64.10	46.42	0.86	-0.29
Koppio	VP	32.44	25.71	1.03	0.03
Port Elliot	VP	25.23	19.26	0.88	0.20
Dookie	Rad	42.77	30.75	1.06	-0.68
Orbost	VP	53.59	39.97	0.88	-0.42
Cape Otway	Rad	41.89	32.44	0.90	-0.01
Peppermint Grove	VP	43.81	29.59	0.79	0.08

Table 5.9: The k-NN model prediction performance with best and worst combinations of input variables in temperate zone.

Table 5.10, summarizes the performance of all eight models in predicting monthly

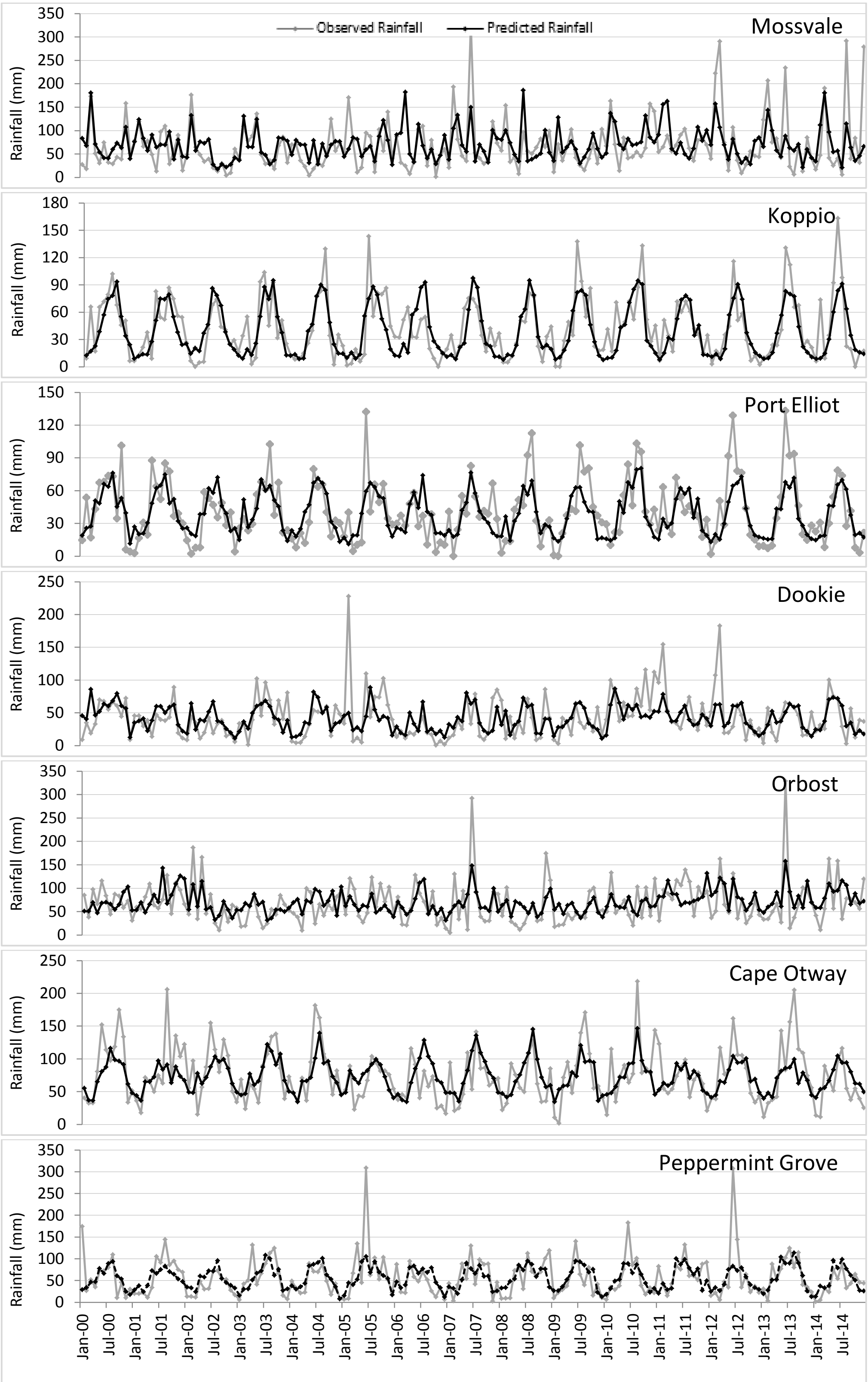

Figure 5.9: Observed rainfall vs. the K-NN model predictions in temperate zone.

rainfall with best combinations of input variables in temperate classification zone. Best results among all models are highlighted in bold. Results presented in this table, show that at least one performance measure indicate that CLR(Opt) is best in two locations compare to all other models; the SVM(RBF) in four locations; ANN(1) in two locations and k-NN model in one location. The CR(EM), SVM(Linear), ANN(0) and MLR models are not best in any location.

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	ANN (0)	ANN (1)	MLR	KNN
Mossvale	RMSE	45.00	46.14	46.48	45.68	46.56	**44.69**	46.51	47.48
	MAE	32.38	33.37	31.03	**29.92**	32.82	31.52	34.72	33.47
	MASE	0.60	0.62	0.57	**0.55**	0.61	0.58	0.64	0.62
	CE	0.36	0.33	0.32	0.34	0.32	**0.37**	0.32	0.29
Koppio	RMSE	22.40	22.99	23.01	22.89	22.75	**22.24**	22.75	22.39
	MAE	16.86	16.95	17.08	16.93	17.37	**16.61**	17.37	17.05
	MASE	0.67	0.68	0.68	0.68	0.69	**0.66**	0.69	0.68
	CE	**0.54**	0.51	0.51	0.51	0.52	**0.54**	0.52	**0.54**
Port Elliot	RMSE	19.37	19.86	19.82	**18.71**	19.60	19.62	19.60	19.08
	MAE	14.66	15.05	14.99	**14.25**	14.96	15.01	14.96	14.45
	MASE	0.67	0.69	0.69	**0.65**	0.68	0.69	0.68	0.66
	CE	0.53	0.51	0.51	**0.56**	0.52	0.52	0.52	0.54
Dookie	RMSE	25.91	25.99	26.57	26.48	26.30	**25.34**	25.91	28.32
	MAE	18.08	17.55	18.37	18.13	18.64	**17.36**	18.08	19.63
	MASE	0.62	0.61	0.63	0.63	0.64	**0.60**	0.62	0.68
	CE	0.38	0.38	0.35	0.35	0.36	**0.41**	0.38	0.26
Orbost	RMSE	**35.27**	36.65	37.57	36.71	36.70	41.54	36.68	37.53
	MAE	**26.00**	27.49	26.97	26.83	28.05	32.59	28.03	29.56
	MASE	**0.57**	0.60	0.59	0.59	0.62	0.72	0.61	0.65
	CE	**0.38**	0.33	0.30	0.33	0.33	0.15	0.33	0.30
Cape Otway	RMSE	33.41	33.40	33.58	**31.27**	32.10	33.32	33.41	32.84
	MAE	25.59	25.58	25.33	**23.81**	24.80	25.72	25.59	24.95
	MASE	0.71	0.71	0.70	**0.66**	0.69	0.72	0.71	0.69
	CE	0.36	0.36	0.35	**0.44**	0.41	0.36	0.36	0.38
Peppermint Grove	RMSE	35.56	36.49	35.80	**35.22**	35.55	36.62	36.41	37.07
	MAE	23.90	24.59	23.52	**22.78**	24.25	25.62	25.20	24.19
	MASE	0.64	0.66	0.63	**0.61**	0.65	0.69	0.68	0.65
	CE	0.39	0.36	0.38	**0.40**	0.39	0.35	0.36	0.34

Table 5.10: The performance of all eight models for monthly rainfall predictions in temperate classification zone.

Results presented in Table 5.10 clearly demonstrate that the CLR(Opt) model is superior to the CR(EM), SVM(Linear), ANN(0) and MLR models for monthly rainfall predictions in temperate classification zone. In comparison with the ANN(1) model, the CLR(Opt) model is best in three locations (Port Elliot, Orbost, Peppermint Grove), ANN(1) model in two locations (Moss Vale and Dookie) and both models performed equally in two locations (Koppio, Cape Otway). According to the RMSE measure, the ANN(1) model performance in Moss Vale is only 0.69% higher and in Dookie 2.2% than CLR(Opt) model.

Similarly, in comparison with the SVM(RBF) model, the CLR(Opt) model is best in three out of seven locations (Moss Vale, Dookie and Orbost); SVM(RBF) is best in three locations (Port Elliot, Cape Otway and Peppermint Grove) and both models have similar results in one location (Koppio). According to the RMSE measure, the SVM(RBF) model performance in Port Elliot is 3.41%, in Cape Otway 6.41% and in Peppermint Grove 0.96% higher than CLR(Opt) model.

Although the ANN(1) model outperformed in two locations, but the overall performance is unreliable. In most locations, the ANN(1) model just predicted a single value for all observations in the test data with most combinations of input variables. For example, in the location Moss Vale, the model predicted 79.45 with five combinations (C2, C5, C8, C10, C12) of input variables. The detail of combinations (c1-c15) is given in Table 5.1. Similarly, in Koppio, the model predicted 43.16 for all 180 observations with ten combinations (C2, C3, C6, C7, C9, C10, C11, C12, 13, 15) and in Dookie, 49.86 with eight combinations (C3, C5, C6, C7, C8, C9, C11, C14).

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8 and 5.9, show that all models follow the series patterns at all locations except Orbost. In Orbost, the ANN(1) model failed to predict rainfall values (see Figure 5.8).

In summary, based on the performance measures, the CLR(Opt) and SVM(RBF) model are the most suitable models in finding the pattern and trends of the observations compared to other models at all sites in temperate classification zone.

5.2 Monthly rainfall predictions in grassland zone

In this section, first we present the monthly rainfall prediction results for each model in predicting monthly rainfall with best and worst combinations of input variables in grassland classification zone. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models using computational results and time series plots.

Table 5.11 summarizes the CLR(Opt) model prediction performance with best and worst combinations of input variables in grassland classification zone. Results presented show that the model provides best predictions at Alexandria with input variables TMax and TMin; at Warren with TMax, TMin and VP; at Newry, Richmond, Annuello, Ningaloo and Dowerine with TMax, TMin, VP and Rad; and at Blinman with a full set of input variables. The model provides the worst predictions with input variable Evap in four locations (Warren, Newry, Alexandria and Annuello); with VP in two locations (Ningaloo and Dowerine); and with Evap and Rad in two locations (Richmond and Blinman).

In grassland zone, the performance measure RMSE for the CLR(Opt) model ranges from 19.43 to 70.15, MAE from 12.80 to 37.91, MASE from 0.48 to 0.68 and CE from 0.24 to 0.68. The performance measures RMSE and MAE indicate the CLR(Opt) model provides best predictions at Dowerine and worst at Newry; the performance measure MASE indicates best at Alexandria and worst at Annuello; and CE indicates best predictions at Richmond and worst at Annuello (Figure 5.10). Graphical display of observed rainfall and the CLR(Opt) model predictions given in Figure 5.11 show that model predictions follow the series patterns and tried well to predict the extreme values at Alexandria and Richmond.

Table 5.3, summarizes the CR(EM) model performance in predicting monthly rainfall with best and worst combinations of input variables in grassland classification zone. These results show that the model provides best predictions with input variables TMax, TMin, Evap and VP at Alexandria and Ningaloo; at Warren, with input variables TMax, TMin, VP and Rad at Newry and Blinman; and with full set of five variables at Richmond, Annuello and Dowerine. The model produced worst predictions in five locations (Newry, Alexandria, Blinman, Annuello and Ningaloo) with input variable Evap; at Richmond with Rad; at Dowerine with VP; and at Warren with Evap and Rad.

The performance measure RMSE for the CR _EM model ranges from 20.25 to 95.77; MAE from 12.98 to 47.50; MASE from 0.56 to 0.69; and CE from 0.23 to 0.51. The performance measures RMSE, MAE and MASE indicates that the CR(EM) model predictions have lowest prediction error at Dowerine and highest at Newry while CE indicates lowest at Annuello and highest at Richmond (see Figure 5.10).

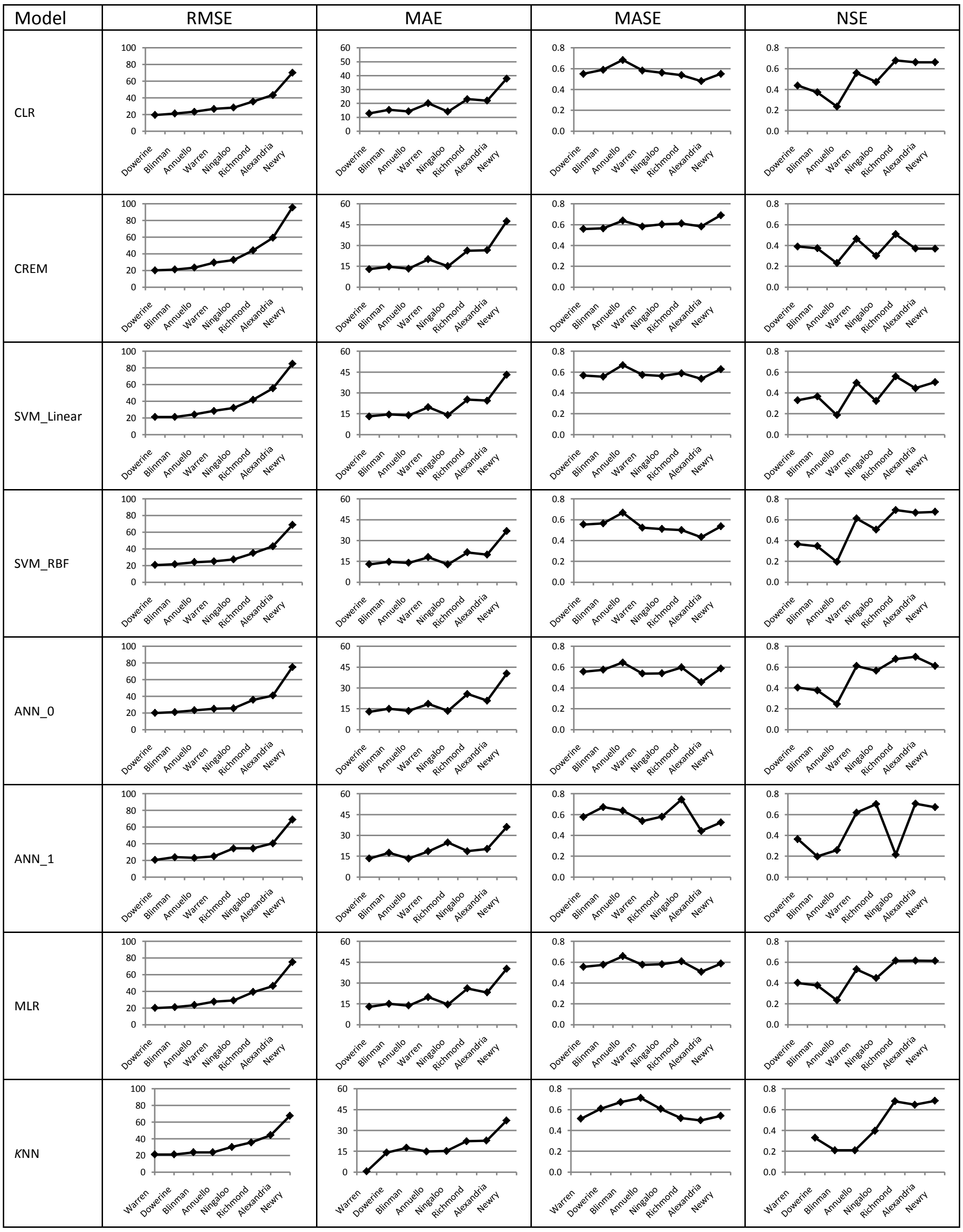

Figure 5.10: The prediction performance of models in grassland zone.

The graphical display of observed rainfall and the model predictions for fifteen years for all eight locations in the grassland classification zone are given in Figure 5.12.

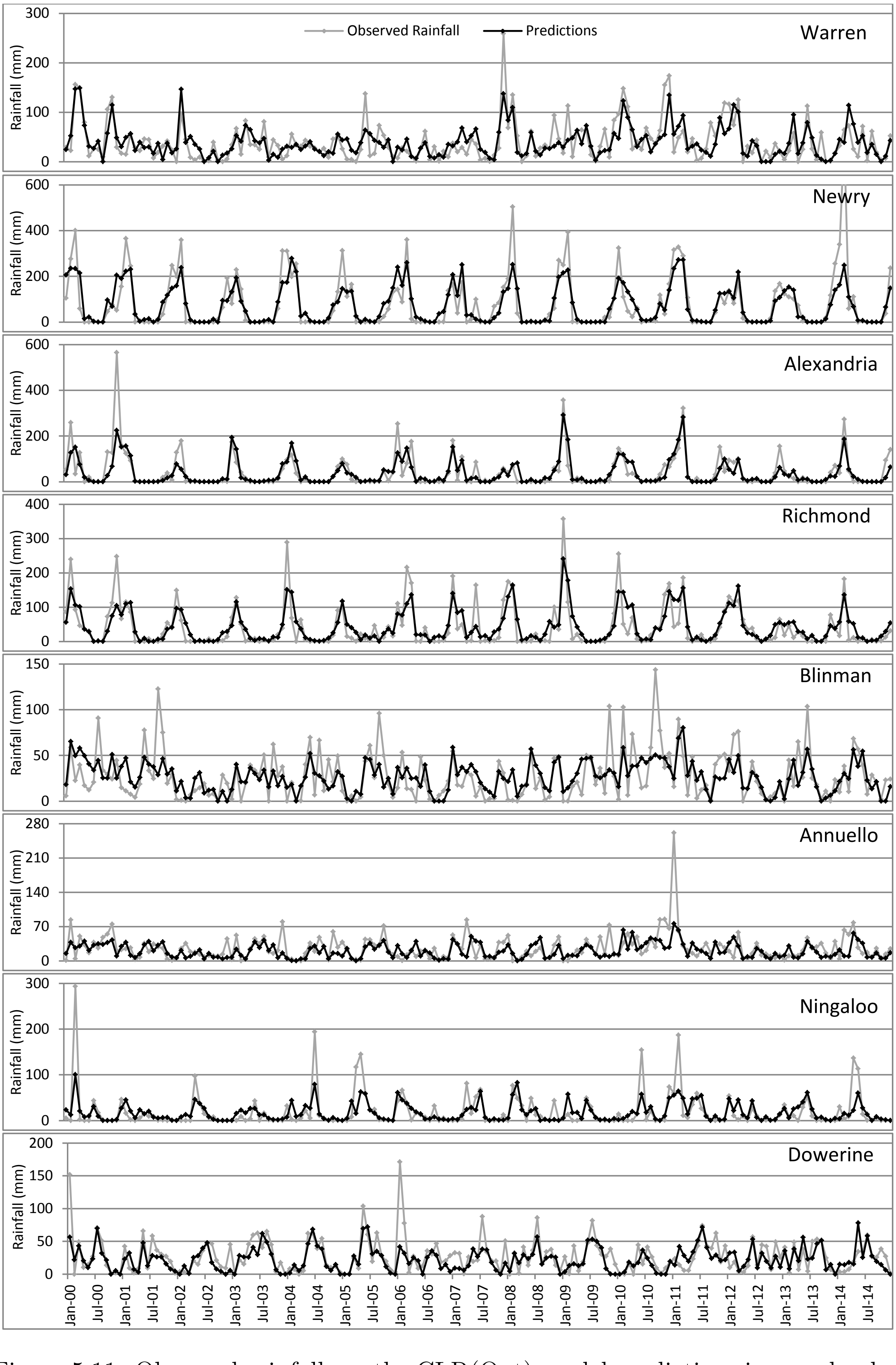

Figure 5.11: Observed rainfall vs. the CLR(Opt) model predictions in grassland zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP	26.80	20.10	0.58	0.56
Newry	TMax, TMin, VP, Rad	70.15	37.91	0.55	0.66
Alexandria	TMax, TMin	43.44	22.02	0.48	0.66
Richmond	TMax, TMin, VP, Rad	35.73	23.10	0.54	0.68
Blinman	TMax, TMin, Evap, VP, Rad	21.14	15.40	0.59	0.37
Annuello	TMax, TMin, VP, Rad	23.42	14.28	0.68	0.24
Ningaloo	TMax, TMin, VP, Rad	28.32	14.06	0.56	0.47
Dowerine	TMax, TMin, VP, Rad	19.43	12.80	0.55	0.44
Performance measures for worst input combinations					
Warren	Evap	39.54	29.38	0.85	0.04
Newry	Evap	118.82	85.24	1.24	0.03
Alexandria	Evap	74.64	47.79	1.04	0.00
Richmond	Evap, Rad	62.52	47.49	1.11	0.02
Blinman	Evap, Rad	26.61	21.21	0.81	0.01
Annuello	Evap	27.28	17.93	0.86	-0.03
Ningaloo	VP	37.36	23.20	0.93	0.08
Dowerine	VP	25.16	19.02	0.82	0.06

Table 5.11: The CLR(Opt) model performance for monthly rainfall predictions with best and worst combination of input variables in grassland zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP, Rad	29.53	20.11	0.58	0.46
Newry	TMax, TMin, VP, Rad	95.77	47.50	0.69	0.37
Alexandria	TMax, TMin, Evap, VP	59.25	26.67	0.58	0.37
Richmond	TMax, TMin, Evap, VP, Rad	44.21	26.25	0.61	0.51
Blinman	TMax, TMin, VP, Rad	21.14	14.75	0.56	0.37
Annuello	TMax, TMin, Evap, VP, Rad	23.51	13.35	0.64	0.23
Ningaloo	TMax, TMin, Evap, VP	32.62	15.12	0.60	0.30
Dowerine	TMax, TMin, Evap, VP, Rad	20.25	12.98	0.56	0.39
Performance measures for worst input combinations					
Warren	Evap, Rad	40.73	30.58	0.89	-0.02
Newry	Evap	119.05	86.41	1.26	0.03
Alexandria	Evap	74.35	49.28	1.08	0.01
Richmond	Rad	63.43	47.70	1.11	-0.01
Blinman	Evap	26.36	20.94	0.80	0.03
Annuello	Evap	27.00	17.88	0.86	-0.01
Ningaloo	Evap	38.08	23.32	0.93	0.05
Dowerine	VP	25.98	19.64	0.85	0.00

Table 5.12: The CR(EM) model prediction performance with best and worst combinations of input variables in grassland zone.

Table 5.13 summarizes the prediction performance of the SVM(Linear) model with best and worst combinations of input variables in grassland zone. These results show that the model provides best predictions at Newry, Alexandria, Ningaloo and

Dowerine with input variables TMax, TMin, Evap and VP; at Warren and Blinman with TMax, TMin, VP and Rad; and at Richmond and Annuello with a full set of five input variables. The model provides worst predictions with input variable Evap at Newry and Annuello; with VP at Ningaloo and Dowerine; with Rad at Alexandria and Richmond; and with Evap and Rad at warren and Blinman.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP, Rad	28.57	19.77	0.57	0.50
Newry	TMax, TMin, Evap, VP	84.98	43.20	0.63	0.50
Alexandria	TMax, TMin, Evap, VP	55.62	24.52	0.54	0.45
Richmond	TMax, TMin, Evap, VP, Rad	41.88	25.30	0.59	0.56
Blinman	TMax, TMin, VP, Rad	21.26	14.52	0.56	0.37
Annuello	TMax, TMin, Evap, VP, Rad	24.16	13.94	0.67	0.19
Ningaloo	TMax, TMin, Evap, VP	32.11	14.09	0.56	0.32
Dowerine	TMax, TMin, Evap, VP	21.23	13.18	0.57	0.33
Performance measures for worst input combinations					
Warren	Evap, Rad	41.53	27.75	0.80	-0.06
Newry	Evap	138.04	78.38	1.14	-0.31
Alexandria	Rad	82.47	42.03	0.92	-0.22
Richmond	Rad	68.57	39.05	0.91	-0.18
Blinman	Evap, Rad	27.01	19.69	0.75	-0.02
Annuello	Evap	28.55	17.44	0.84	-0.13
Ningaloo	VP	40.58	18.91	0.76	-0.08
Dowerine	VP	28.03	19.48	0.84	-0.17

Table 5.13: The SVM(Linear) model prediction performance with best and worst combinations of input variables in grassland zone.

The performance measure RMSE for the SVM(Linear) model ranges from 21.23 to 84.98; MAE from 13.18 to 43.20; MASE from 0.54 to 0.67; and CE from 0.19 to 0.56. The performance measures RMSE and MAE indicates that the SVM(Linear) model provides best predictions at Dowerine and worst at Newry; MASE indicates best at Alexandria and worst at Annuello, and CE indicates best at Richmond and worst at Annuello (Figure 5.10). The graphical illustration of model predictions with the actual observations presented in Figure 5.13, shows that model follows the series patterns at all locations.

Table 5.14 summarizes the prediction performance of the SVM(RBF) model with best and worst combinations of input variables in grassland classification zone. The results presented in this table show that the model provides best predictions with input variables TMax, TMin and VP at Newry and Ningaloo; with TMax, TMin, VP and Rad at Warren, Richmond, Annuello and Dowerine; and with a full set of input variables in the remaining locations Alexandria and Blinman. The model provides worst predictions with input variable Evap at Newry, Alexandria, Richmond, Annuello and Ningaloo; with VP at Blinman and Dowerine; and with Rad at Warren.

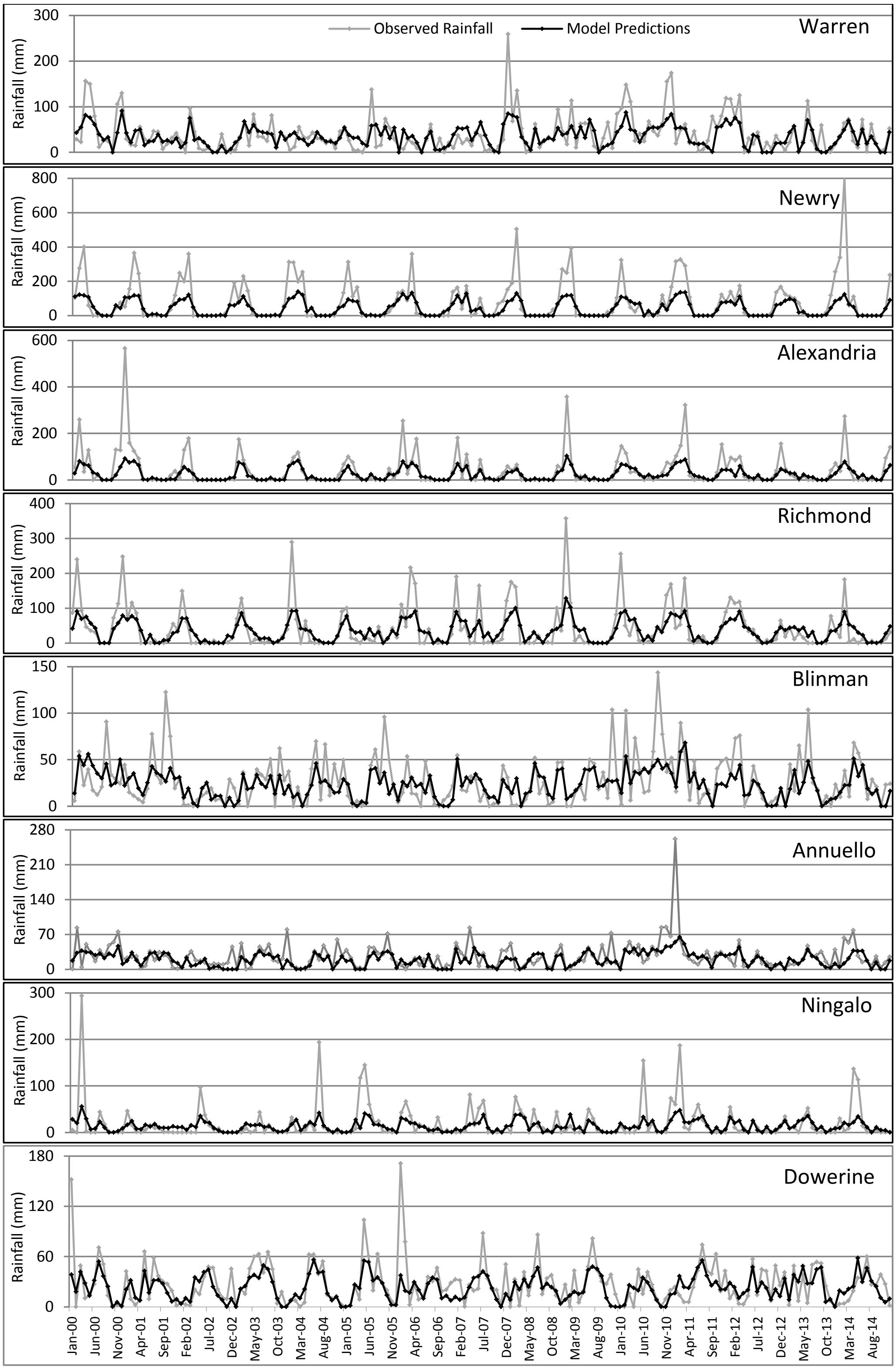

Figure 5.12: Observed rainfall vs. the CR(EM) model predictions in grassland zone.

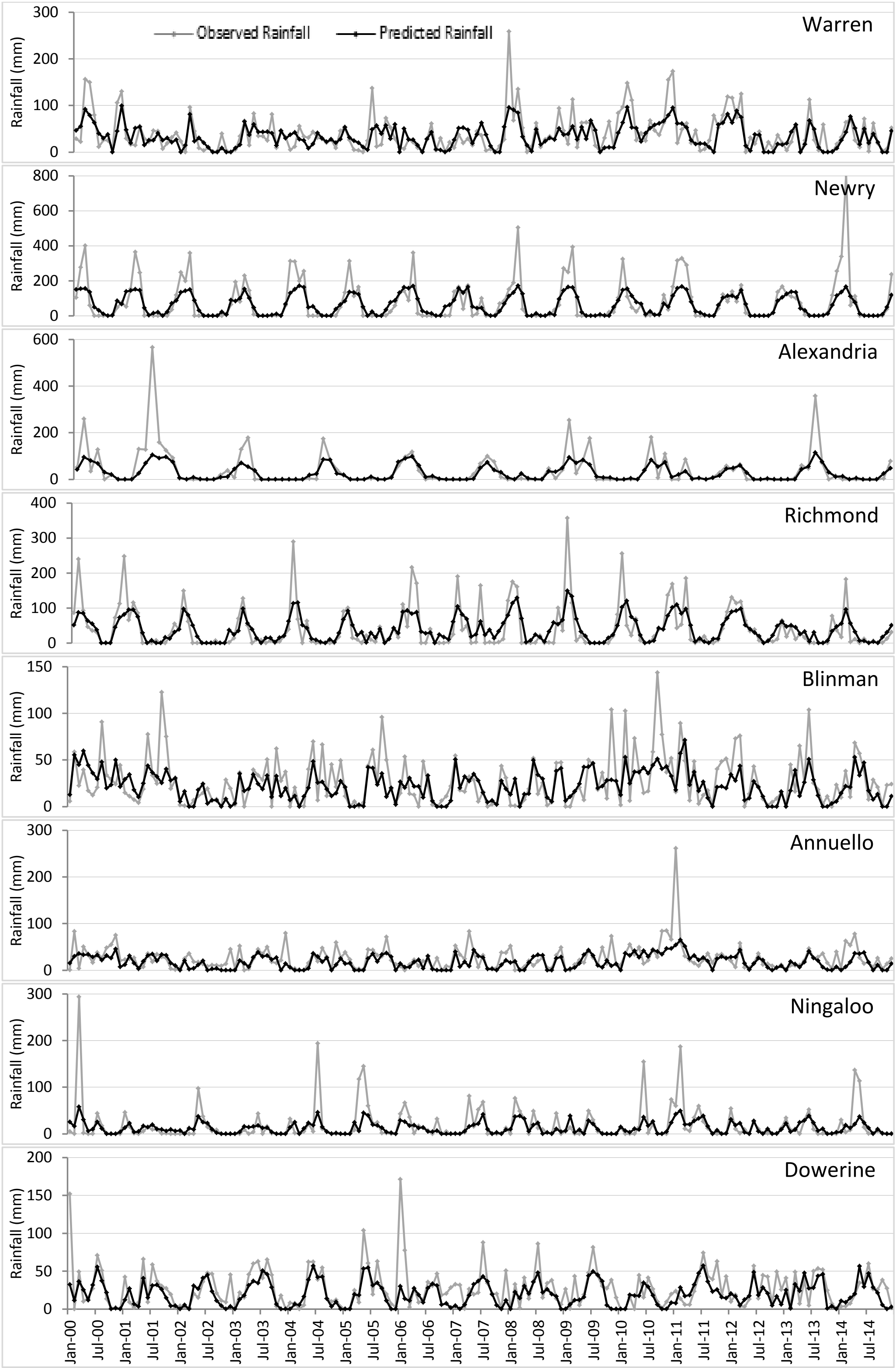

Figure 5.13: Observed rainfall vs. the SVM(Linear) model predictions in grassland zone.

The performance measure RMSE for the SVM(RBF) model ranges from 20.66 to 84.98, MAE from 12.79 to 36.84, MASE from 0.50 to 0.67 and CE from 0.20 to 0.69. The performance measure RMSE indicates that the model has lowest prediction error at Dowerine; MAE indicates at Ningaloo; MASE indicates at Alexandria, and CE indicates at Richmond. The performance measures RMSE and MAE indicates the model provides worst predictions at the location Newry and MASE and CE indicates at the location Annuello (see Figure 5.10). The Graphical display of observed rainfall and the SVM(Linear) model predictions given in Figure 5.14, show that the model predictions follow the series patterns and at location Richmond the model successfully predicted the extreme values.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP, Rad	25.13	18.05	0.52	0.61
Newry	TMax, TMin, VP	68.68	36.84	0.54	0.68
Alexandria	TMax, TMin, Evap, VP, Rad	43.09	19.82	0.43	0.67
Richmond	TMax, TMin, VP, Rad	34.94	21.47	0.50	0.69
Blinman	TMax, TMin, Evap, VP, Rad	21.59	14.73	0.56	0.35
Annuello	TMax, TMin, VP, Rad	24.05	13.93	0.67	0.20
Ningaloo	TMax, TMin, VP	27.43	12.79	0.51	0.51
Dowerine	TMax, TMin, VP, Rad	20.66	12.87	0.55	0.37
Performance measures for worst input combinations					
Warren	Rad	40.39	26.53	0.77	0.00
Newry	Evap	125.86	79.05	1.15	-0.09
Alexandria	Evap	81.71	41.65	0.91	-0.20
Richmond	Evap	66.57	39.90	0.93	-0.11
Blinman	VP	27.80	19.00	0.73	-0.08
Annuello	Evap	28.22	17.21	0.82	-0.11
Ningaloo	Evap	37.69	16.96	0.68	0.07
Dowerine	VP	25.87	18.66	0.80	0.01

Table 5.14: The SVM(RBF) model prediction performance with best and worst combinations of input variables in grassland zone.

Table 5.15 summarizes the prediction performance of the MLR model with best and worst combinations of input variables in grassland zone. The results show that the model provides best predictions with combination of input variables TMax, TMin, Evap and VP at Ningaloo and Dowerine; with TMax, TMin, VP and Rad at Warren and Blinman; with TMax, TMin and VP at Newry; and with full set of input variables at the remaining three locations Alexandria, Richmond and Annuello. The MLR model provides worst predictions with input variable Evap in four locations (Newry, Alexandria, Richmond and Annuello); with VP in two locations (Ningaloo and Dowerine); and with Evap and Rad in the remaining two locations (Warren and Blinman).

The performance measure RMSE for the MLR model ranges from 20.07 to 75.10,

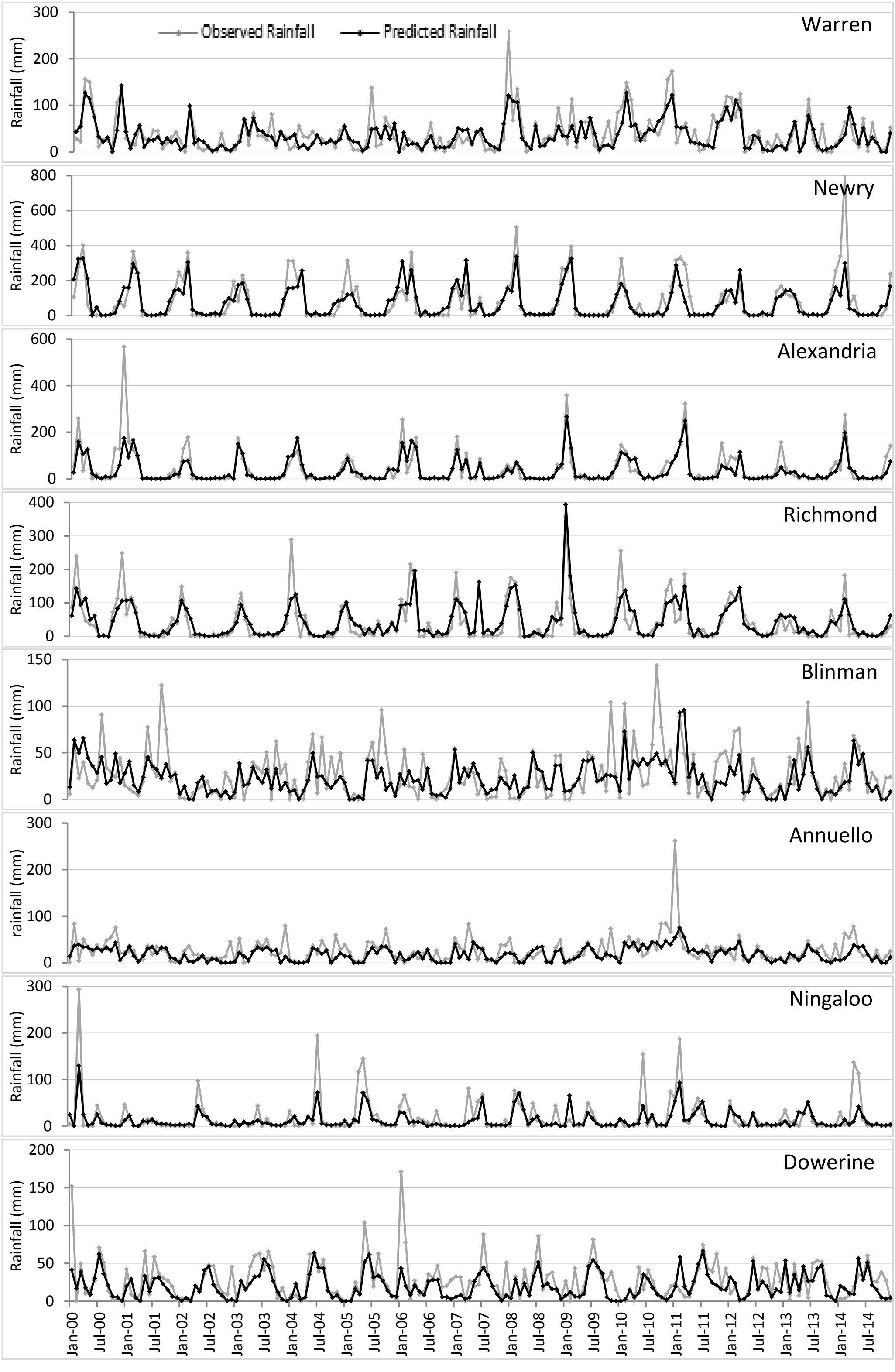

Figure 5.14: Observed rainfall vs. the SVM(RBF) model predictions in grassland zone.

MAE from 12.88 to 40.31, MASE from 0.51 to 0.66 and CE from 0.23 to 0.61. The performance measures RMSE and MAE indicates that the MLR model has best predictions at Dowerine; MASE indicates at Alexandria; CE indicates at Richmond, Alexandria and Newry. Similarly, the performance measures RMSE, and MAE indicates the model provides worst predictions at Newry and MASE and CE indicates at Annuello (see Figure 5.10). The time series plot of observed rainfall and the MLR model predictions given in Figure 5.15, show that the model predictions follow the series patterns but unable to predict extreme values.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP, Rad	27.59	19.85	0.58	0.53
Newry	TMax, TMin, VP	75.10	40.31	0.59	0.61
Alexandria	TMax, TMin, Evap, VP, Rad	46.38	23.21	0.51	0.61
Richmond	TMax, TMin, Evap, VP, Rad	39.21	26.12	0.61	0.61
Blinman	TMax, TMin, VP, Rad	21.08	15.02	0.58	0.38
Annuello	TMax, TMin, Evap, VP, Rad	23.46	13.72	0.66	0.23
Ningaloo	TMax, TMin, Evap, VP	29.02	14.54	0.58	0.45
Dowerine	TMax, TMin, Evap, VP	20.07	12.88	0.55	0.40
Performance measures for worst input combinations					
Warren	Evap, Rad	40.72	30.51	0.88	-0.02
Newry	Evap	119.45	85.92	1.25	0.02
Alexandria	Evap	74.64	47.79	1.04	0.00
Richmond	Evap	62.97	45.62	1.06	0.00
Blinman	Evap, Rad	26.61	21.21	0.81	0.01
Annuello	Evap	27.28	17.92	0.86	-0.03
Ningaloo	VP	37.64	23.62	0.94	0.07
Dowerine	VP	26.31	19.39	0.84	-0.03

Table 5.15: The MLR model prediction performance with best and worst combinations of input variables in grassland zone.

Table 5.16 summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables for location in grassland zone. The results presented show that the model performance was best with input variables TMax and TMin at Richmond; with TMax, TMin, Evap and VP at Dowerine; with TMax, TMin, VP and Rad at Warren, Alexandria, Blinman and Ningaloo; and with a full set of input variables at Newry and Annuello. The ANN(0) model provides the worst predictions at four locations (Newry, Alexandria, Richmond and Annuello) with input variable Evap; at two locations (Ningaloo and Dowerine) with VP; and at the remaining two locations (Warren and Blinman) with Rad.

The performance measure RMSE for the ANN(0) model ranges from 20.02 to 75.09, MAE from 12.91 to 40.39, MASE from 0.46 to 0.64 and CE from 0.25 to 0.70. The performance measures RMSE and MAE indicates that the model has lowest prediction error at the location Dowerine and highest at Newry while performance

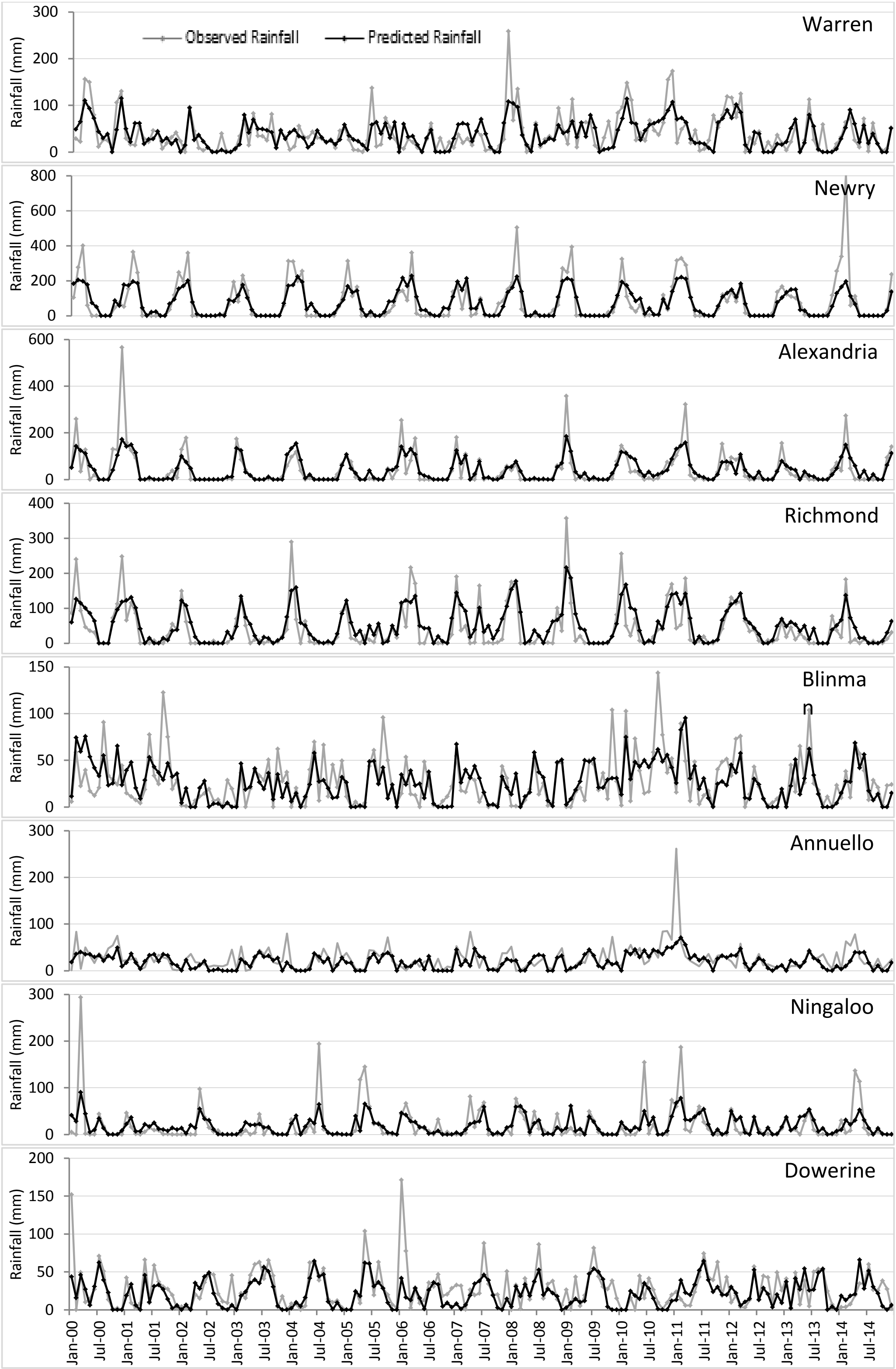

Figure 5.15: Observed rainfall vs. the MLR model predictions in grassland zone.

measures MASE and CE indicates lowest at Alexandria and highest at Annuello (see Figure 5.10). The graphical display of observed rainfall and the model predictions given in Figure 5.16, show that model predictions follow the series patterns.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP, Rad	25.13	18.58	0.54	0.61
Newry	TMax, TMin, Evap, VP, Rad	75.09	40.39	0.59	0.61
Alexandria	TMax, TMin, VP, Rad	40.97	20.84	0.46	0.70
Richmond	TMax, TMin	35.81	25.66	0.60	0.68
Blinman	TMax, TMin, VP, Rad	21.08	15.02	0.58	0.38
Annuello	TMax, TMin, Evap, VP, Rad	23.27	13.46	0.64	0.25
Ningaloo	TMax, TMin, VP, Rad	25.70	13.53	0.54	0.57
Dowerine	TMax, TMin, Evap, VP	20.02	12.91	0.56	0.40
Performance measures for worst input combinations					
Warren	Rad	40.41	30.16	0.87	0.00
Newry	Evap	119.45	85.92	1.25	0.02
Alexandria	Evap	74.64	47.79	1.04	0.00
Richmond	Evap	62.97	45.62	1.06	0.00
Blinman	Rad	27.79	19.60	0.75	-0.08
Annuello	Evap	27.28	17.92	0.86	-0.03
Ningaloo	VP	37.88	23.70	0.95	0.06
Dowerine	VP	26.31	19.39	0.84	-0.03

Table 5.16: The ANN(0) model prediction performance with best and worst combinations of input variables in grassland zone.

Table 5.17 summarizes the prediction performance of the ANN(1) model with best and worst combinations of input variables using test data sets for all eight locations in grassland zone. The results presented in this table show that the model with input variables TMax and TMin provides best predictions at the location Richmond; with TMax, TMin and Rad at two locations (Newry and Alexandria); with Tmax, TMin and VP at two locations (Blinman and Ningaloo); and with input variables TMax, TMin, VP and Rad at three locations (Warren, Annuello and Dowerine).

The performance measure RMSE for the ANN _1 model ranges from 20.67 to 69.18, MAE from 13.32 to 36.10, MASE from 0.44 to 0.75 and CE from 0.20 to 0.71. The performance measures RMSE and MAE indicates that the ANN(1) model have lowest prediction error at the location Dowerine, and MASE and CE indicate best at Alexandria while RMSE and MAE indicate worst predictions at Newry; MASE at Ningaloo, and CE at Blinman and Ningaloo (see Figure 5.10). The graphical display of observed rainfall and the ANN(1) model predictions for all eight locations in the grassland classification zone is given in Figure 5.17. The figure shows that the model predictions follow the series pattern and tried well to predict the extreme values.

Table 5.18 summarizes the prediction performance of the k-NN model with the best and worst combinations of input variables for locations in grassland zone. The

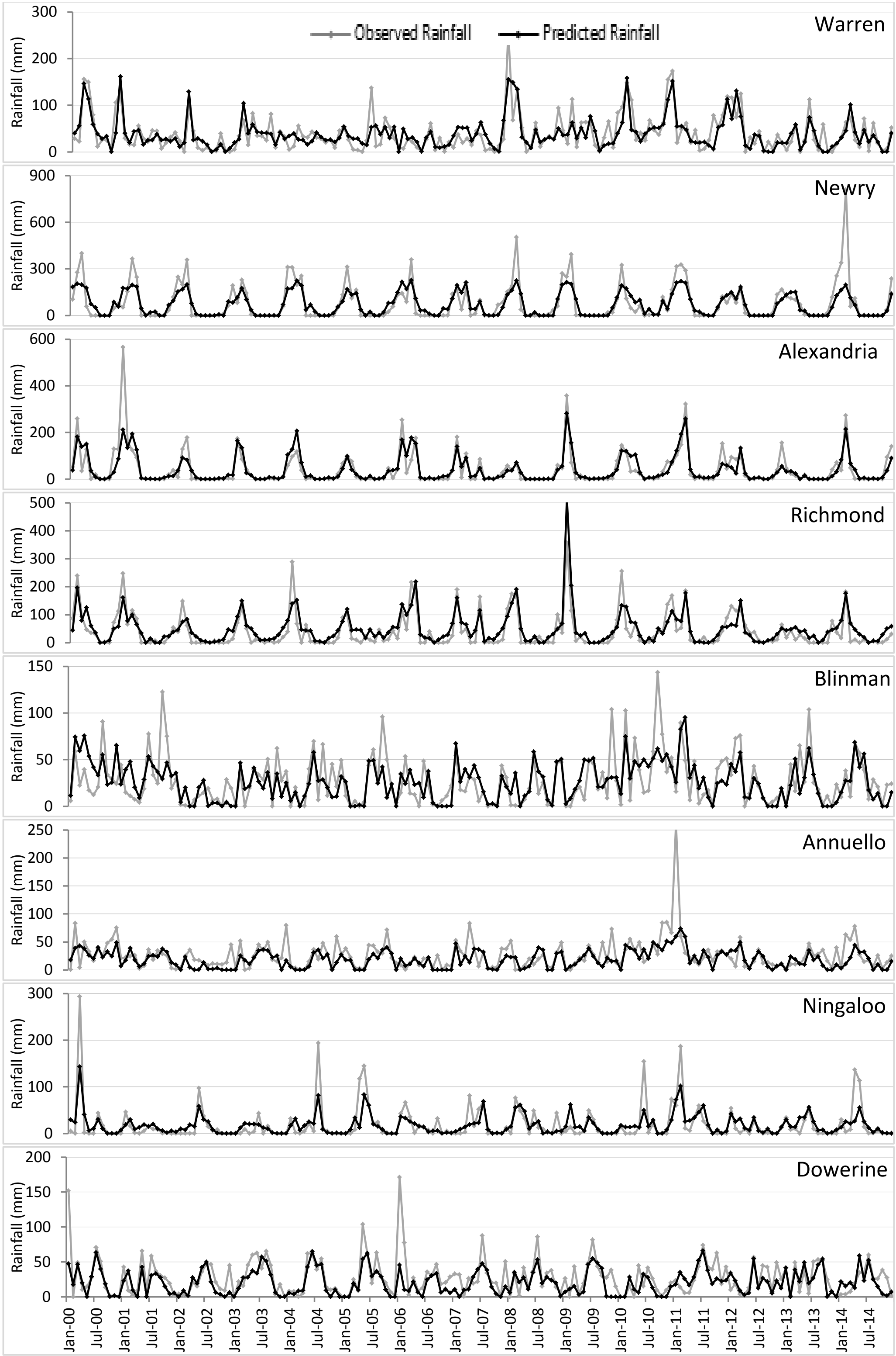

Figure 5.16: Observed rainfall vs. the ANN(0) model predictions in grassland zone.

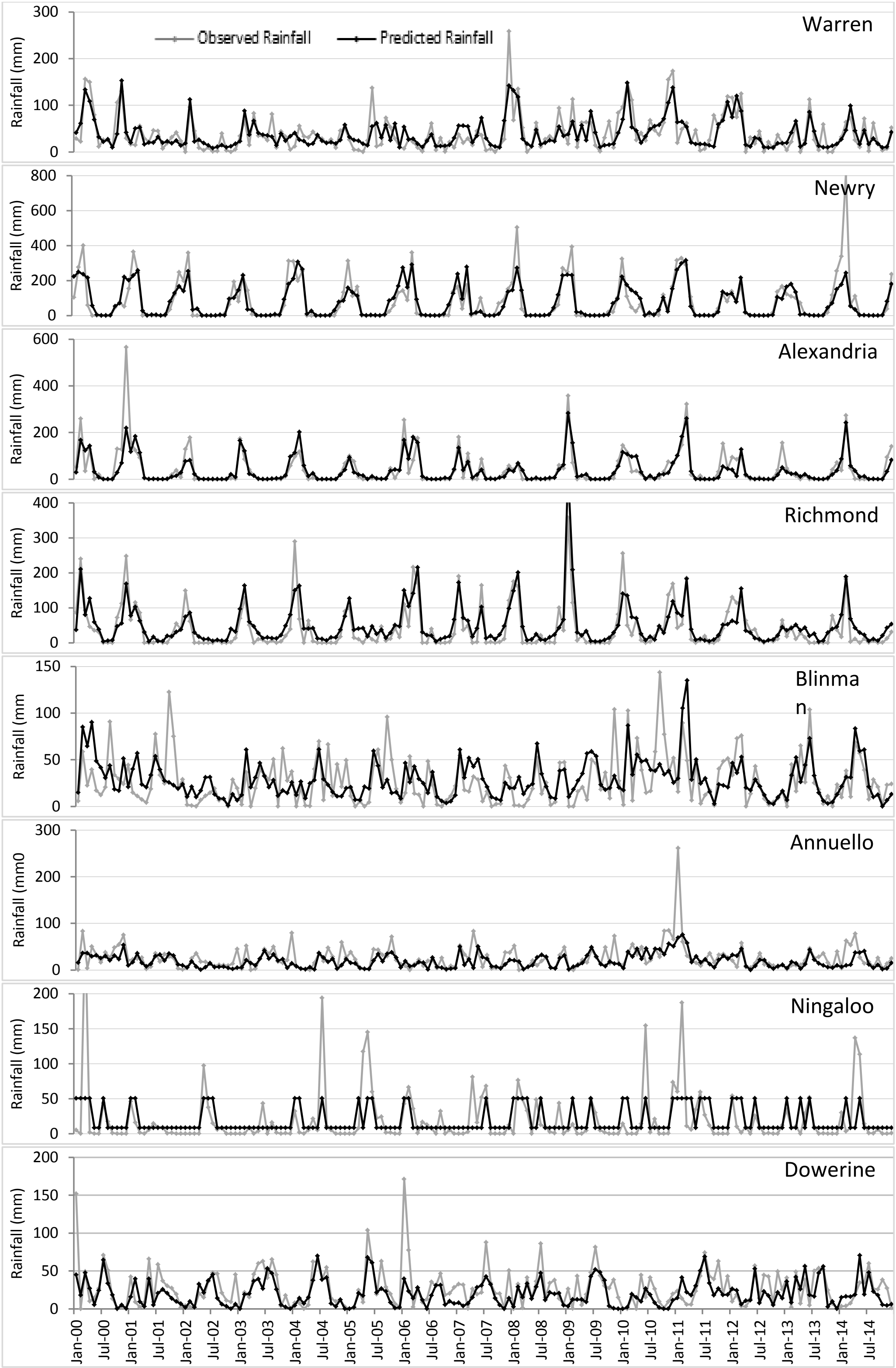

Figure 5.17: Observed rainfall vs. the ANN(1) model predictions in grassland zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP, Rad	24.88	18.56	0.54	0.62
Newry	TMax, TMin, Rad	69.18	36.10	0.53	0.67
Alexandria	TMax, TMin, Rad	40.53	20.28	0.44	0.71
Richmond	TMax, TMin	34.50	24.91	0.58	0.70
Blinman	TMax, TMin, VP	23.90	17.56	0.67	0.20
Annuello	TMax, TMin, VP, Rad	23.10	13.32	0.64	0.26
Ningaloo	TMax, TMin, VP	34.55	18.67	0.75	0.22
Dowerine	TMax, TMin, VP, Rad	20.67	13.40	0.58	0.37
Performance measures for worst input combinations					
Warren	Evap	40.40	30.27	0.88	0.00
Newry	Evap	121.72	85.14	1.24	-0.02
Alexandria	Evap	74.97	47.39	1.04	-0.01
Richmond	Rad	103.56	51.02	1.19	-1.70
Blinman	TMax, TMin, Evap, Rad	27.89	20.73	0.79	-0.09
Annuello	TMax, TMin, Evap	31.37	17.59	0.84	-0.37
Ningaloo	TMax, TMin, Evap	39.09	24.97	1.00	0.00
Dowerine	Rad	25.99	19.83	0.85	0.00

Table 5.17: The ANN(1) model prediction performance with best and worst combinations of input variables in grassland zone.

results presented in this table show that the model with input variables TMax and TMin provides best predictions at Newry and Ningaloo; with TMax, TMin and VP at Warren and Alexandria; with TMax, TMin and Rad at Blinman and Dowerine; with TMax, TMin and Evap at Annuello; and with TMax, TMin, VP and Rad at the location Richmond. The k-NN model provides the worst predictions with input variable Evap at three locations (Warren, Newry and Alexandria), with VP at three locations (Blinman, Ningaloo and Dowerine) and with Rad at two locations (Richmond and Annuello).

In grassland zone, the performance measure RMSE for the k-NN model ranges from 21.24 to 67.75, MAE from 14.13 to 37.17, MASE from 0.50 to 0.71 and CE from 0.21 to 0.68. The performance measures RMSE and MAE indicates that the k-NN model have best predictions at the location Dowerine while MASE indicates at the location Alexandria and CE indicate at the locations Newry and Richmond. Similarly, both performance measures RMSE and MAE indicates the model provides worst predictions at the location Newry; the performance measure MASE indicates at the location Annuello and CE indicates at Blinman and Annuello (see Figure 5.10). The graphical display of observed rainfall and the k-NN model predictions for all eight locations in the grassland classification zone given in Figure 5.18, show that model follow the series patterns.

Tables 5.19, summarizes the performance of all eight models in predicting monthly rainfall with best combinations of input variables in grassland classification zone. Best

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP	28.13	21.19	0.61	0.51
Newry	TMax, TMin	67.75	37.17	0.54	0.68
Alexandria	TMax, TMin, VP	44.40	22.72	0.50	0.65
Richmond	TMax, TMin, VP, Rad	35.67	22.27	0.52	0.68
Blinman	TMax, TMin, Rad	23.73	17.56	0.67	0.21
Annuello	TMax, TMin, Evap	23.84	14.87	0.71	0.21
Ningaloo	TMax, TMin	30.29	15.19	0.61	0.40
Dowerine	TMax, TMin, Rad	21.24	14.13	0.61	0.33
Performance measures for worst input combinations					
Warren	Evap	43.44	31.93	0.93	-0.16
Newry	Evap	120.01	88.83	1.29	0.01
Alexandria	Evap	77.06	46.91	1.03	-0.06
Richmond	Rad	66.08	45.46	1.06	-0.10
Blinman	VP	30.32	20.94	0.80	-0.29
Annuello	Rad	31.55	20.96	1.00	-0.38
Ningaloo	VP	39.93	23.97	0.96	-0.05
Dowerine	VP	28.59	21.36	0.92	-0.21

Table 5.18: The kNN model prediction performance with best and worst combinations of input variables in grassland zone.

results among all models are highlighted in bold. The results show that at least one performance measure indicates that the SVM(RBF) model is best compare to other models in six out of eight locations; ANN(1) model in five locations; ANN(0) in three locations; CR(EM) and MLR in two locations; while SVM(Linear) and CLR(Opt) in one location.

According to the performance measures RMSE, the SVM(RBF) is not best in any location and the ANN(1) model is best at four locations (Warren, Alexandria, Richmond and Annuello). In these four locations, the CLR(Opt) model performance is 7.16% lower in Warren, 6.70% in Alexandria, 3.44% in Richmond and 1.37% lower in Annuello than the ANN(1) model. Similarly, the ANN(0) model is best at Ningaloo with 9.25% and at Blinman with 0.28% and k-NN model at Newry with 3.42% higher performance than CLR(Opt) model. At location Dowerine the CLR(Opt) model outperformed other models.

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 5.11, 5.12, 5.13, 5.14, 5.15, 5.16, 5.17 and 5.18, show that all models follow the series patterns at all locations except Ningaloo. In this location, the ANN(1) model failed to follow the series patterns (see Figure 5.17).

Although ANN(1) model outperformed other models at most locations, but overall performance is unreliable. For example at location Warren, the ANN(1) model predicted single value (42.29) for all 180 observations with four combinations (C8, C9, C11 and C12)(see Table 5.1 for combinations detail) of input variables. Similarly at

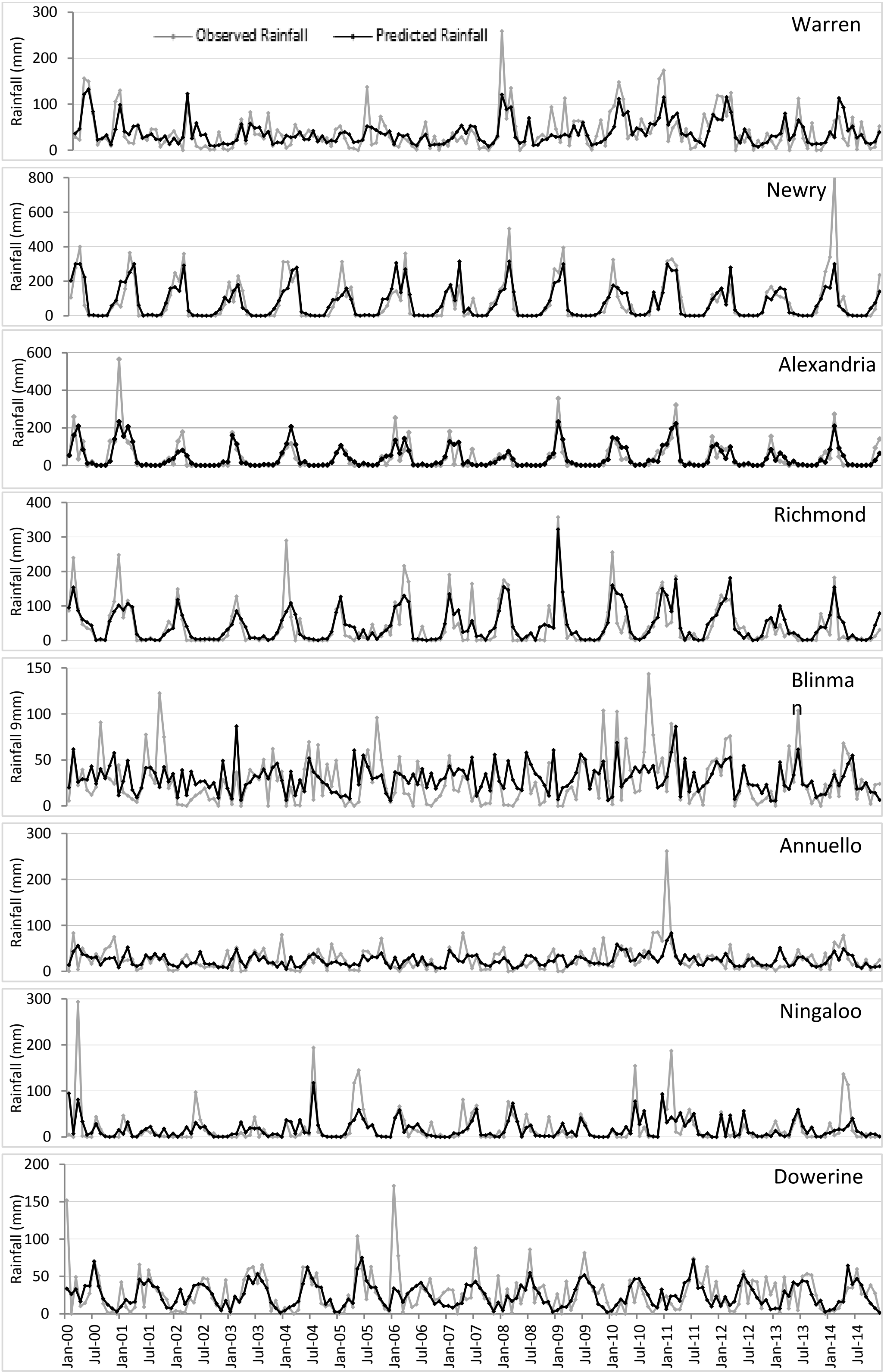

Figure 5.18: Observed rainfall vs. the k-NN model predictions in grassland zone.

the location Alexandria, the ANN(1) model predicted single value (35.62) for all 180 observations with nine combinations (C2, C3, C4, C8, C9, C10, C11, C12, C13 and C14) of input variables; at Richmond predicted 43.17 with eight combinations (C2, C3, C8, C9, C11, C12, C14 and C15); and at Annuello, with seven combinations (C2, C3, C4, C8, C11, C12 and C15).

In summary, based on the performance measure RMSE, the ANN(1) and ANN(0) models are the most suitable models in predicting monthly rainfall in grassland classification zone. The CLR(Opt), MLR and k-NN models were determined to be the second in ranking.

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	ANN (0)	ANN (1)	MLR	KNN
Warren	RMSE	26.80	29.53	28.57	25.13	25.13	**24.88**	27.59	28.13
	MAE	20.10	20.11	19.77	**18.05**	18.58	18.56	19.85	21.19
	MASE	0.58	0.58	0.57	**0.52**	0.54	0.54	0.58	0.61
	CE	0.56	0.46	0.50	0.61	0.61	**0.62**	0.53	0.51
Newry	RMSE	70.15	95.77	84.98	68.68	75.09	69.18	75.10	**67.75**
	MAE	37.91	47.50	43.20	36.84	40.39	**36.10**	40.31	37.17
	MASE	0.55	0.69	0.63	0.54	0.59	**0.53**	0.59	0.54
	CE	0.66	0.37	0.50	0.68	0.61	0.67	0.61	**0.68**
Alexandria	RMSE	43.44	59.25	55.62	43.09	40.97	**40.53**	46.38	44.40
	MAE	22.02	26.67	24.52	**19.82**	20.84	20.28	23.21	22.72
	MASE	0.48	0.58	0.54	**0.43**	0.46	0.44	0.51	0.50
	CE	0.66	0.37	0.45	0.67	0.70	**0.71**	0.61	0.65
Richmond	RMSE	35.73	44.21	41.88	34.94	35.81	**34.50**	39.21	35.67
	MAE	23.10	26.25	25.30	**21.47**	25.66	24.91	26.12	22.27
	MASE	0.54	0.61	0.59	**0.50**	0.60	0.58	0.61	0.52
	CE	0.68	0.51	0.56	0.69	0.68	**0.70**	0.61	0.68
Blinman	RMSE	21.14	21.14	21.26	21.59	**21.08**	23.90	**21.08**	23.73
	MAE	15.40	14.75	**14.52**	14.73	15.02	17.56	15.02	17.56
	MASE	0.59	**0.56**	**0.56**	**0.56**	0.58	0.67	0.58	0.67
	CE	0.37	0.37	0.37	0.35	**0.38**	0.20	**0.38**	0.21
Annuello	RMSE	23.42	23.51	24.16	24.05	23.27	**23.10**	23.46	23.84
	MAE	14.28	13.35	13.94	13.93	13.46	**13.32**	13.72	14.87
	MASE	0.68	**0.64**	0.67	0.67	**0.64**	**0.64**	0.66	0.71
	CE	0.24	0.23	0.19	0.20	0.25	**0.26**	0.23	0.21
Ningaloo	RMSE	28.32	32.62	32.11	27.43	**25.70**	34.55	29.02	30.29
	MAE	14.06	15.12	14.09	**12.79**	13.53	18.67	14.54	15.19
	MASE	0.56	0.60	0.56	**0.51**	0.54	0.75	0.58	0.61
	CE	0.47	0.30	0.32	0.51	**0.57**	0.22	0.45	0.40
Dowerine	RMSE	**19.43**	20.25	21.23	20.66	20.02	20.67	20.07	21.24
	MAE	**12.80**	12.98	13.18	12.87	12.91	13.40	12.88	14.13
	MASE	**0.55**	0.56	0.57	**0.55**	0.56	0.58	**0.55**	0.61
	CE	**0.44**	0.39	0.33	0.37	0.40	0.37	0.40	0.33

Table 5.19: Prediction performance of models with best combination of input variables in grassland classification zone.

5.3 Monthly rainfall predictions in desert zone

In this section, first we present the monthly rainfall prediction results for each model in predicting monthly rainfall with best and worst combinations of input variables in desert classification zone. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models using computational results and time series plots.

Table 5.20 summarizes the prediction performance of the CLR(Opt) model with best and worst combinations of input variables in desert classification zone. Results presented in this table show that the model provides best predictions at Henbury with a combination of input variables TMax, TMin and VP; at Boulia and Marree with TMax, TMin, VP and Rad; at Wiluna with TMax, TMin, Evap and VP; and at Wilcannia with full set of five input variables. The CLR(Opt) model provides worst predictions at Wilcannia and Henbury with input variable Rad; at Wiluna with VP; and at Boulia and Marree with a combination of input variable Evap and Rad.

In the desert zone, the performance measure RMSE for the CLR(Opt) model ranges from 13.15 to 28.93, MAE from 8.80 to 16.88, MASE from 0.51 to 0.74 and CE from 0.36 to 0.59. The performance measures RMSE and MAE indicate that the model provides best predictions at Marree; MASE indicates at Wilcannia; and CE indicates at Boulia while the performance measures MAE, MASE and CE indicate worst predictions at Wiluna and RMSE at Henbury (Figure 5.19). A visual comparison of observed rainfall and model predictions given in Figure 5.20 show that the model follows the series patterns at all locations.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	21.29	12.31	0.51	0.53
Henbury	TMax, TMin, VP	28.93	16.88	0.57	0.49
Boulia	TMax, TMin, VP, Rad	24.78	15.65	0.60	0.59
Marree	TMax, TMin, VP, Rad	13.15	8.80	0.54	0.45
Wiluna	TMax, TMin, Evap, VP	26.12	18.45	0.74	0.36
Performance measures for worst input combinations					
Wilcannia	Rad	30.18	20.97	0.87	0.05
Henbury	Rad	39.90	24.13	0.81	0.04
Boulia	Evap, Rad	40.33	29.87	1.15	-0.10
Marree	Evap, Rad	18.02	14.50	0.89	-0.04
Wiluna	VP	38.73	27.96	1.12	-0.41

Table 5.20: The CLR(Opt) model prediction performance for best and worst combination of input variables in desert zone.

Table 5.21 summarizes the prediction performance of the CR(EM) model with best and worst combinations of input variables for locations in desert classification zone. The results presented in this table show that the model provides best predictions with

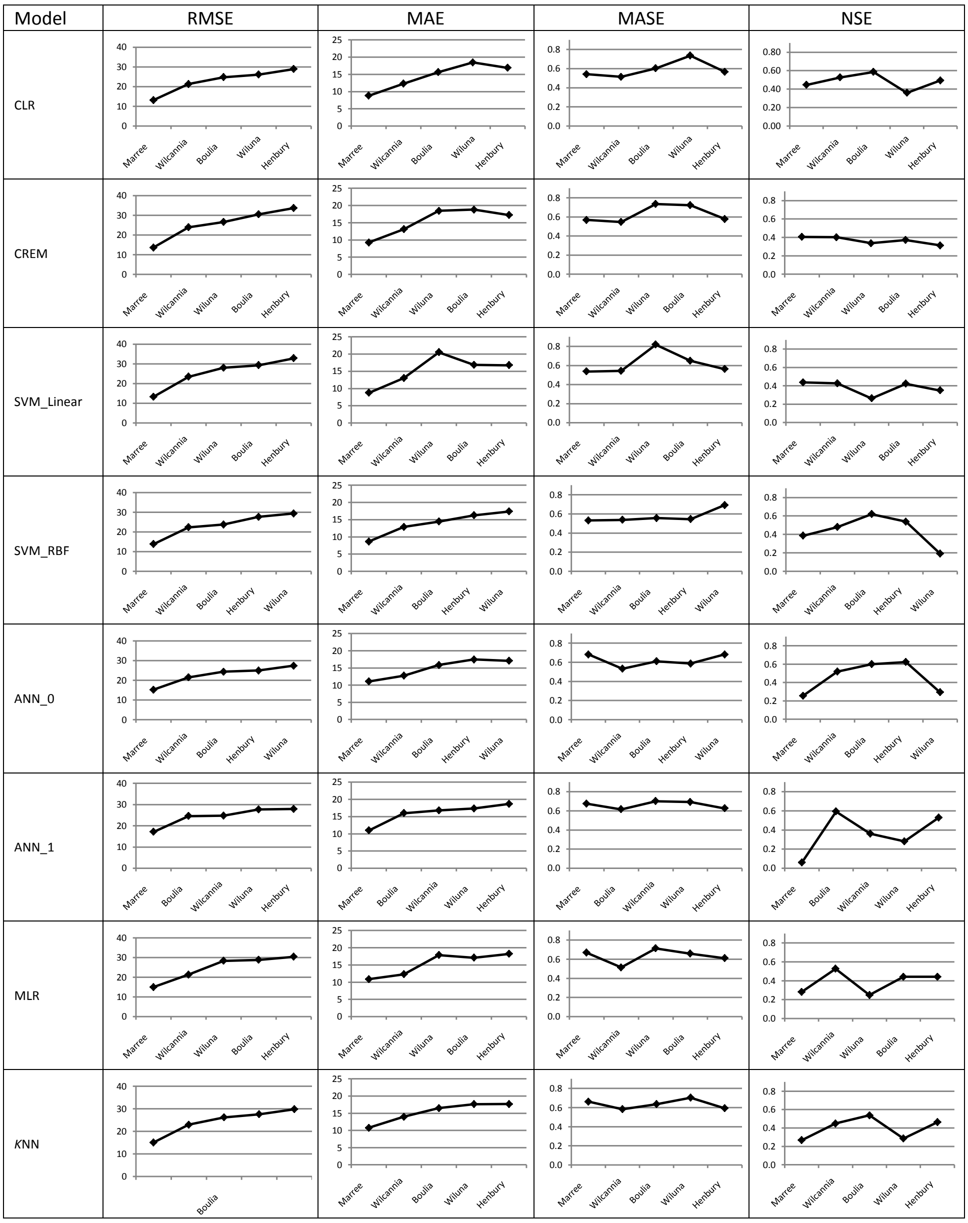

Figure 5.19: The prediction performance of each model in desert zone.

input variables TMax, TMin, Evap and VP at location Wiluna; with TMax, TMin, VP and Rad at two locations Henbury and Marree; and with all five input variables at the remaining two locations Wilcannia and Boulia. The CR(EM) model provides worst predictions at two locations (Henbury and Marree) with input variable Evap; at Wiluna with input variable Rad; and at Wilcannia and Boulia with input variables Evap and Rad.

The performance measure RMSE for the CR(EM) model in desert classification

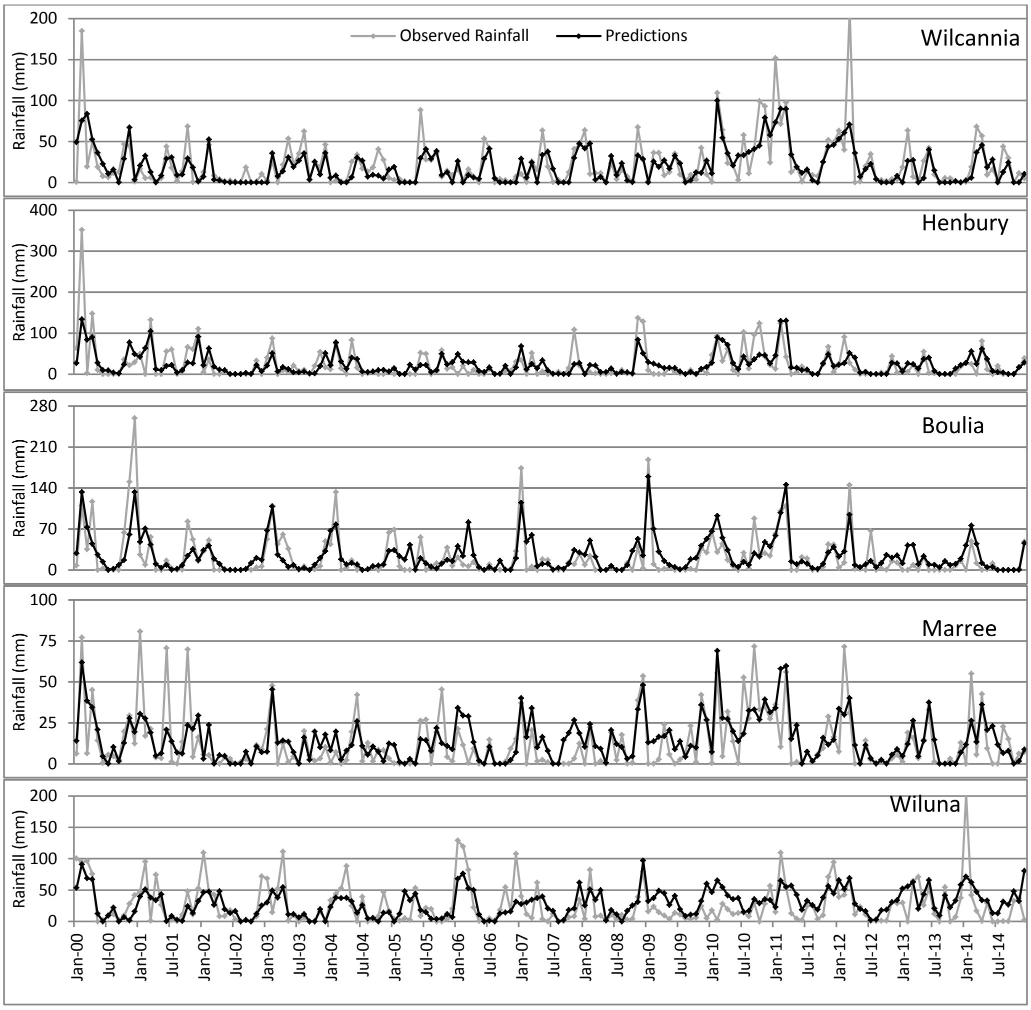

Figure 5.20: Observed rainfall vs. the CLR(Opt) model predictions in desert zone.

zone ranges from 13.61 to 33.70, MAE from 9.25 to 18.80, MASE from 0.55 to 0.74 and CE from 0.31 to 0.41. The performance measures RMSE, MAE and CE indicates that the CR(EM) model provides best predictions at location Marree and MASE indicates at Wilcannia while RMSE and CE indicate worst predictions at Henbury; MAE indicates at Boulia, and MASE indicates at Wiluna (see Table 5.28 and Figure 5.19). The graphical display of observed rainfall and the CR(EM) model predictions for all five locations in the desert classification zone is given in Figure 5.21.

Table 5.22 summarizes the prediction performance of the SVM(Linear) model with best and worst combinations of input variables for locations in desert classification zone. According to the results, the model provides best predictions with input variables TMax, TMin and VP at the location Wiluna and adding Rad in TMax, TMin

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	23.92	13.13	0.55	0.40
Henbury	TMax, TMin, VP, Rad	33.70	17.24	0.58	0.31
Boulia	Evap, VP	30.53	18.80	0.72	0.37
Marree	TMax, TMin, Evap, VP, Rad	13.61	9.25	0.57	0.41
Wiluna	TMax, TMin, Evap, VP	26.61	18.46	0.74	0.34
Performance measures for worst input combinations					
Wilcannia	Evap, Rad	31.56	22.69	0.95	-0.04
Henbury	Evap	40.71	26.03	0.87	0.00
Boulia	Evap, Rad	42.39	34.38	1.32	-0.21
Marree	Evap	17.99	14.45	0.89	-0.04
Wiluna	Rad	33.12	24.07	0.96	-0.03

Table 5.21: The CR(EM) model prediction performance for best and worst combinations of input variables for locations in desert zone.

and VP increased the model performance in the remaining four locations (Wilcannia, Henbury, Boulia and Wiluna). The model provides the worst predictions at four locations (Wilcannia, Henbury, Marree and Wiluna) with input variable Evap and with Rad at the location Boulia.

In desert zone, the performance measure RMSE for the SVM(Linear) model ranges from 13.26 to 32.80, MAE from 8.73 to 20.53, MASE from 0.54 to 0.82 and CE from 0.26 to 0.44. All four performance measures RMSE, MAE, MASE and CE indicate that the SVM(Linear) model provides best predictions at the location Marree while the performance measures MAE, MASE and CE indicate worst predictions at Wiluna and RMSE indicates at Henbury (see Table 5.28 and Figure 5.19). The time series plot of observed rainfall and the model predictions of all five locations of desert classification zone is given in Figure 5.22.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, VP, Rad	23.43	13.02	0.54	0.43
Henbury	TMax, TMin, VP, Rad	32.80	16.77	0.56	0.35
Boulia	TMax, TMin, VP, Rad	29.27	16.86	0.65	0.42
Marree	TMax, TMin, VP, Rad	13.26	8.73	0.54	0.44
Wiluna	TMax, TMin, VP	28.02	20.53	0.82	0.26
Performance measures for worst input combinations					
Wilcannia	Evap	32.35	18.84	0.79	-0.09
Henbury	Evap	43.17	22.00	0.74	-0.13
Boulia	Rad	40.89	21.14	0.81	-0.13
Marree	Evap	18.94	11.84	0.73	-0.15
Wiluna	Evap	36.37	22.91	0.91	-0.24

Table 5.22: The SVM _Linear model prediction performance for best and worst combinations of input variables for locations in desert zone.

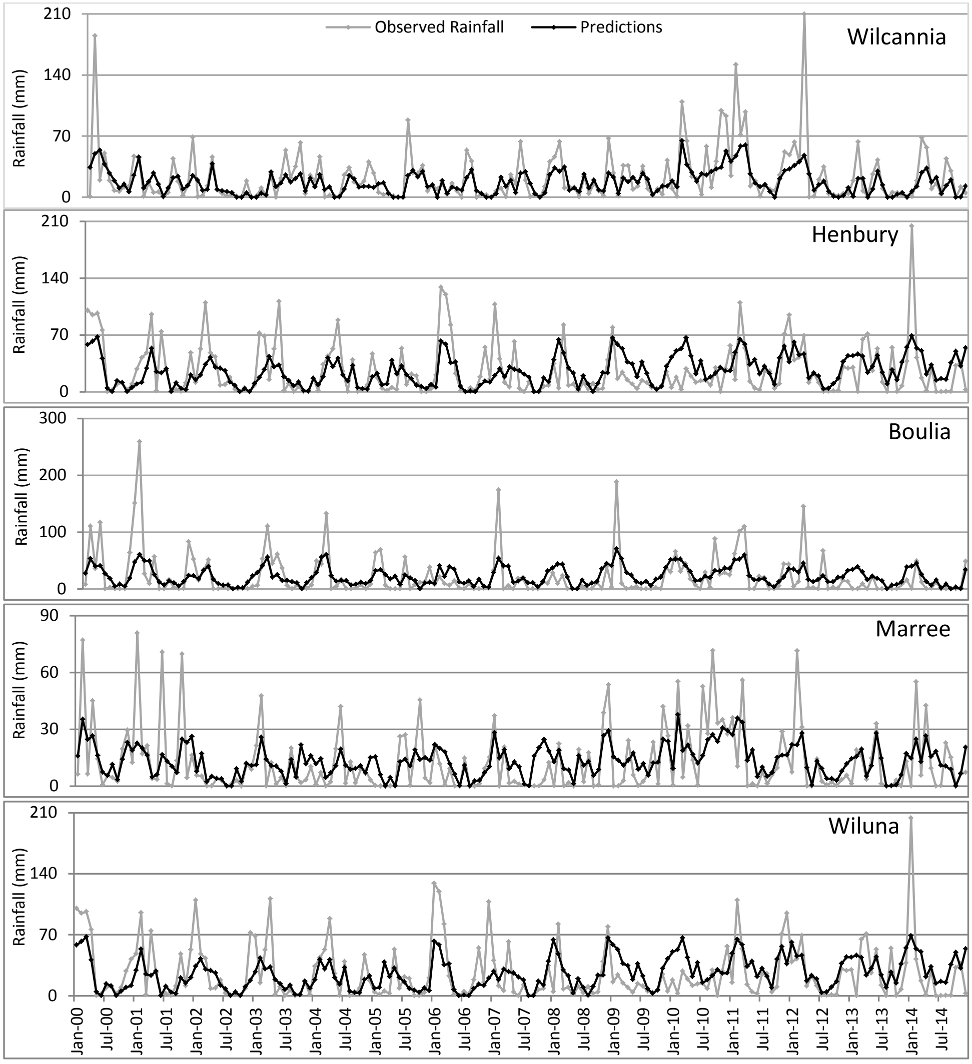

Figure 5.21: Observed rainfall vs. the CR(EM) model predictions in desert zone.

Table 5.23 summarizes the prediction performance of the SVM(RBF) model with best and worst combinations of input variables for locations in desert classification zone. The results presented in this table show that the model provides best predictions with input variables TMax, TMin and Rad at the location Wiluna; with TMax, TMin, VP and Rad at Boulia; with Evap, VP and Rad at Henbury; and with a full set of five variables at Wilcannia and Marree. The SVM(RBF) model provides worst predictions with single input variable Rad at Henbury; with Evap at Marree and Henbury; and with a combination of Evap and Rad at Wilcannia and Boulia.

In the desert zone, the performance measure RMSE for the SVM(RBF) model ranges from 13.84 to 29.34; MAE from 8.63 to 17.37; MASE from 0.53 to 0.69; and CE from 0.19 to 0.62. The performance measures RMSE, MAE and MASE indicate

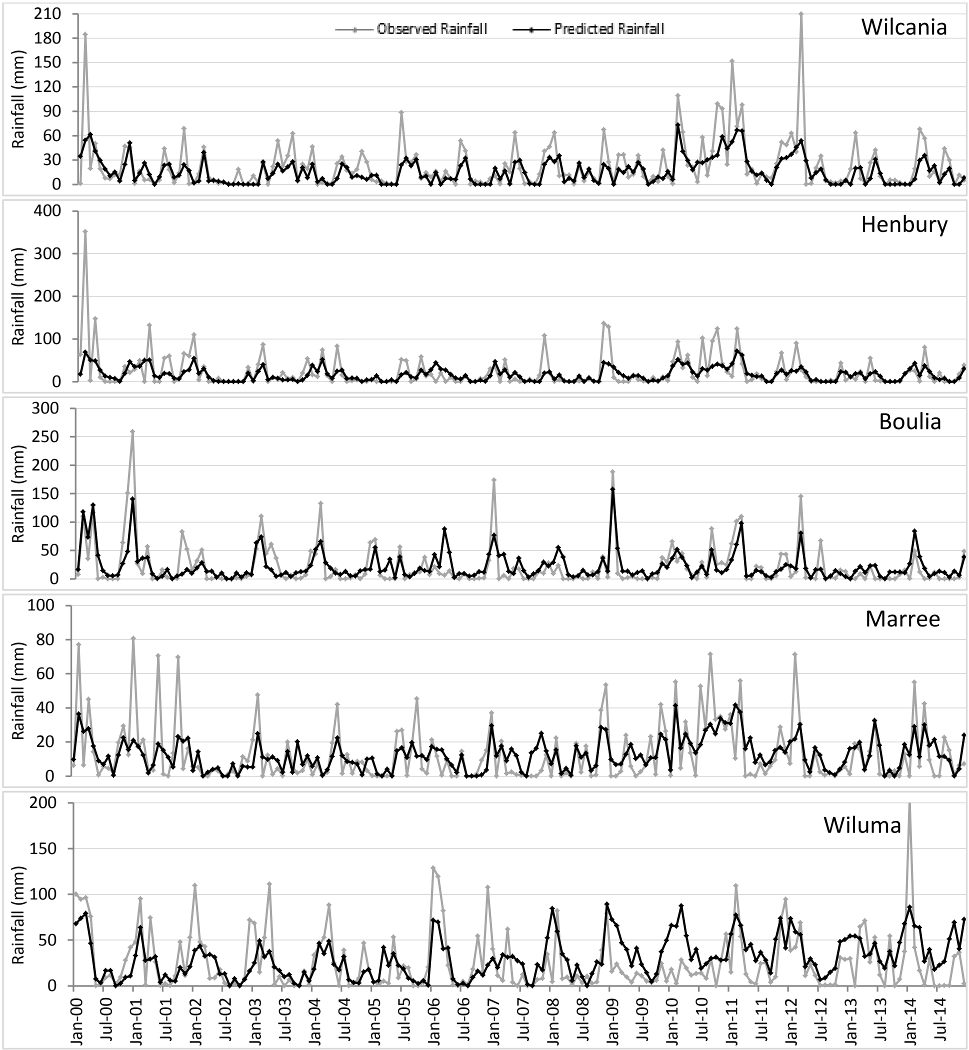

Figure 5.22: Observed rainfall vs. the SVM(Linear) model predictions in desert zone.

that the SVM(RBF) model predictions have lowest prediction error at location Marree and CE indicates at Boulia while all four performance measures indicate the model provides worst predictions at location Wiluna (see Table 5.28 and Figure 5.19). The graphical display of observed rainfall and the model predictions for all five locations in the desert classification zone is given in Figure 5.23.

Table 5.24 presents the prediction performance results of the MLR model with best and worst combinations of input variables for locations in desert classification zone. According to these results, the model provides best predictions with input variables TMax and TMin at the location Wiluna; with TMax, TMin and Evap at the location Boulia; with TMax, TMin and Rad in the location Marree; with TMax, TMin, VP and Rad at the location Henbury; and with all five input variables at the

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	22.32	12.89	0.54	0.48
Henbury	Evap, VP, Rad	27.64	16.26	0.55	0.54
Boulia	TMax, TMin, VP, Rad	23.76	14.46	0.56	0.62
Marree	TMax, TMin, Evap, VP, Rad	13.84	8.63	0.53	0.39
Wiluna	TMax, TMin, Rad	29.34	17.37	0.69	0.19
Performance measures for worst input combinations					
Wilcannia	Evap, Rad	32.80	20.30	0.85	-0.12
Henbury	Evap	42.26	21.47	0.72	-0.08
Boulia	Evap, Rad	41.35	28.69	1.10	-0.15
Marree	Evap	41.20	18.19	1.12	-4.44
Wiluna	Rad	61.53	37.02	1.48	-2.55

Table 5.23: The SVM(RBF) model prediction performance for best and worst combinations of input variables for locations in desert zone.

location Wilcannia. The MLR model provides worst predictions at Henbury with input variable Evap; at Marree with VP; at Wiluna with a combination of input variables Evap and VP; and at Wilcannia and Boulia with a combination of input variables Evap and Rad.

In the desert zone, the performance measure RMSE for the MLR model ranges from 14.98 to 30.39; MAE from 10.86 to 18.21; MASE from 0.51 to 0.71; and CE from 0.25 to 0.53. The performance measures RMSE, and MAE indicate that the MLR model provides best predictions at the location Marree and worst at Henbury while the performance measure MASE and CE indicate best at Wilcannia and worst at Wiluna (see Table 5.28 and Figure 5.19). The graphical display of observed rainfall and the MLR model predictions over the test period for all five locations of desert classification zone is given in Figure 5.24.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	21.29	12.31	0.51	0.53
Henbury	TMax, TMin, VP, Rad	30.39	18.21	0.61	0.44
Boulia	TMax, TMin, Evap	28.78	17.11	0.66	0.44
Marree	TMax, TMin, Rad	14.98	10.86	0.67	0.28
Wiluna	TMax, TMin	28.33	17.88	0.71	0.25
Performance measures for worst input combinations					
Wilcannia	Evap, Rad	31.19	22.15	0.92	-0.02
Henbury	Evap	40.69	26.03	0.87	0.00
Boulia	Evap, Rad	47.38	40.43	1.56	-0.51
Marree	VP	18.41	13.37	0.82	-0.09
Wiluna	Evap, VP	39.91	29.58	1.18	-0.49

Table 5.24: The MLR model prediction performance for best and worst combinations of input variables for locations in desert zone.

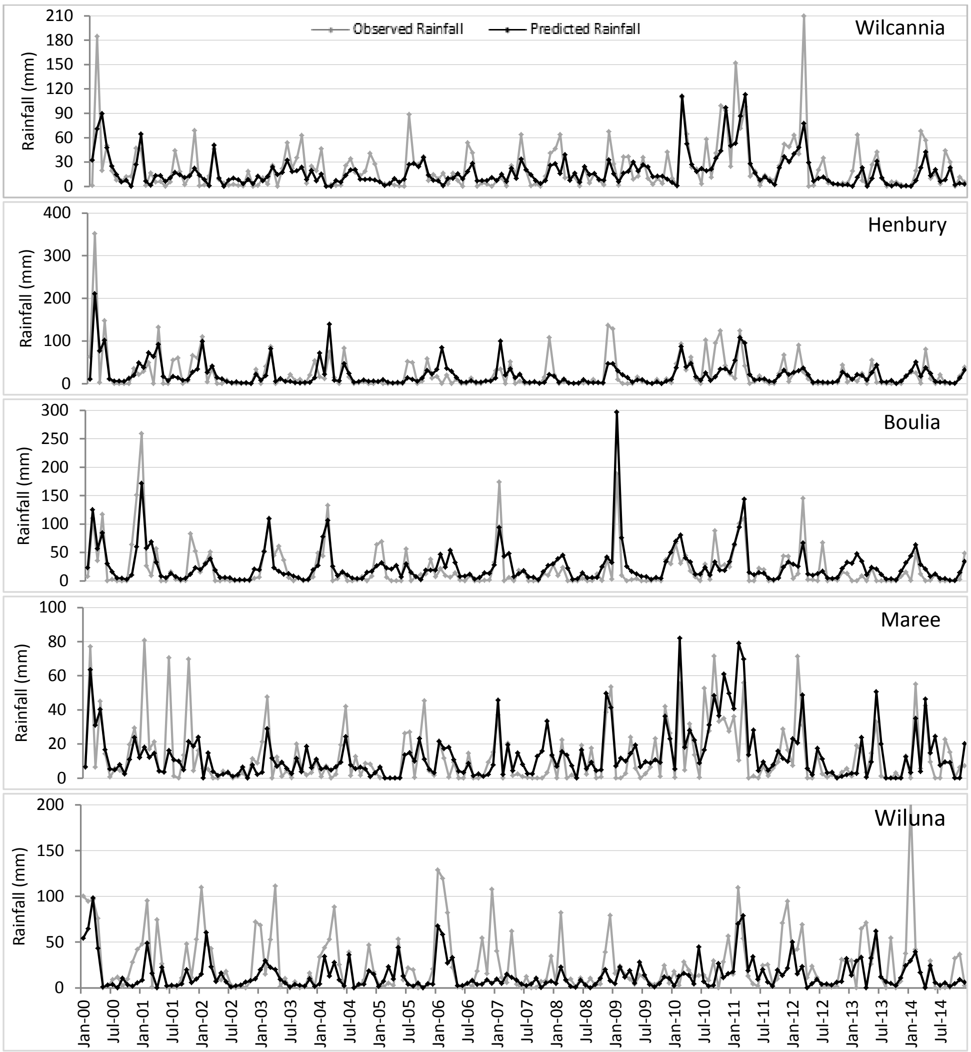

Figure 5.23: Observed rainfall vs. the SVM(RBF) model predictions in desert zone.

Table 5.25 summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables for locations in desert classification zone. According to these results, the model provides best predictions with input variables TMax and TMin in four out of five locations (Henbury, Boulia, Marree and Wiluna) and in the remaining fifth location (Wilcannia), the model provides best predictions with a full set of five input variables. The ANN(0) model provides the worst predictions at Wilcannia and Henbury with input variable Rad; at Marree and Wiluna with VP; and at Boulia with a combination of input variables Evap and Rad.

The performance measure RMSE for the ANN(0) model ranges from 15.23 to 27.43; MAE from 11.09 to 17.48; MASE from 0.53 to 0.68; and CE from 0.26 to 0.62. The performance measures RMSE, and MAE indicates that the ANN(0) model

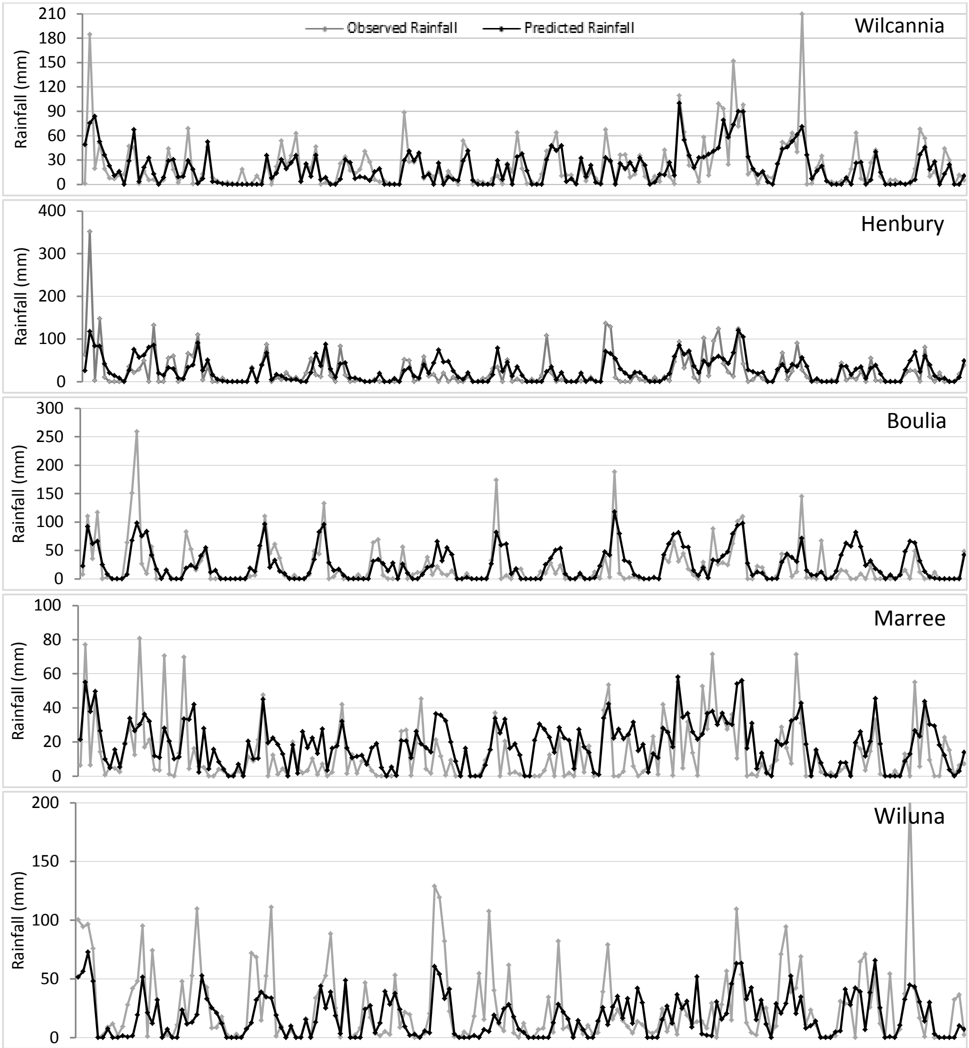

Figure 5.24: Observed rainfall vs. the MLR model predictions in desert zone.

predictions have the lowest error at Marree; MASE indicates at Wilcannia; and CE indicates at Henbury while RMSE indicates the model provides worst predictions at Wiluna; MAE indicates at Henbury; MASE and CE indicate at Marree (see Table 5.28 and Figure 5.19). The graphical display of observed rainfall and the ANN(0) model predictions over the test period for all five locations in the desert classification zone is given in Figure 5.24.

Table 5.26 summarizes the prediction performance of the ANN(1) model with best and worst combinations of input variables of all five locations in desert classification zone. The results presented in this table show that the model with input variables TMax and TMin provides best predictions at location Henbury; with TMax, TMin and Rad at Marree and Wiluna; with TMax, TMin and VP at Boulia; and with VP

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	21.47	12.77	0.53	0.52
Henbury	TMax, TMin	24.95	17.48	0.59	0.62
Boulia	TMax, TMin	24.34	15.85	0.61	0.60
Marree	TMax, TMin	15.23	11.09	0.68	0.26
Wiluna	TMax, TMin	27.43	17.10	0.68	0.29
Performance measures for worst input combinations					
Wilcannia	Rad	30.08	20.64	0.86	0.06
Henbury	Rad	40.74	25.55	0.86	0.00
Boulia	Evap, Rad	43.45	34.18	1.32	-0.27
Marree	VP	22.57	15.21	0.94	-0.63
Wiluna	VP	72.29	41.42	1.65	-3.90

Table 5.25: The ANN(0) model performance in predicting monthly rainfall with best and worst combinations of input variables in desert classification zone.

and Rad at Wilcannia. The model provides worst predictions with input variable Evap at Wilcannia and Boulia; with VP at Marree and Wiluna; and with Rad at Henbury.

The performance measure RMSE for the ANN(1) model ranges from 17.14 to 27.92, MAE from 10.97 to 18.65, MASE from 0.62 to 0.70 and CE from 0.06 to 0.59. The performance measures RMSE and MAE indicates that the ANN(1) model predictions have lowest prediction error at Marree while MASE and CE at Boulia. The performance measures RMSE and MAE indicates the model provides worst predictions at Henbury; MASE indicates at Wilcannia, and CE indicates at Marree.(see Table 5.28 and Figure 5.19). The graphical display of observed rainfall and the model predictions over the test period for all five locations in the desert classification zone is given in Figure 5.26.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	VP, Rad	24.73	16.80	0.70	0.36
Henbury	TMax, TMin	27.92	18.65	0.63	0.53
Boulia	TMax, TMin, VP	24.59	15.99	0.62	0.59
Marree	TMax, TMin, Rad	17.14	10.97	0.67	0.06
Wiluna	TMax, TMin, Rad	27.69	17.36	0.69	0.28
Performance measures for worst input combinations					
Wilcannia	Evap	31.00	21.76	0.91	0.00
Henbury	Rad	40.70	25.33	0.85	0.00
Boulia	Evap	38.57	26.14	1.01	0.00
Marree	VP	24.15	15.54	0.96	-0.87
Wiluna	VP	60.15	37.77	1.51	-2.39

Table 5.26: The ANN(1) model prediction performance for best and worst combinations of input variables for locations in desert zone.

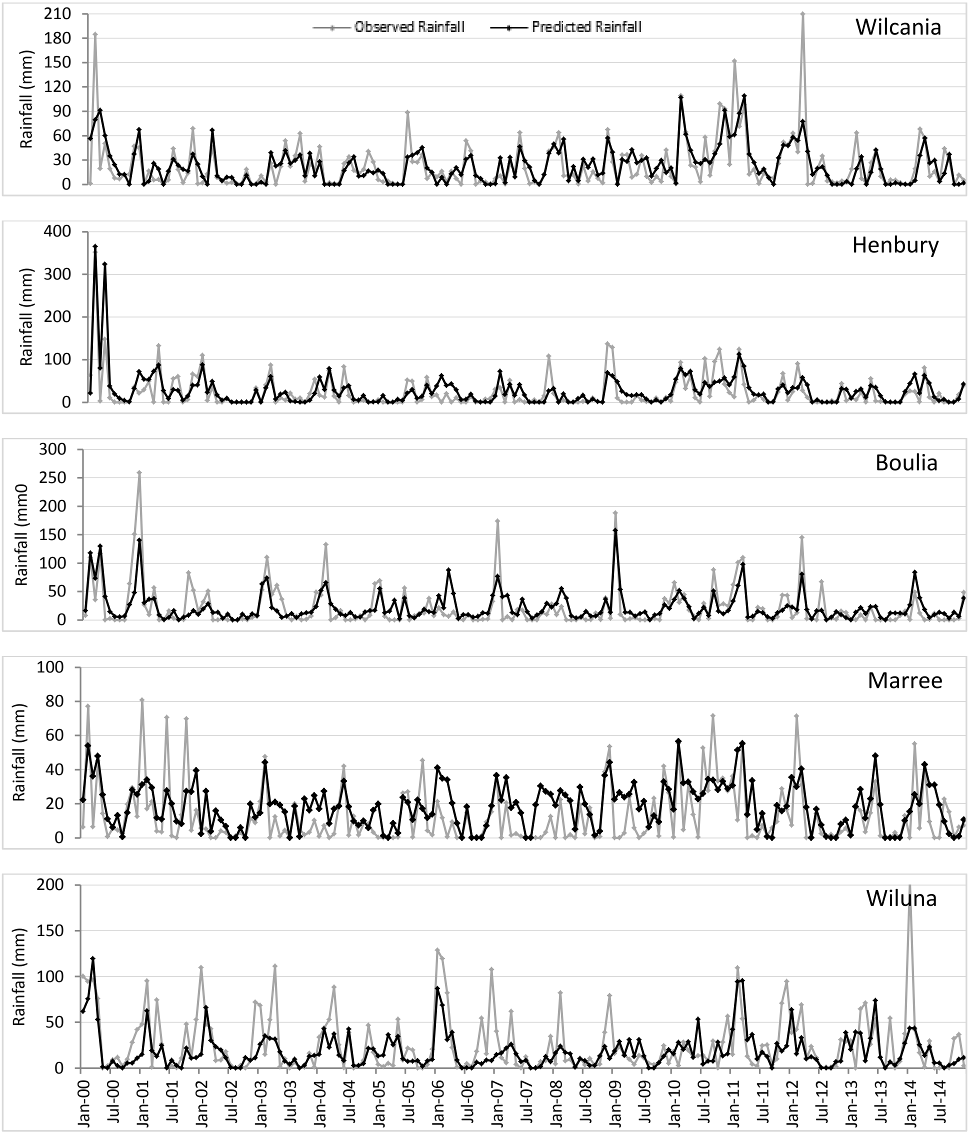

Figure 5.25: Observed rainfall vs. the ANN(0) model predictions in desert zone.

Table 5.27 summarizes the prediction performance of the k-NN model with best and worst combinations of input variables for all five locations in desert classification zone. The results show that the k-NN model with input variables TMax, TMin and Evap provides best predictions at the location Wiluna; with TMax, TMin and VP at Henbury; with TMax, TMin and Rad at Marree; with VP and Rad at Boulia; and with TMax, TMin, Evap and VP at Wilcannia. The model provides worst predictions with input variable Rad at Wilcannia, Henbury and Boulia; with input variable VP at Marree; and with a combination of input variables Evap and VP at the location Wiluna.

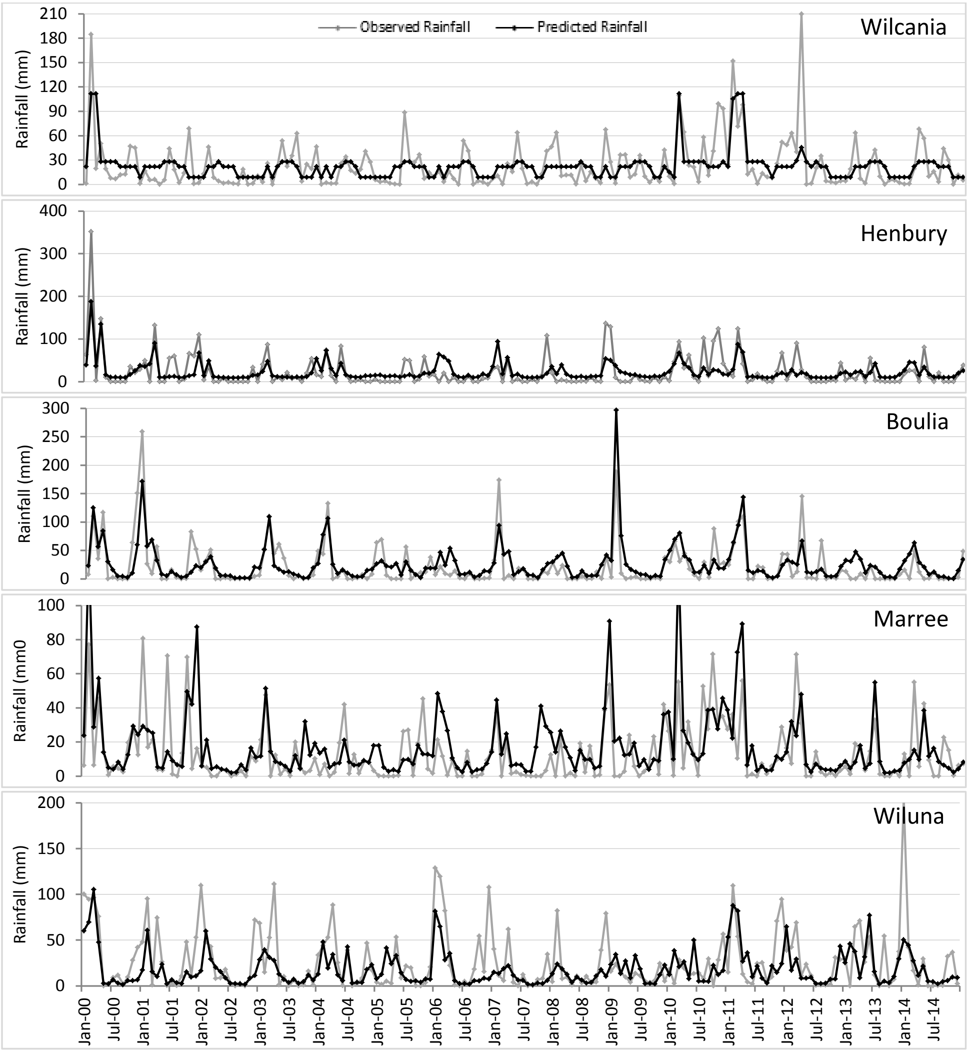

Figure 5.26: Observed rainfall vs. the ANN(1) model predictions in desert zone.

In the desert zone, the performance measure RMSE for the KNN model ranges from 15.12 to 29.80, MAE from 10.77 to 17.67, MASE from 0.58 to 0.70 and CE from 0.27 to 0.54. The performance measures RMSE, MAE and CE indicates that the k-NN model provides best predictions at Marree and MASE indicates at Wilcannia while RMSE and MAE indicate the model provides worst predictions at Henbury; MASE indicates at Wiluna, and CE at Wiluna (see Table 5.28 and Figure 5.19). The graphical display of observed rainfall and the k-NN model predictions over the test period for all five locations of desert classification zone is given in Figure 5.27.

Tables 5.28, summarizes the performance of all eight models in predicting monthly rainfall with best combinations of input variables in desert classification zone. Best results among all models are highlighted in bold. Results presented in this table, show

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	22.96	13.99	0.58	0.45
Henbury	TMax, TMin	29.80	17.67	0.59	0.46
Boulia	TMax, TMin	26.20	16.50	0.64	0.54
Marree	TMax, TMin	15.12	10.77	0.66	0.27
Wiluna	TMax, TMin	27.61	17.64	0.70	0.29
Performance measures for worst input combinations					
Wilcannia	Rad	36.38	22.73	0.95	-0.38
Henbury	Rad	42.26	25.60	0.86	-0.08
Boulia	Rad	42.31	26.35	1.01	-0.21
Marree	VP	22.25	15.78	0.97	-0.59
Wiluna	Evap, VP	37.50	28.19	1.12	-0.32

Table 5.27: The k-NN model prediction performance for best and worst combinations of input variables for locations in desert zone.

that at least one performance measure indicate that CLR(Opt) is best in three locations compare to all other models; the SVM(RBF) also in three locations; ANN(0) in two locations and MLR model in one location. The CR(EM), SVM(Linear), ANN(1) and k-NN models are not best in any location.

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	ANN (0)	ANN (1)	MLR	KNN
Wilcannia	RMSE	**21.29**	23.92	23.43	22.32	21.47	24.73	**21.29**	22.96
	MAE	**12.31**	13.13	13.02	12.89	12.77	16.80	**12.31**	13.99
	MASE	**0.51**	0.55	0.54	0.54	0.53	0.70	**0.51**	0.58
	CE	**0.53**	0.40	0.43	0.48	0.52	0.36	**0.53**	0.45
Henbury	RMSE	28.93	33.70	32.80	27.64	**24.95**	27.92	30.39	29.80
	MAE	16.88	17.24	16.77	**16.26**	17.48	18.65	18.21	17.67
	MASE	0.57	0.58	0.56	**0.55**	0.59	0.63	0.61	0.59
	CE	0.49	0.31	0.35	0.54	**0.62**	0.53	0.44	0.46
Boulia	RMSE	24.78	30.53	29.27	**23.76**	24.34	24.59	28.78	26.20
	MAE	15.65	18.80	16.86	**14.46**	15.85	15.99	17.11	16.50
	MASE	0.60	0.72	0.65	**0.56**	0.61	0.62	0.66	0.64
	CE	0.59	0.37	0.42	**0.62**	0.60	0.59	0.44	0.54
Marree	RMSE	**13.15**	13.61	13.26	13.84	15.23	17.14	14.98	15.12
	MAE	8.80	9.25	8.73	**8.63**	11.09	10.97	10.86	10.77
	MASE	0.54	0.57	0.54	**0.53**	0.68	0.67	0.67	0.66
	CE	**0.45**	0.41	0.44	0.39	0.26	0.06	0.28	0.27
Wiluna	RMSE	**26.12**	26.61	28.02	29.34	27.43	27.69	28.33	27.61
	MAE	18.45	18.46	20.53	17.37	**17.10**	17.36	17.88	17.64
	MASE	0.74	0.74	0.82	0.69	**0.68**	0.69	0.71	0.70
	CE	**0.36**	0.34	0.26	0.19	0.29	0.28	0.25	0.29

Table 5.28: Prediction performance of models in desert classification zone.

According to our primary performance measure RMSE, the CLR(Opt) model outperformed other models at three locations (Wilcannia, Marree and Wiluna). The

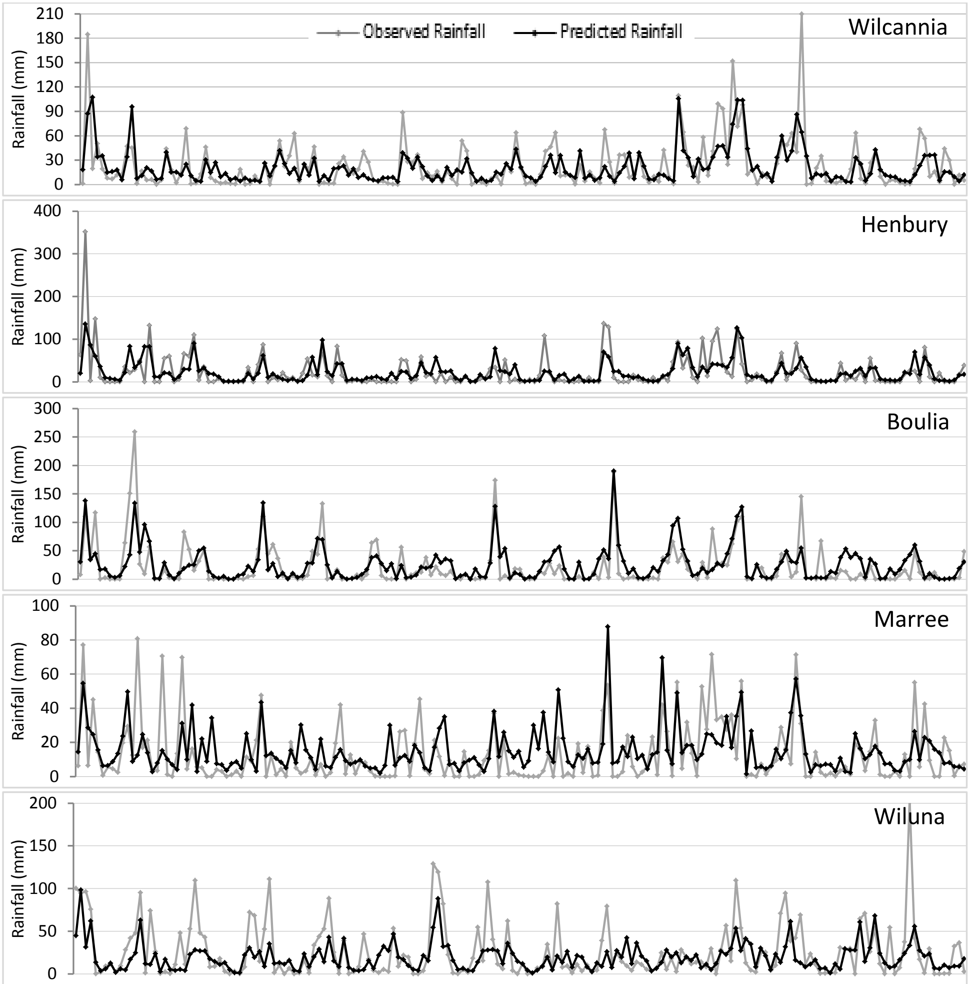

Figure 5.27: Observed rainfall vs. the k-NN model predictions in desert zone.

SVM(RBF) model is best at Boulia with 4.12% higher performance and the ANN(0) model at Henbury with 13.76% higher performance. Results demonstrate that the CLR(Opt) model is superior to other models in predicting monthly rainfall in desert classification zone.

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 5.20, 5.21, 5.22, 5.23, 5.24, 5.25, 5.26 and 5.27, show that all models follow the series patterns at all sites in desert classification zone.

In summary, based on the performance measure RMSE, the CLR(Opt) model is the most suitable model in finding the pattern and trends of the observations compared to other models at all sites in desert classification zone.

5.4 Monthly rainfall predictions in tropical and subtropical zones

In this section, first we present the monthly rainfall prediction results for each model in predicting monthly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models using computational results and time series plots.

Table 5.29 summarizes the prediction performance of the CLR(Opt) model with best and worst combinations of input variables in tropical and subtropical zones. Results presented in this table show that the model provides best predictions at Palmerville and Yamba with input variables TMax, TMin and Rad; at Katherine with TMax, TMin, VP and Rad; and at Fairymead with VP and Rad. The CLR(Opt) model provides worst predictions in three out of four locations (Katherine, Yamba and Fairymead) with input variable Evap and at Palmerville with input variable Rad.

In tropical and subtropical zones, the performance measure RMSE for CLR(Opt) model ranges from 58.17 to 94.89, MAE from 34.42 to 54.81, MASE from 0.43 to 0.58 and CE from 0.34 to 0.81. All four performance measures indicate the model provides best predictions at Katherine while the performance measures RMSE, and CE indicates the model provides worst predictions at Fairymead and MAE and MASE indicates at Yamba (Figure 5.28). The graphical display of observed rainfall and the model predictions given in Figure 5.29, show that the model predictions follow the series patterns in all locations.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP, Rad	58.17	34.42	0.43	0.81
Palmerville	Tropical	TMax, TMin, Rad	70.15	41.21	0.52	0.71
Yamba	Subtropical	TMax, TMin, Rad	75.04	54.81	0.58	0.38
Fairymead	Subtropical	VP, Rad	94.89	47.04	0.55	0.34
Performance measures for worst input combinations						
Katherine	Tropical	Evap	122.49	92.30	1.15	0.17
Palmerville	Tropical	Rad	127.27	97.99	1.24	0.04
Yamba	Subtropical	Evap	95.16	77.33	0.81	0.01
Fairymead	Subtropical	Evap	116.60	70.60	0.83	0.02

Table 5.29: The CLR(Opt) model prediction performance with best and worst combinations of input variables in tropical and subtropical zones.

Table 5.30 summarizes the prediction performance of the CR(EM) model with best and worst combinations of input variables in tropical and subtropical zones. The results presented in this table show that the model provides best predictions with

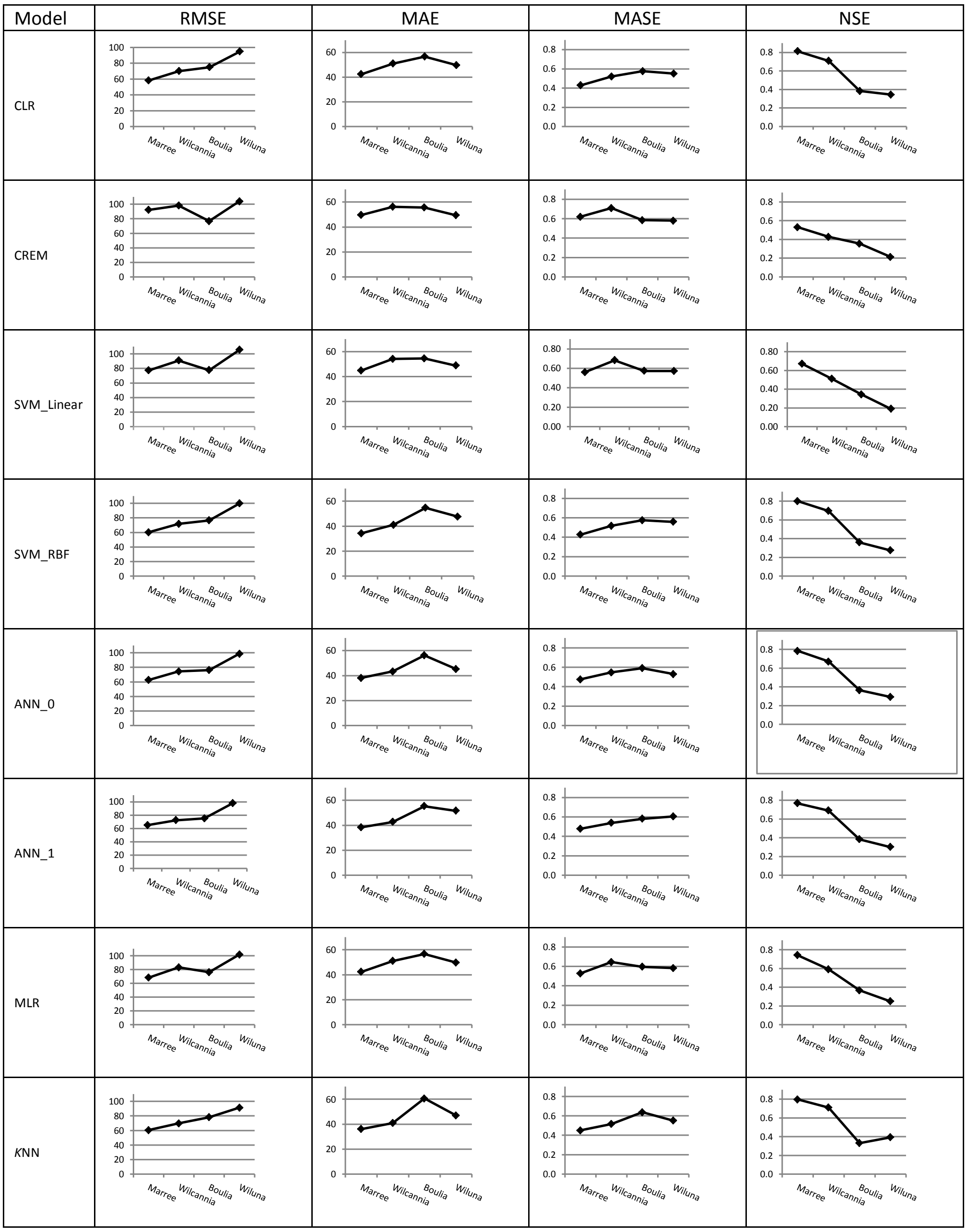

Figure 5.28: The prediction performance of models in tropical and subtropical zones.

input variables TMax, TMin and VP at Katherine; with TMax, TMin and Rad at Yamba; with TMax, TMin, VP and Rad at Fairymead; and with input variables Evap, VP and Rad at Palmerville. The model provides worst predictions in three out of four locations (Katherine, Palmerville and Yamba) with input variable Evap and in the fourth location Fairymead with input variable Rad.

The performance measure RMSE in tropical and subtropical zones for the CR(EM)

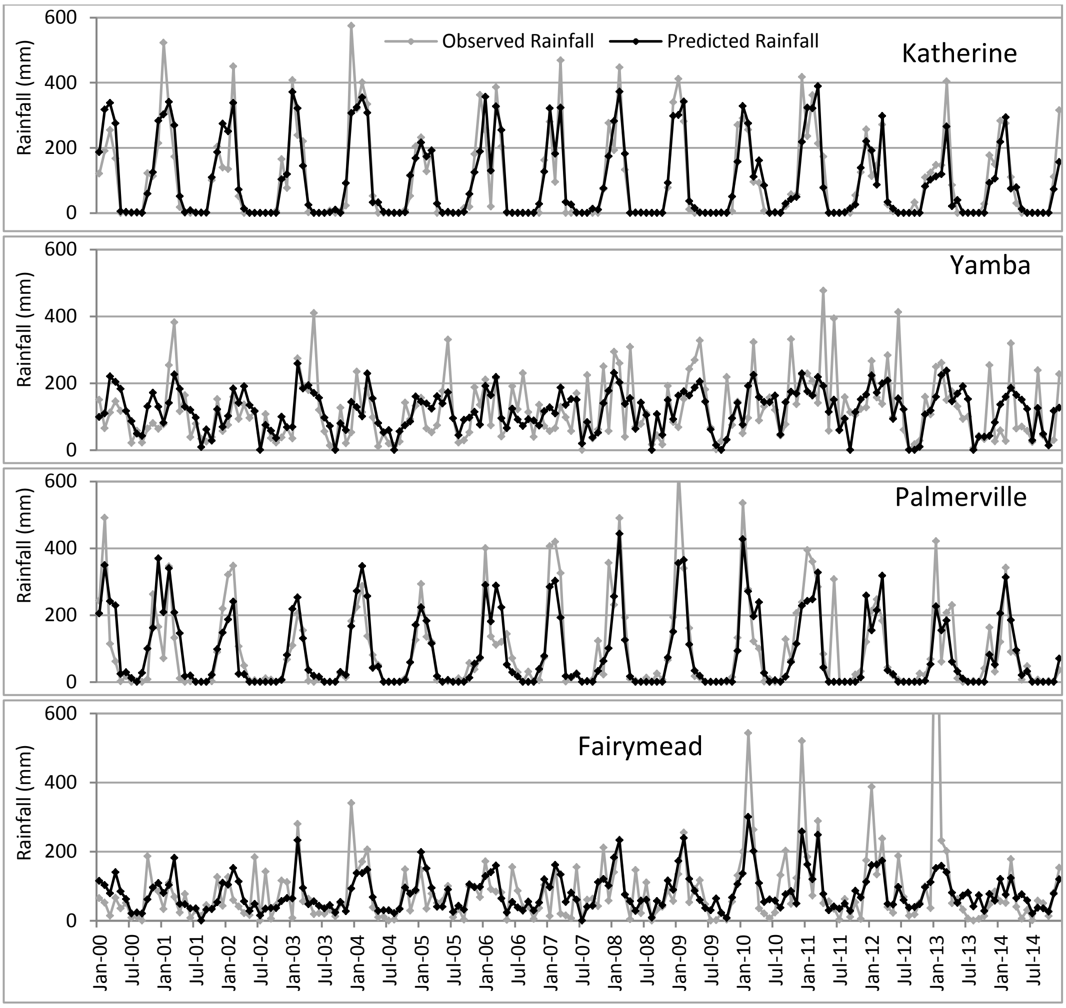

Figure 5.29: Observed rainfall vs. the CLR(Opt) model predictions in tropical and subtropical classification zones.

model ranges from 76.77 to 103.97, MAE from 49.47 to 56.21, MASE from 0.58 to 0.71 and CE from 0.21 to 0.53. The performance measures RMSE indicates that the model provides best predictions at Yamba; MAE and MASE indicate at Fairymead and CE indicates at Katherine while the performance measures RMSE, and CE indicates worst predictions at Fairymead and MAE and MASE indicates at Palmerville (see Figure 5.28). The time series plot of observed rainfall and the model predictions of all four locations in the tropical and subtropical classification zones is given in Figure 5.30.

Table 5.31 summarizes the prediction performance of the SVM(Linear) model with best and worst combinations of input variables in tropical and subtropical zones. According to the results, the model provides best predictions at Fairymead with input

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP	92.08	49.67	0.62	0.53
Palmerville	Tropical	Evap, VP, Rad	98.18	56.21	0.71	0.43
Yamba	Subtropical	TMax, TMin, Rad	76.77	55.59	0.58	0.35
Fairymead	Subtropical	TMax, TMin, VP, Rad	103.97	49.47	0.58	0.21
Performance measures for worst input combinations						
Katherine	Tropical	Evap	125.29	97.19	1.21	0.13
Palmerville	Tropical	Evap	127.20	97.56	1.23	0.04
Yamba	Subtropical	Evap	95.03	76.84	0.81	0.01
Fairymead	Subtropical	Rad	117.29	71.20	0.83	0.00

Table 5.30: The CR(EM) model prediction performance with best and worst combinations of input variables in tropical and subtropical zones.

variables TMax, TMin and VP; at Katherine with TMax, TMin, Evap and VP; at Palmerville with Evap, VP and Rad; and at Yamba with a full set of input variables. The SVM(Linear) model provides worst predictions in three out of four locations (Katherine, Palmerville and Yamba) with input variable Evap and with input variable Rad in the remaining fourth location Fairymead.

In tropical and subtropical zones, the performance measure RMSE for the SVM(Linear) model ranges from 77.19 to 105.46, MAE from 44.88 to 54.57, MASE from 0.56 to 0.68 and CE from 0.19 to 0.67. All four performance measures indicate that the SVM(Linear) model provide best predictions at location Katherine while RMSE and CE indicate the model provides worst predictions at Fairymead; MAE indicates at Yamba; MASE indicates at Palmerville (see Figure 5.28). The graphical display of observed rainfall and the model predictions for all four locations of tropical and subtropical zones is given in Figure 5.31.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP	77.19	44.88	0.56	0.67
Palmerville	Tropical	Evap, VP, Rad	90.77	54.15	0.68	0.51
Yamba	Subtropical	TMax, TMin, Evap, VP, Rad	77.48	54.57	0.57	0.34
Fairymead	Subtropical	TMax, TMin, VP	105.46	48.85	0.57	0.19
Performance measures for worst input combinations						
Katherine	Tropical	Evap	141.41	95.85	1.19	-0.11
Palmerville	Tropical	Evap	143.86	89.11	1.12	-0.23
Yamba	Subtropical	Evap	96.16	72.47	0.76	-0.01
Fairymead	Subtropical	Rad	119.81	62.75	0.74	-0.05

Table 5.31: The SVM(Linear) model prediction performance with best and worst combination of input variables in tropical and subtropical classification zones.

Table 5.32 summarizes the prediction performance of the SVM(RBF) model with best and worst combinations of input variables of selected locations in tropical and

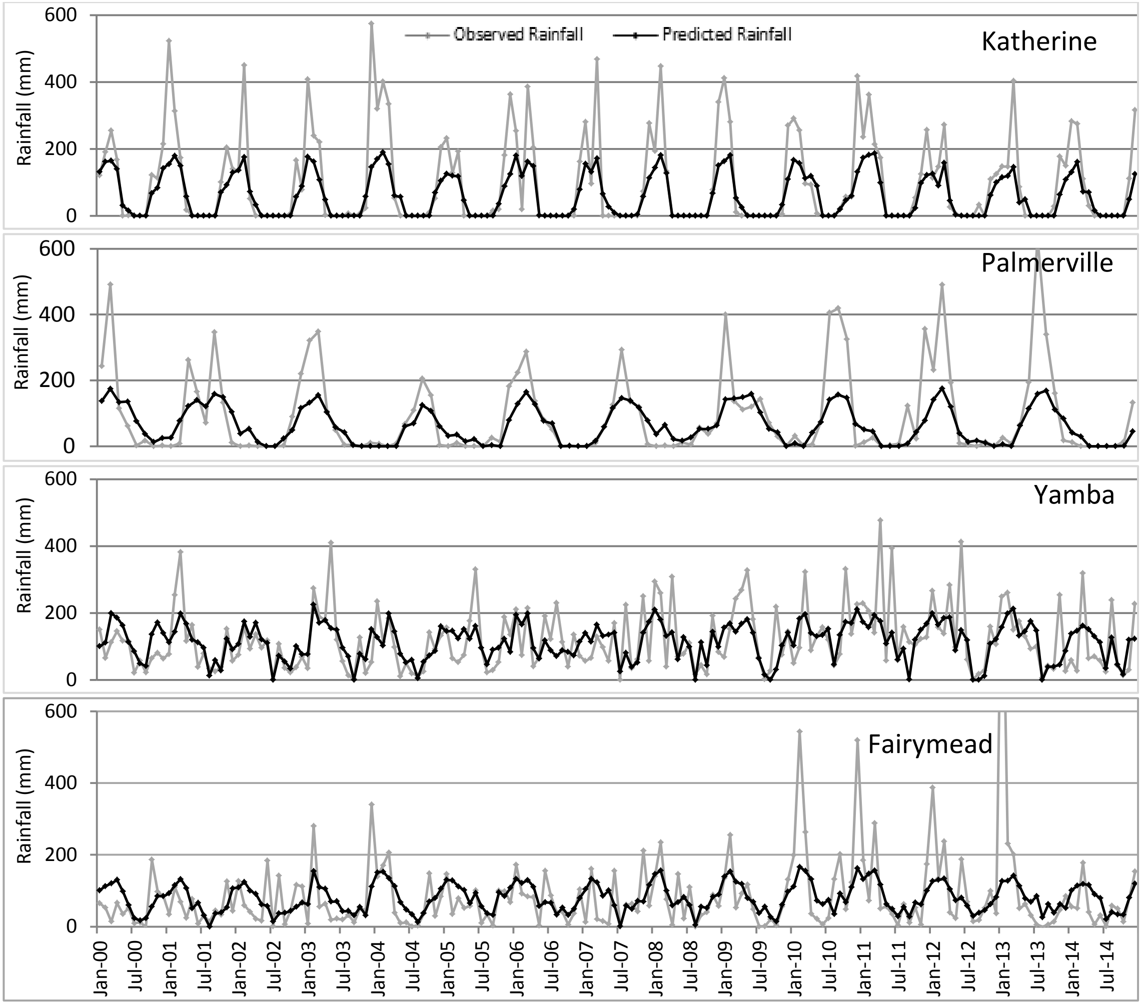

Figure 5.30: Observed rainfall vs. the CR(EM) model predictions in tropical and subtropical classification zones.

subtropical zones. The results presented in this table show that the model with input variables TMax, TMin and Rad provides best predictions at two locations (Palmerville and Yamba), with input variables TMax, TMin and VP at Katherine and with VP and Rad at the location Fairymead. The SVM(RBF) model provides worst predictions with input variable Evap at the location Katherine; with Rad at Yamba and fairymead; and with combinations of Evap and Rad at Palmerville.

In tropical and subtropical zones, the performance measure RMSE for the SVM(RBF) model ranges from 60.15 to 99.81, MAE from 34.26 to 54.65, MASE from 0.43 to 0.57 and CE from 0.27 to 0.80. All four performance measures indicate that the SVM(RBF) model provides best predictions at location Katherine while the performance measures

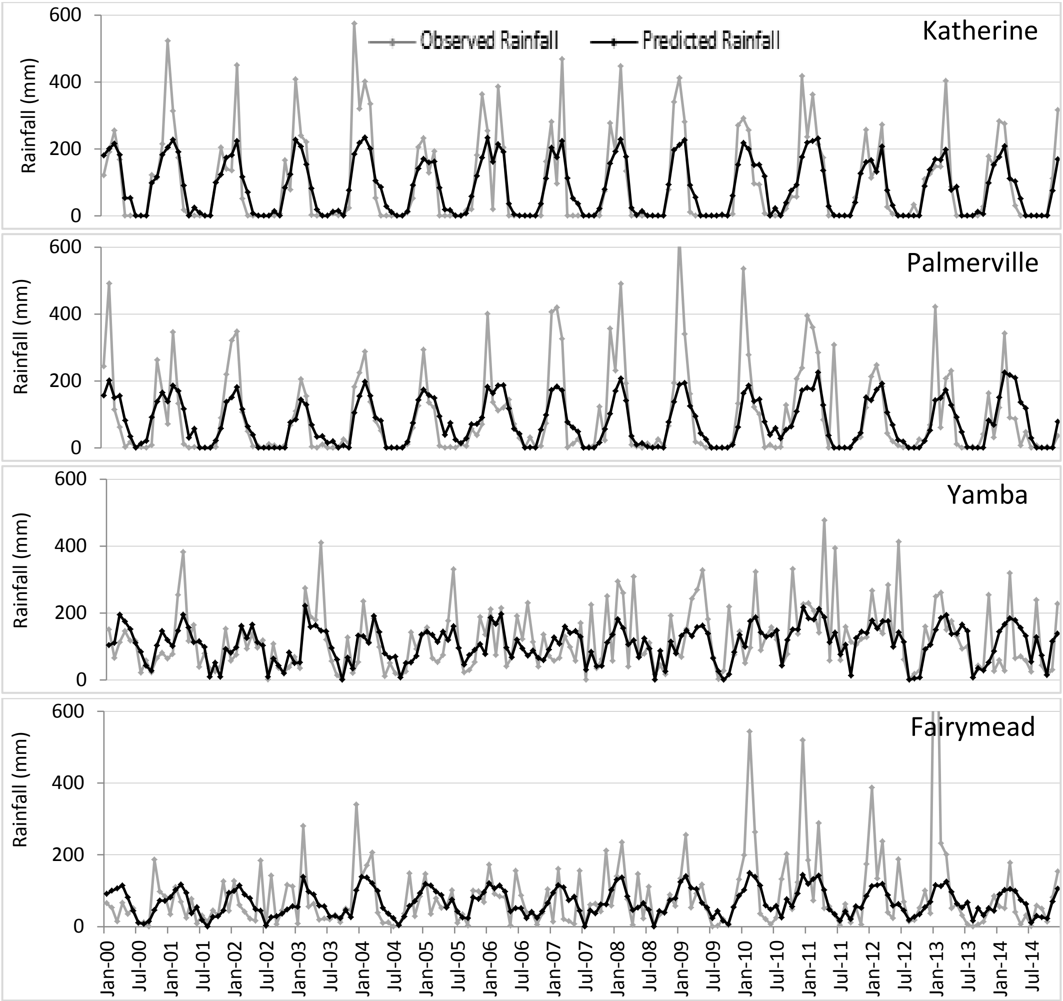

Figure 5.31: Observed rainfall vs. the SVM(Linear) model predictions in tropical and subtropical classification zones.

RMSE, and CE indicates the model provides worst predictions at location Fairymead; MAE indicates at location Yamba; MASE indicates at location Palmerville (see Figure 5.28). The graphical display of observed rainfall and the SVM(RBF) model predictions for all four locations of the tropical and subtropical classification zones are given in Figure 5.32.

Table 5.33 summarizes the prediction performance of the MLR model with best and worst combinations of input variables for all selected locations of tropical and subtropical zones. According to the results, the MLR model provides best predictions at Palmerville with input variable VP; at Fairymead with TMax, TMin and VP; at Yamba with TMax, TMin and Rad; and at Katherine with TMax, TMin, Evap

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP	60.15	34.26	0.43	0.80
Palmerville	Tropical	TMax, TMin, Rad	71.67	41.03	0.52	0.69
Yamba	Subtropical	TMax, TMin, Rad	76.48	54.65	0.57	0.36
Fairymead	Subtropical	VP, Rad	99.81	47.64	0.56	0.27
Performance measures for worst input combinations						
Katherine	Tropical	Evap	131.99	87.22	1.09	0.03
Palmerville	Tropical	Evap, Rad	148.60	94.58	1.19	-0.31
Yamba	Subtropical	Rad	95.57	71.54	0.75	0.00
Fairymead	Subtropical	Rad	118.97	62.49	0.73	-0.03

Table 5.32: The SVM(RBF) model prediction performance for best and worst combination of input variables for locations in tropical and subtropical zones.

and VP. The model provides worst predictions in Katherine and Yamba with input variable Evap; in Palmerville and Fairymead with input variable Rad.

In tropical and subtropical zones, the performance measure RMSE for the MLR model ranges from 68.26 to 101.59, MAE from 42.32 to 56.64, MASE from 0.53 to 0.64 and CE from 0.25 to 0.74. All four performance measures indicate that the MLR model provides best predictions at location Katherine while the performance measures RMSE, and CE indicate the model provides worst predictions at Fairymead; MAE indicates at location Yamba; MASE indicates at location Palmerville (see Figure 5.28). The graphical display of observed rainfall and the MLR model predictions for locations in the tropical and subtropical classification zones are given in Figure 5.33.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP	68.26	42.32	0.53	0.74
Palmerville	Tropical	VP	83.00	51.02	0.64	0.59
Yamba	Subtropical	TMax, TMin, Rad	76.06	56.64	0.60	0.37
Fairymead	Subtropical	TMax, TMin, VP	101.59	49.71	0.58	0.25
Performance measures for worst input combinations						
Katherine	Tropical	Evap	122.64	95.58	1.19	0.17
Palmerville	Tropical	Rad	127.27	98.00	1.24	0.04
Yamba	Subtropical	Evap	95.16	77.33	0.81	0.01
Fairymead	Subtropical	Rad	117.17	70.85	0.83	0.00

Table 5.33: The MLR model prediction performance with best and worst combination of input variables in tropical and subtropical zones.

Table 5.34 summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables in tropical and subtropical zones. According to the results, the model provides best predictions with TMax, TMin and Rad at Palmerville and Yamba; with TMax, TMin, VP and Rad at Katherine; with TMax, TMin, Evap and VP at Fairymead. The model provides worst predictions

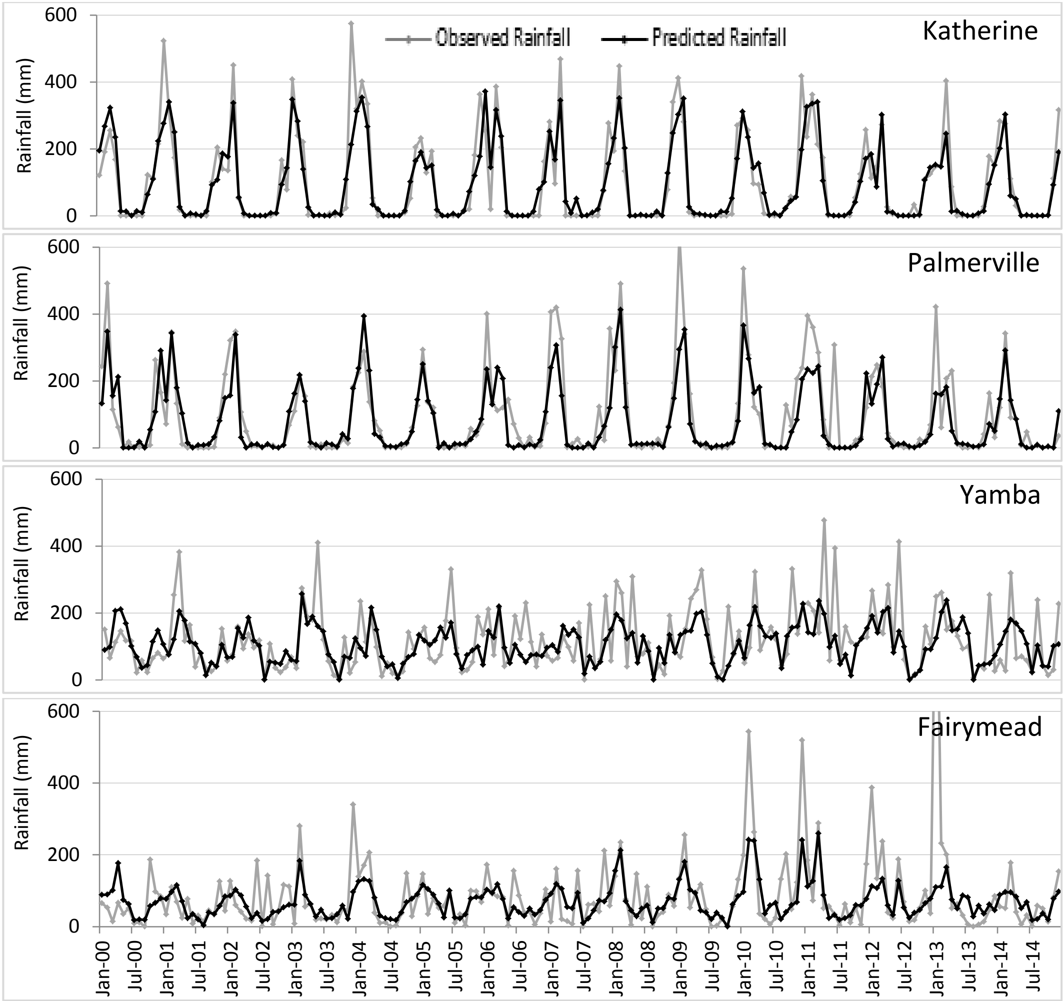

Figure 5.32: Observed rainfall vs. the SVM(RBF) model predictions in tropical and subtropical classification zones.

with input variable Evap in Katherine and Yamba; with Rad in Fairymead; and with a combination of input variables Evap and Rad in the location Palmerville.

In tropical/subtropical zones, the performance measure RMSE for the ANN(0) model ranges from 62.51 to 98.59, MAE from 38.15 to 56.20, MASE from 0.48 to 0.59 and CE from 0.29 to 0.78. All four performance measures indicate that the ANN(0) model provides best predictions in Katherine while the performance measures RMSE, and CE indicates the model provides worst predictions in Fairymead and MAE and MASE indicates in Yamba (Figure 5.28). The graphical display of observed rainfall and the ANN(0) model predictions for locations in tropical and subtropical classification zones is given in Figure 5.34.

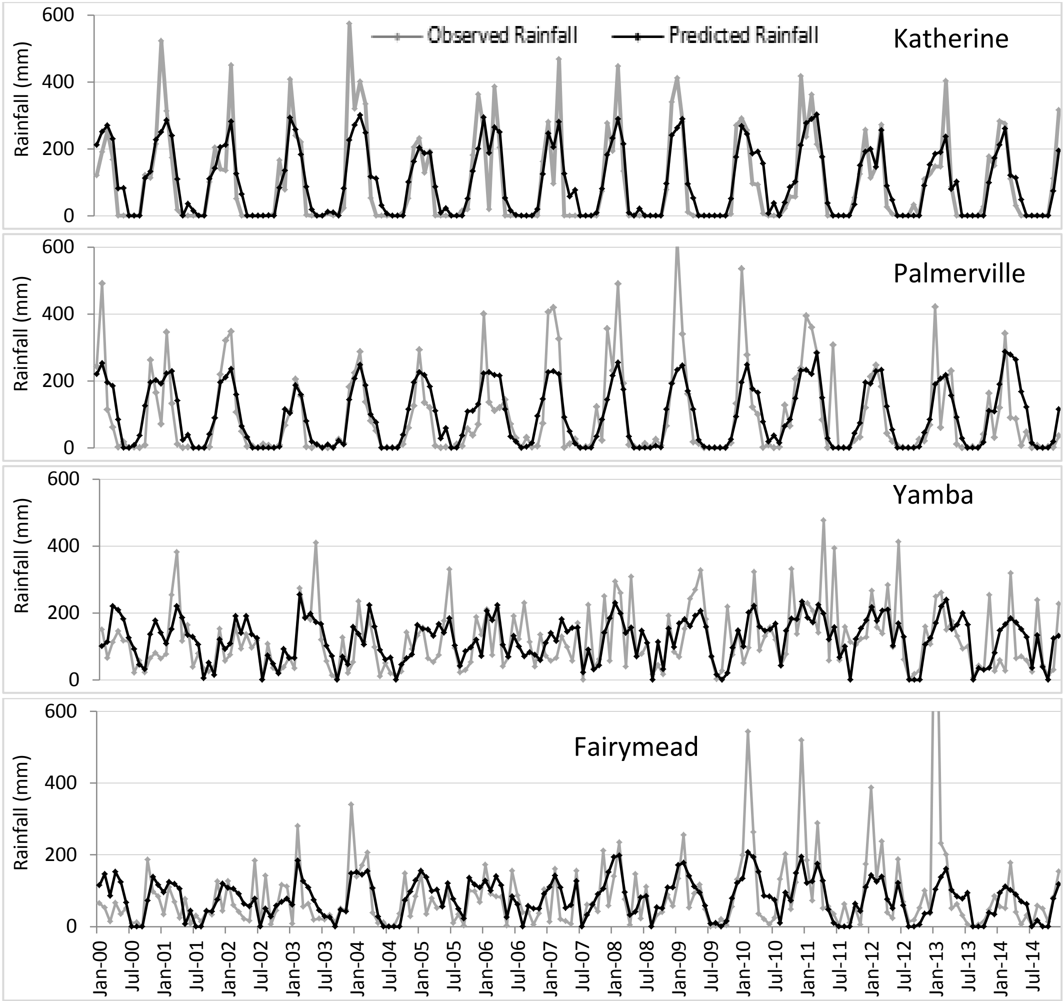

Figure 5.33: Observed rainfall vs. the MLR model predictions in tropical and subtropical classification zones.

Table 5.35 summarizes the prediction performance of the ANN(1) model with best and worst combinations of input variables locations in tropical and subtropical zones. The ANN(1) model provides best predictions with input variables TMax, TMin and Evap at Palmerville; with TMax, TMin and Rad at Yamba; with VP and Rad at Katherine and Fairymead. The model provides worst predictions with input variable Rad in all four locations of tropical and subtropical classification zones.

In tropical/subtropical zones, the performance measure RMSE for the ANN(1) model ranges from 64.85 to 98.00, MAE from 38.28 to 55.25, MASE from 0.48 to 0.60 and CE from 0.30 to 0.77. All four performance measures indicate that the ANN(1) model provides best predictions in Katherine while the performance measures RMSE,

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP, Rad	62.51	38.15	0.48	0.78
Palmerville	Tropical	TMax, TMin, Rad	74.43	43.38	0.55	0.67
Yamba	Subtropical	TMax, TMin, Rad	76.14	56.20	0.59	0.37
Fairymead	Subtropical	TMax, TMin, Evap, VP	98.59	45.26	0.53	0.29
Performance measures for worst input combinations						
Katherine	Tropical	Evap	122.64	95.58	1.19	0.17
Palmerville	Tropical	Evap, Rad	127.98	98.36	1.24	0.03
Yamba	Subtropical	Evap	95.16	77.33	0.81	0.01
Fairymead	Subtropical	Rad	117.17	70.85	0.83	0.00

Table 5.34: The ANN(0) model prediction performance with best and worst combinations of input variables in tropical and subtropical zones.

MASE and CE indicate the model provides worst predictions in Fairymead and MAE indicates in Yamba (see Figure 5.28). The graphical display of observed rainfall and the ANN(1) model predictions are given in Figure 5.35.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	VP, Rad	64.85	38.28	0.48	0.77
Palmerville	Tropical	TMax, TMin, Evap	72.34	42.64	0.54	0.69
Yamba	Subtropical	TMax, TMin, Rad	75.11	55.25	0.58	0.38
Fairymead	Subtropical	VP, Rad	98.00	51.61	0.60	0.30
Performance measures for worst input combinations						
Katherine	Tropical	Rad	134.75	107.62	1.34	-0.01
Palmerville	Tropical	Rad	129.94	98.79	1.25	0.00
Yamba	Subtropical	Rad	95.90	76.77	0.81	-0.01
Fairymead	Subtropical	Rad	120.39	72.76	0.85	-0.06

Table 5.35: The ANN(1) model prediction performance with best and worst combinations of input variables in tropical and subtropical zones.

Table 5.36 summarizes the prediction performance of the k-NN model with best and worst combinations of input variables in tropical and subtropical zones. The results presented show that the k-NN model with input variables TMax, TMin and VP provides best predictions at Katherine; with input variables TMax, TMin and Rad at Palmerville and Yamba; with VP and Rad at Fairymead. The model provides worst predictions with input variable Evap in three locations (Katherine, Yamba and Fairymead) and with Rad in the remaining location Palmerville.

In tropical and subtropical zones, the performance measure RMSE for the k-NN model ranges from 60.61 to 91.28, MAE from 36.09 to 60.53, MASE from 0.45 to 0.64 and CE from 0.33 to 0.80. All four performance measures indicate that the kNN model provides best predictions in Katherine while the performance measures RMSE indicates the model provides worst predictions at Fairymead and MAE, MASE and

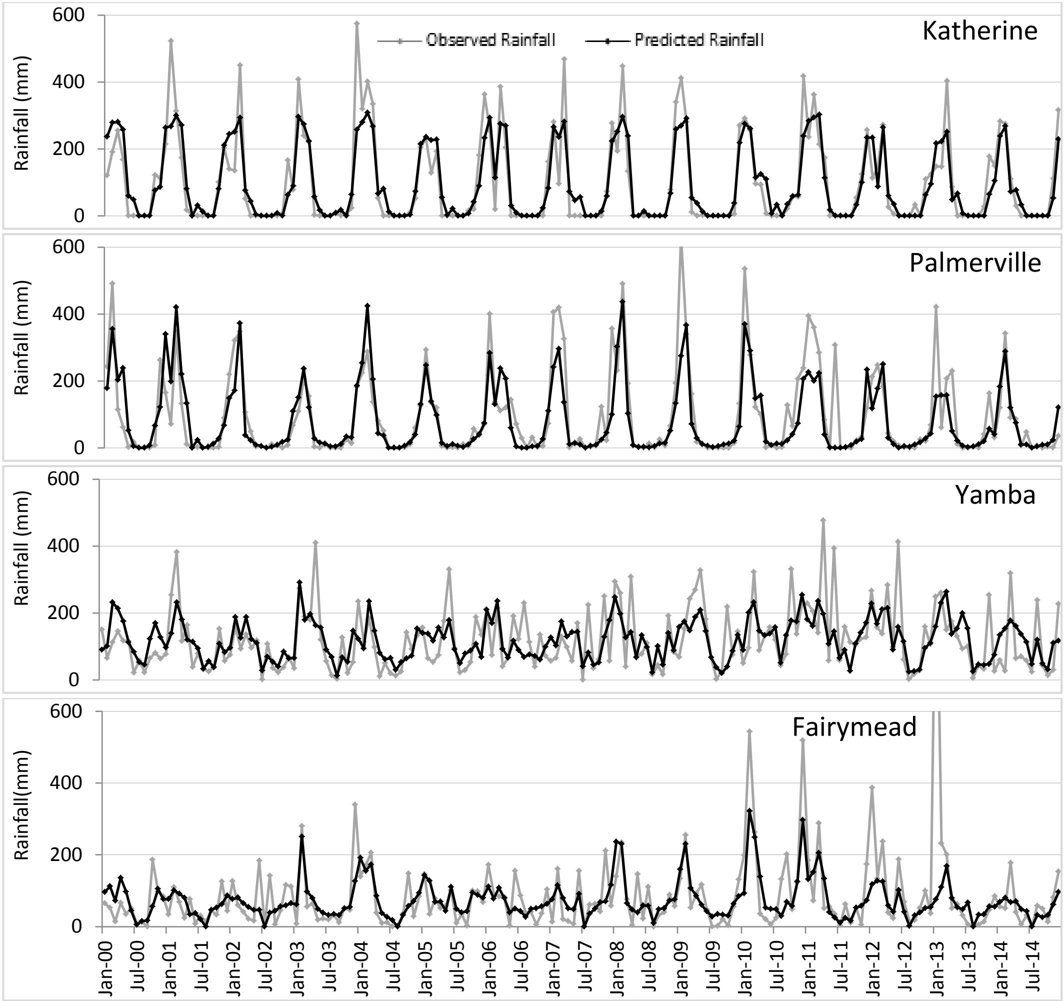

Figure 5.34: Observed rainfall vs. the ANN(0) model predictions in tropical and subtropical classification zones.

CE indicate at Yamba (see Figure 5.28). The graphical display of observed rainfall and the k-NN model predictions for all four selected locations in the tropical and subtropical classification zones are given in Figure 5.36.

Table 5.37, summarizes the performance of all eight models in predicting monthly rainfall with best combinations of input variables in tropical and subtropical zones. Best results among all models are highlighted in bold. Results presented in this table, show that at least one performance measure indicate that CLR(Opt) is best in three locations compare to all other models; the SVM(RBF) also in three locations; the k-NN in two locations; the SVM(Linear) in one location; the ANN(0) in one and ANN(1) model is also in one location. The CR(EM) and MLR models are not best

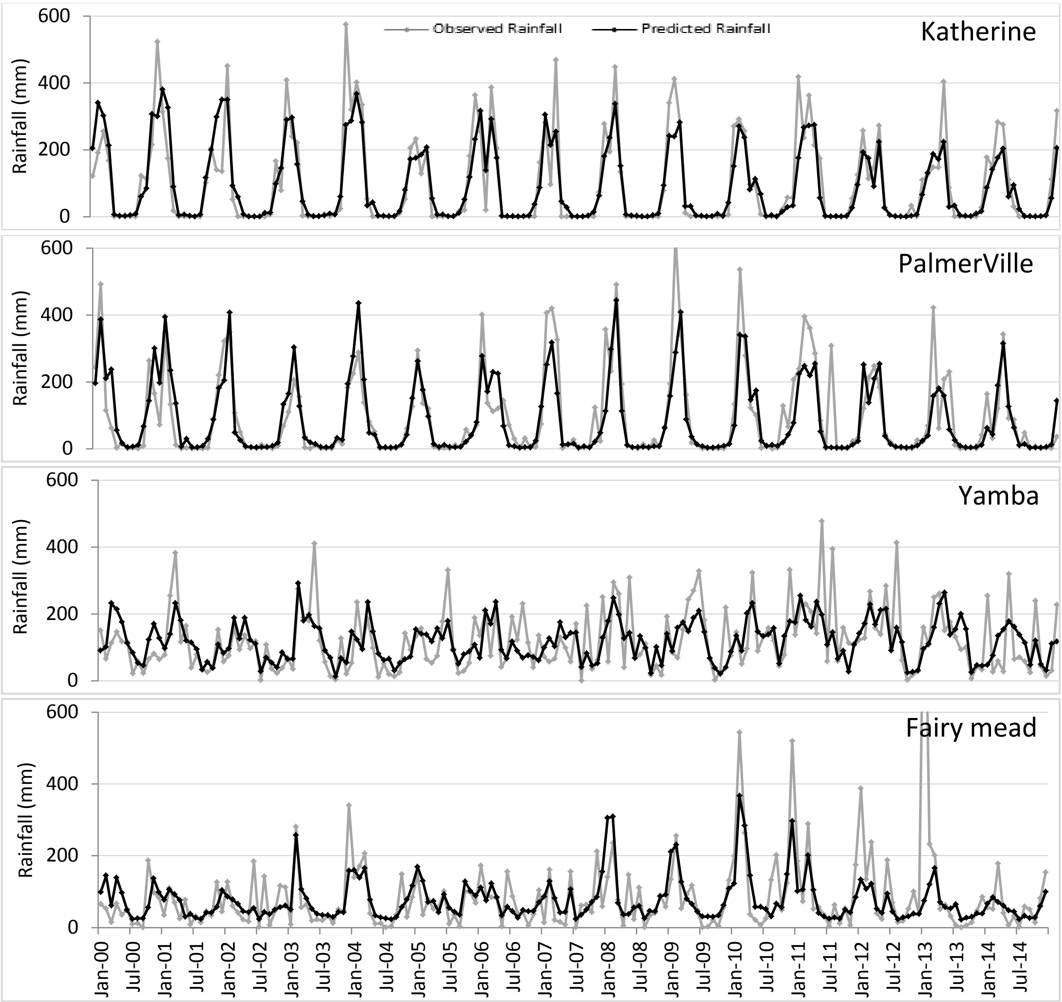

Figure 5.35: Observed rainfall vs. the ANN(1) model predictions in tropical and subtropical classification zones.

in any location.

According to our primary performance measure RMSE, the CLR(Opt) model outperformed other models at two locations (Katherine and Yamba) and k-NN model in the remaining two locations (Palmerville and Fairymead). The k-NN model performance is 0.5% higher at Palmerville and 3.8% higher at Fairymead comparing to the CLR(Opt) model. Results confirmed the superiority of CLR(Opt) model over CR(EM), SVM(Linear), SVM(RBF), ANN(0), ANN(1) and MLR models in tropical

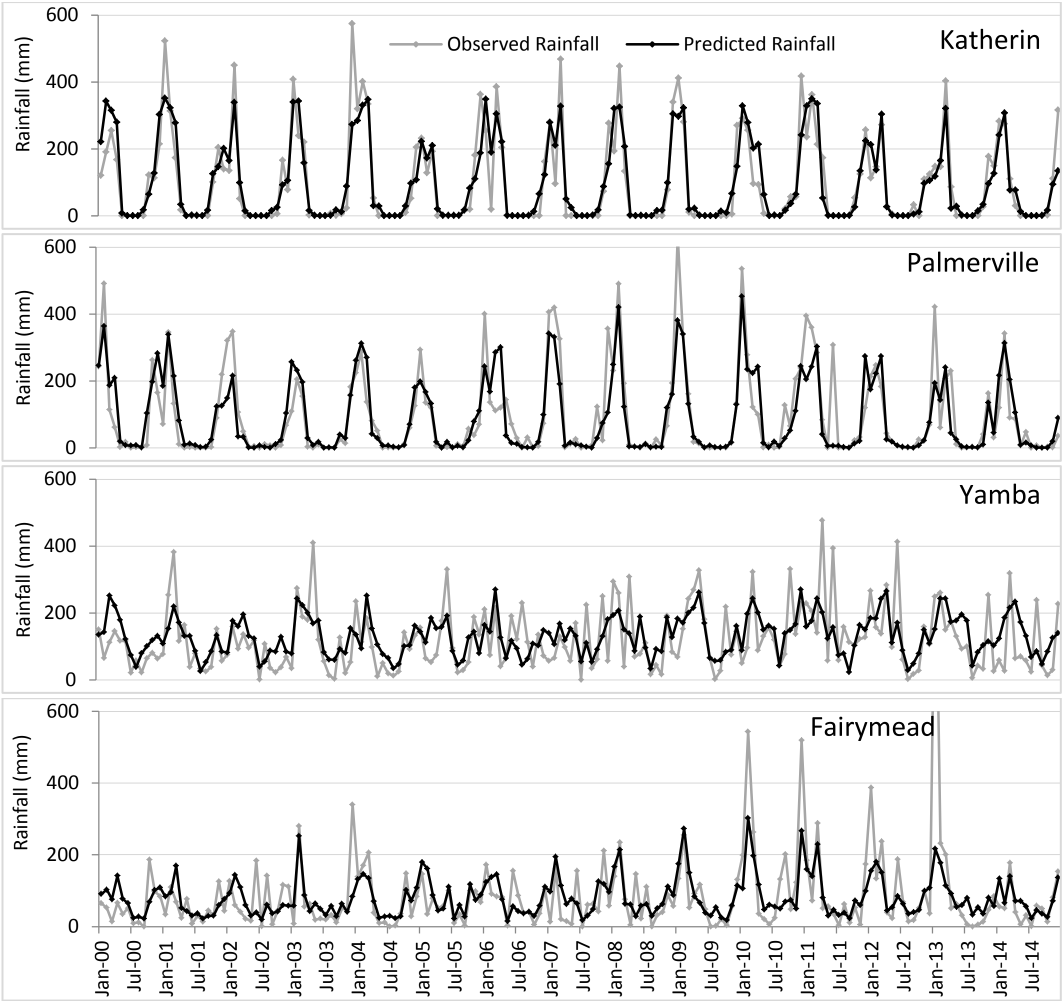

Figure 5.36: Observed rainfall vs. the k-NN model predictions in tropical and subtropical classification zones.

and subtropical classification zones.

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 5.29, 5.30, 5.31, 5.32, 5.33, 5.34, 5.35 and 5.36, show that all models follow the series patterns at all locations in tropical and subtropical classification zones.

In summary, based on the performance measure RMSE, the CLR(Opt) and k-NN models are the most suitable models in finding the pattern and trends of the observations compared to other models at all sites in tropical and subtropical classification zones.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP	60.51	36.09	0.45	0.80
Palmerville	Tropical	TMax, TMin, Rad	69.80	40.91	0.52	0.71
Yamba	Subtropical	TMax, TMin, Rad	78.15	60.53	0.64	0.33
Fairymead	Subtropical	VP, Rad	91.28	47.04	0.55	0.39
Performance measures for worst input combinations						
Katherine	Tropical	Evap	124.58	94.81	1.18	0.14
Palmerville	Tropical	Rad	141.74	106.16	1.34	-0.19
Yamba	Subtropical	Evap	106.50	83.86	0.88	-0.24
Fairymead	Subtropical	Evap	121.80	73.95	0.87	-0.08

Table 5.36: The k-NN model prediction performance with best and worst combinations of input variables in tropical and subtropical zones.

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	ANN (0)	ANN (1)	MLR	KNN
Katherine (Tropical)	RMSE	**58.17**	92.08	77.19	60.15	62.51	64.85	68.26	60.51
	MAE	34.42	49.67	44.88	**34.26**	38.15	38.28	42.32	36.09
	MASE	**0.43**	0.62	0.56	**0.43**	0.48	0.48	0.53	0.45
	CE	**0.81**	0.53	0.67	0.80	0.78	0.77	0.74	0.80
Palmerville (Tropical)	RMSE	70.15	98.18	90.77	71.67	74.43	72.34	83.00	**69.80**
	MAE	41.21	56.21	54.15	41.03	43.38	42.64	51.02	**40.91**
	MASE	**0.52**	0.71	0.68	**0.52**	0.55	0.54	0.64	**0.52**
	CE	**0.71**	0.43	0.51	0.69	0.67	0.69	0.59	**0.71**
Yamba (Subtropical)	RMSE	**75.04**	76.77	77.48	76.48	76.14	75.11	76.06	78.15
	MAE	54.81	55.59	**54.57**	54.65	56.20	55.25	56.64	60.53
	MASE	0.58	0.58	**0.57**	**0.57**	0.59	0.58	0.60	0.64
	CE	**0.38**	0.35	0.34	0.36	0.37	**0.38**	0.37	0.33
Fairymead (Subtropical)	RMSE	94.89	103.97	105.46	99.81	98.59	98.00	101.59	**91.28**
	MAE	47.04	49.47	48.85	47.64	**45.26**	51.61	49.71	47.04
	MASE	0.55	0.58	0.57	0.56	**0.53**	0.60	0.58	0.55
	CE	0.34	0.21	0.19	0.27	0.29	0.30	0.25	**0.39**

Table 5.37: Prediction performance of models with best combination of input variables in tropical and subtropical zones.

5.5 Summary of chapter

This chapter reports on the results of the proposed CLR(Opt) and seven other models for monthly rainfall predictions over fifteen years. Considering classification zones, twenty-four geographically diverse weather stations were used: seven from temperate zone, eight from grassland, five from the desert, two from tropical and two from subtropical zones. The purpose of choosing these locations was to investigate the performance of the prediction models in zones which experience different climate and hydrological regimes. All the selected prediction models were developed for each weather station using training sets and evaluated by using test sets. The prediction performance of models was evaluated by comparing observed and predicted rainfall using performance measures RMSE, MAE, MASE and CE. The computational results of the proposed model were compared with those obtained from CR(EM), SVM(Linear), SVM(RBF), MLR, ANN(0), ANN(1) and k-NN models.

The results reported in this chapter show that in the temperate climate zone, the CLR(Opt) and SVM(RBF) models are the most capable of finding the patterns and trends of the observations. In the grassland climate zone, the ANN(1) model is the most suitable for monthly rainfall predictions compared to other models. In the desert climate zone, the CLR(Opt) is the best, and in tropical and subtropical zones, the CLR(Opt) and k-NN models are better at predicting monthly rainfall when compared with other models. Results demonstrate that the CLR(Opt) model is an efficient method for monthly rainfall predictions and is superior to CR(EM), SVM(Linear), MLR, ANN(0) and k-NN models in most locations used in this study.

The prediction performance of all models varied considerably with changes of geographic regions. Predictions have the lowest deviation from the actual observations in the desert classification zone. The temperate classification zone was found to be second, with the lowest deviation between predictions and actual observations, and the grassland zone was ranked third. In the tropical and subtropical zones, predictions have the highest deviation from the actual rainfall observations.

The results also confirmed that no single meteorological parameter as an input variable provides the best predictions. Adding more input variables improved the performance of models in most locations used in this study. The best combinations of input variables in most temperate zone locations are: C12, C13 and C15; in grassland zone: C6, C11, C13 and C15; in desert zone: C1, C6, C7, C13 and C15 and in tropical and subtropical zones are: C6, C7, C10, C12 and C15 (see Table 5.1 for combinations detail). The results presented in this chapter suggest that these five meteorological parameters are suitable predictors for monthly rainfall predictions in Australia.

Chapter 6

Weekly Rainfall Predictions

In this chapter, first we present the weekly rainfall prediction results of the CLR(Opt), CR(EM), MLR, SVMreg(RBF), SVMreg(Linear), ANN(0), ANN(1) and K-NN models within each classification zone then we assess the CLR(Opt) model performance comparing to other models. For weekly rainfall predictions, all models were developed using four combinations of input variables: (1) minimum temperature, maximum temperature and vapour pressure (2) minimum temperature, maximum temperature and solar radiation, (3) minimum temperature, maximum temperature, vapour pressure and solar radiation and (4) minimum temperature, maximum temperature, evaporation, vapour pressure and solar radiation. We used these combinations of input variables as these provided best monthly rainfall predictions in most locations (See Chapter 5). The models were trained using data from Jan 1970 to Feb 2005 and tested using data from Feb 2005 to Feb 2015 in all 24 locations. In all cases, negative predicted values were adjusted to zero rainfall before performance measures calculated.

The prediction performance of models with each combination was evaluated by comparing observed and predicted rainfall. The performance measure RMSE used as a primary measure to determine the best combination of input variables for each model at each location. Then we compare the results of the proposed method with those obtained using the CR(EM), MLR, SVMreg(RBF), SVMreg(Linear), ANN(0), ANN(1) and K-NN methods. We also compare the predictions zone-wise to assess the influence of geographic regions on the performance of models.

The structure of the chapter is as follows. We present the weekly rainfall prediction results for temperate classification zone in Section 6.1; for grassland in Section 6.2; for desert in Section 6.3; for tropical and subtropical in Section 6.4. We also provide the comparative assessment results of prediction models within each classification zone at the end of each section. Section 6.5 concludes the chapter.

6.1 Weekly rainfall predictions in temperate zone

In this section, first we present the weekly rainfall prediction results of each model in predicting weekly rainfall with best and worst combinations of input variables in temperate classification zone. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models using computational results and time series plots.

Table 6.1, summarizes the performance of the CLR(Opt) model in predicting weekly rainfall with best and worst combinations of input variables in temperate classification zone. According to these results, the model provides best predictions with input variables TMax, TMin and VP in Orbost; with TMax, TMin and Rad in Port Elliot and Peppermint Grove; with TMax, TMin, VP and Rad in Dookie; with full set of input variables in the remaining three locations Mossvale, Koppio and Cape Otway. The CLR(Opt) model provides worst predictions in four locations (Mossvale, Dookie, Orbost and Cape Otway) with input variable TMax, TMin and Rad; in two locations (Koppio and Peppermint Grove) with TMax, TMin and VP; and in one location (Port Elliot) with TMax, TMin, VP and Rad.

The performance measure RMSE for the CLR(Opt) in predicting weekly rainfall ranges from 10.07 to 20.04; MAE from 6.51 to 12.42; MASE from 0.54 to 0.68; and CE from 0.18 to 0.42. The performance measures RMSE, and MAE indicate that the model provides best predictions with lowest prediction error in the location Port Elliot; MASE indicates in the location Dookie; CE indicates in the location Mossvale while the performance measures RMSE, and MAE indicate worst predictions in Mossvale; MASE and CE indicate in Peppermint Grove (see Figure 6.1). The graphical display of observed rainfall and CLR(Opt) model predictions for ten years from Feb 2005 to Feb 2015 is given in Figure 6.2.

Table 6.2, summarizes the performance of the CR(EM) model in predicting weekly rainfall with best and worst combinations of input variables using test data over the period Feb 2005 to Feb 2015 in temperate classification zone. Results show that the CR(EM) model with input variables TMax, TMin, VP and Rad provides best predictions in three out of seven locations (Mossvale, Koppio and Port Elliot) and with full set of input variables provides best predictions in the remaining four locations (Dookie, Orbost, Cape Otway, and Peppermint Grove). The model provides worst predictions with input variable TMax, TMin and Rad in three locations (Mossvale, Dookie and Orbost); with TMax, TMin and VP in two locations (Cape Otway and Peppermint Grove); with a full set of five input variables in the remaining two locations (Koppio and Port Elliot).

The performance measure RMSE for the CR(EM) in predicting weekly rainfall ranges from 10.88 to 21.02, MAE from 7.12 to 11.89, MASE from 0.54 to 0.70 and

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	20.04	12.42	0.63	0.42
Koppio	TMax, TMin, Evap, VP, Rad	10.55	6.59	0.58	0.39
Port Elliot	TMax, TMin, Rad	10.07	6.51	0.61	0.31
Dookie	TMax, TMin, VP, Rad	13.49	7.43	0.54	0.39
Orbost	TMax, TMin, VP	17.13	10.86	0.61	0.37
Cape Otway	TMax, TMin,Evap, VP, Rad	14.34	9.91	0.65	0.25
Peppermint Grove	TMax, TMin, Rad	16.60	10.01	0.68	0.18
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	21.02	13.74	0.70	0.36
Koppio	TMax, TMin, VP	10.86	6.67	0.59	0.35
Port Elliot	TMax, TMin, VP, Rad	10.24	6.80	0.64	0.29
Dookie	TMax, TMin, Rad	13.83	7.60	0.56	0.35
Orbost	TMax, TMin, Rad	18.40	12.29	0.69	0.28
Cape Otway	TMax, TMin, Rad	15.11	10.81	0.71	0.16
Peppermint Grove	TMax, TMin, VP	16.74	10.05	0.68	0.17

Table 6.1: The CLR(Opt) model prediction performance for weekly rainfall using best and worst combinations of input variables in temperate classification zone.

CE from 0.14 to 0.36. The performance measures RMSE indicates that the CR(EM) model provides best predictions with lowest prediction error in the location Port Elliot; MAE indicates in the location Koppio; MASE indicates in the location Dookie, and CE indicates in the location Mossvale while performance measures RMSE, and MAE indicate worst predictions at Mossvale; MASE indicates in Cape Otway; and CE in Peppermint Grove(see Figure 6.1). The graphical display of observed rainfall and CR(EM) model predictions for ten years from Feb 2005 to Feb 2015 is given in Figure 6.3.

Table 6.3, summarizes the prediction performance of the SVM(Linear) model with best and worst combinations of input variables using test data sets over the period Feb 2005 to Feb 2015 in temperate classification zone. The results presented in this table show that the model with input variables TMax, TMin, and VP provides best predictions in the location Mossvale; with TMax, TMin and Rad in two locations (Koppio and Peppermint Grove); with TMax, TMin, VP and Rad in the location Port Elliot and with full set of input variables provides best predictions in the remaining three locations (Dookie, Orbost and Cape Otway). The SVM(Linear) model provides the worst predictions with input variable TMax, TMin and Rad in three locations (Mossvale, Dookie and Orbost); with TMax, TMin and VP also in three locations (Port Elliot, Cape Otway and Peppermint Grove) and with full set of five input variables in the remaining location Koppio.

The performance measure RMSE for the SVM(Linear) model in predicting weekly rainfall ranges from 10.92 to 23.84, MAE from 6.64 to 12.64, MASE from 0.53 to 0.65

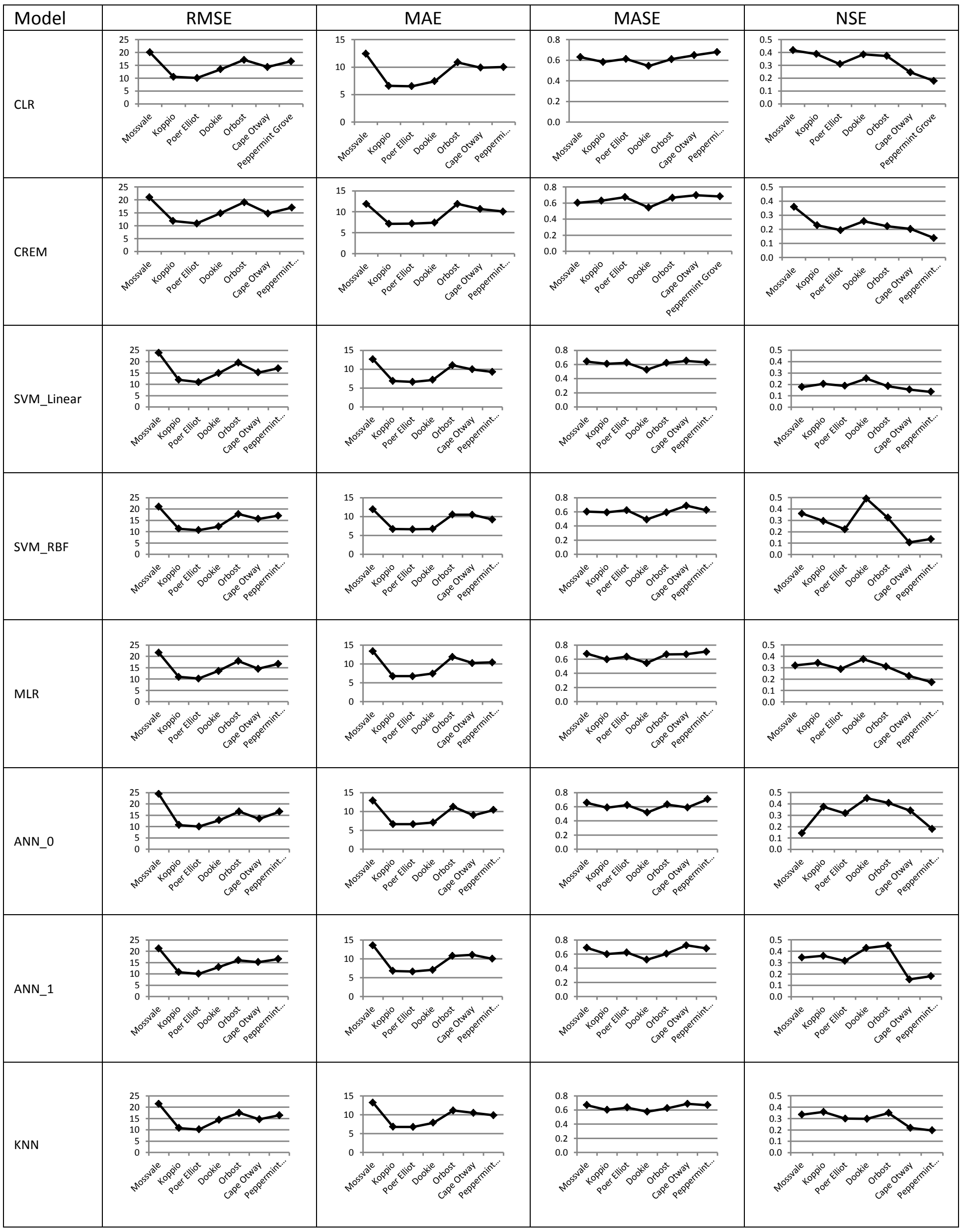

Figure 6.1: Illustration of models performance in predicting weekly rainfall in temperate zone.

and CE from 0.13 to 0.25. The performance measures RMSE, and MAE indicate the model provides best predictions in the location Port Elliot; MAE and CE indicate in the location Dookie while RMSE and MAE indicate worst predictions in Mossvale; MASE in Cape Otway and CE indicate in Peppermint Grove (see Figure 6.1). The graphical display of observed rainfall and SVM(Linear) model predictions for ten years from Feb 2005 to Feb 2015 is given in Figure 6.4.

Table 6.4, summarizes the prediction performance of the SVM(RBF) model with

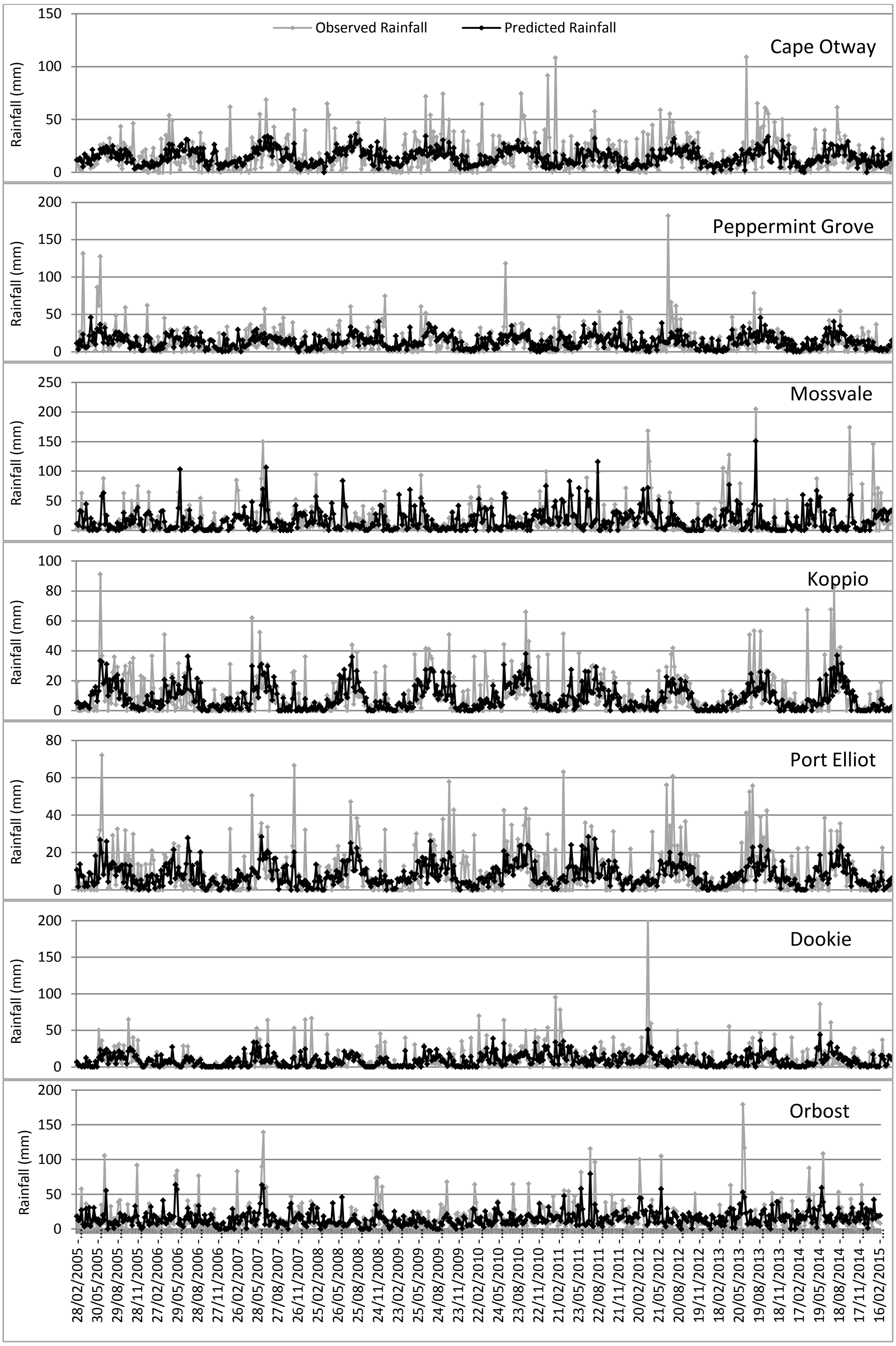

Figure 6.2: Observed rainfall vs. CLR(Opt) model weekly predictions in temperate zone.

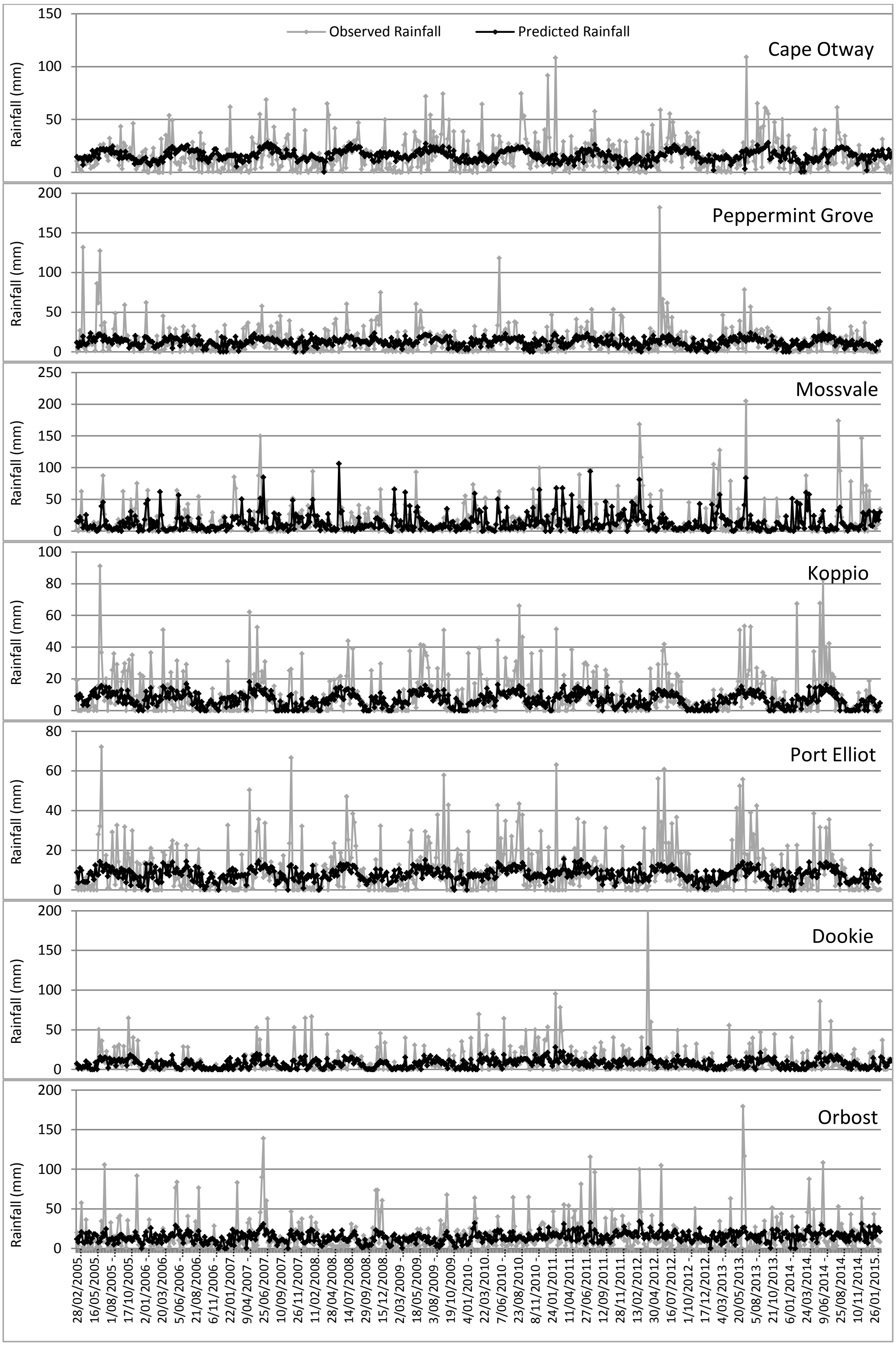

Figure 6.3: Observed rainfall vs. CR(EM) model weekly predictions in temperate zone.

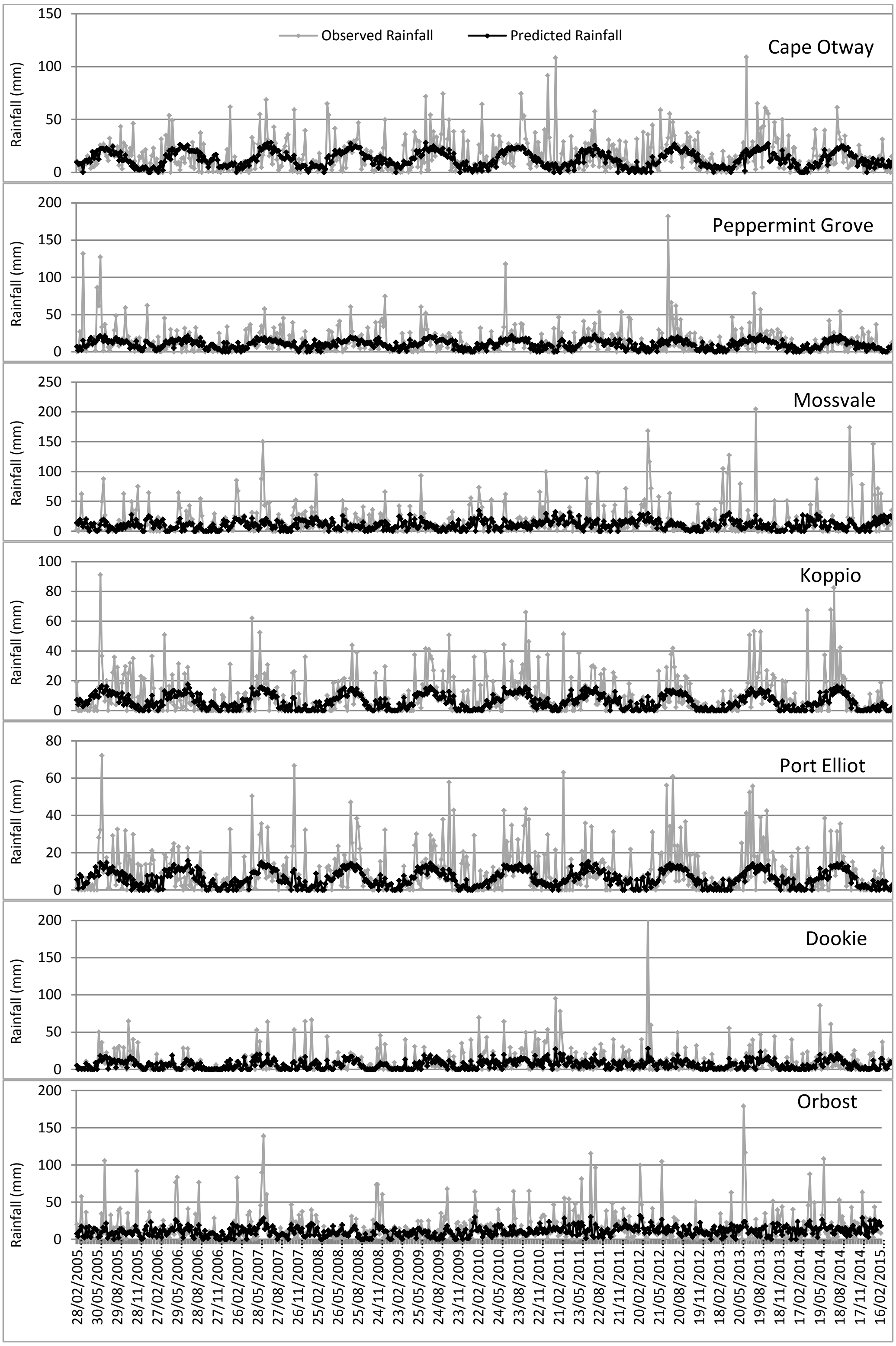

Figure 6.4: Observed rainfall vs. the SVM(Linear) model weekly predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, VP, Rad	21.01	11.89	0.60	0.36
Koppio	TMax, TMin, VP, Rad	11.84	7.12	0.63	0.23
Port Elliot	TMax, TMin, VP, Rad	10.88	7.18	0.67	0.19
Dookie	TMax, TMin, Evap, VP, Rad	14.83	7.41	0.54	0.26
Orbost	TMax, TMin, Evap, VP, Rad	19.09	11.88	0.67	0.22
Cape Otway	TMax, TMin, Evap, VP, Rad	14.74	10.66	0.70	0.20
Peppermint Grove	TMax, TMin, Evap, VP, Rad	17.01	10.06	0.68	0.14
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	21.71	11.99	0.61	0.32
Koppio	TMax, TMin, Evap, VP, Rad	11.93	7.12	0.63	0.22
Port Elliot	TMax, TMin, Evap, VP, Rad	10.96	6.96	0.65	0.18
Dookie	TMax, TMin, Rad	15.00	7.87	0.58	0.24
Orbost	TMax, TMin, Rad	19.54	12.30	0.69	0.18
Cape Otway	TMax, TMin, Vp	15.18	11.16	0.73	0.15
Peppermint Grove	TMax, TMin, VP	17.13	10.15	0.69	0.13

Table 6.2: The CR(EM) model prediction performance for weekly rainfall using best and worst combinations of input variables in temperate zone.

best and worst combinations of input variables using test data sets over the period Feb 2005 to Feb 2015 in temperate classification zone. The results presented in this table show that the model with input variables TMax, TMin, and VP provides best predictions in three out of seven locations (Koppio, Orbost and Peppermint Grove); with TMax, TMin and Rad in three locations (Port Elliot, Dookie and Cape Otway) and with TMax, TMin, VP and Rad in the remaining location Mossvale. The model provides the worst predictions with input variable TMax, TMin and Rad in three locations (Mossvale, Dookie and Orbost); with TMax, TMin and VP in three locations (Port Elliot, Cape Otway and Peppermint Grove) and with a full set of five input variables in the remaining location Koppio.

The performance measure RMSE for the SVM(RBF) model in predicting weekly rainfall ranges from 10.69 to 21.01, MAE from 6.62 to 11.89, MASE from 0.49 to 0.69 and CE from 0.11 to 0.49. The performance measures RMSE, and MAE indicate the model provides best predictions in the location Port Elliot; MAE and CE indicate in the location Dookie while RMSE and MAE indicate the model provide worst predictions in Port Elliot and MASE and CE indicate in Cape Otway (see Figure 6.1). The graphical display of observed rainfall and SVM(RBF) model predictions for ten years over the period Feb 2005 to Feb 2015 is given in Figure 6.5.

Table 6.5, summarizes the prediction performance of the MLR model in predicting weekly rainfall with best and worst combinations of input variables in temperate classification zone. According to these results the model with input variables TMax, TMin, and VP provides best predictions in the location Koppio; with TMax, TMin

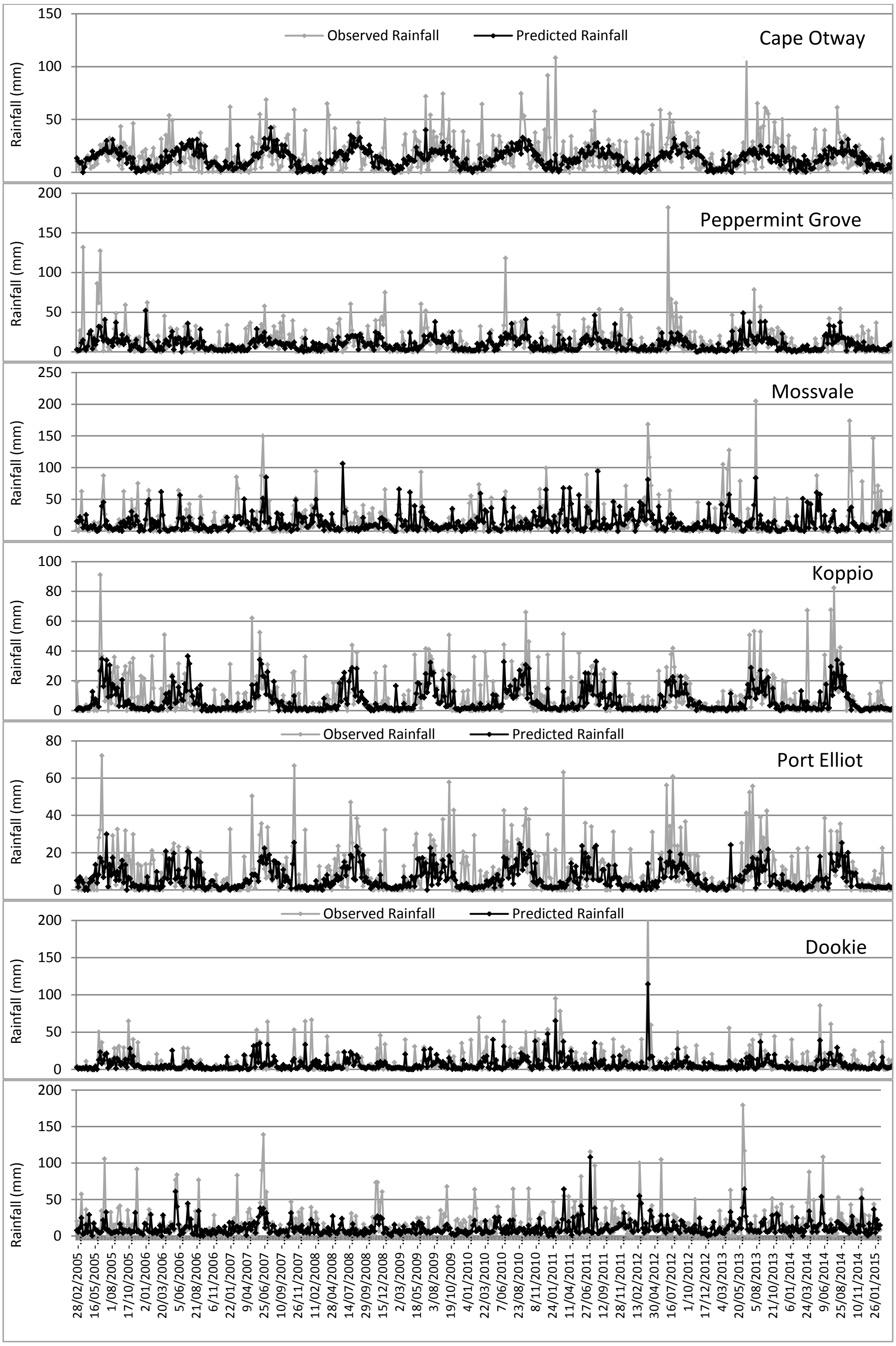

Figure 6.5: Observed rainfall vs. SVM(RBF) model weekly predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, VP	23.84	12.64	0.64	0.18
Koppio	TMax, TMin, Rad	12.02	6.89	0.61	0.21
Port Elliot	TMax, TMin, VP, Rad	10.92	6.64	0.62	0.19
Dookie	TMax, TMin, Evap, VP, Rad	14.89	7.16	0.53	0.25
Orbost	TMax, TMin, Evap, VP, Rad	19.53	11.06	0.62	0.18
Cape Otway	TMax, TMin, Evap, VP, Rad	15.19	9.94	0.65	0.15
Peppermint Grove	TMax, TMin, Rad	17.05	9.27	0.63	0.13
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	24.19	12.67	0.64	0.15
Koppio	TMax, TMin, Evap, VP, Rad	12.23	6.86	0.61	0.18
Port Elliot	TMax, TMin, VP	11.10	6.75	0.63	0.16
Dookie	TMax, TMin, Rad	15.13	7.35	0.54	0.23
Orbost	TMax, TMin, Rad	20.17	11.45	0.64	0.13
Cape Otway	TMax, TMin, VP	15.52	10.41	0.68	0.12
Peppermint Grove	TMax, TMin, VP	17.18	9.37	0.64	0.12

Table 6.3: The SVM_Linear model prediction performance for weekly rainfall using best and worst combinations of input variables in temperate zone.

and Rad in Port Elliot and Peppermint Grove; with a full set of input variables in Mossvale, Dookie, Orbost and Cape Otway. The MLR model provides worst predictions with input variable TMax, TMin and Rad in Mossvale, Dookie, Orbost and Cape Otway; with TMax, TMin and VP in Port Elliot and Peppermint Grove and with a full set of five input variables in the remaining location Koppio.

The performance measure RMSE for the MLR model in predicting weekly rainfall ranges from 10.22 to 21.65, MAE from 6.77 to 13.35, MASE from 0.55 to 0.71 and CE from 0.17 to 0.38. The performance measures RMSE, and MAE indicate the MLR model provides best predictions in Port Elliot and worst in Mossvale while MAE and CE indicate best in Dookie and worse in Peppermint Grove (see Figure 6.1). The graphical display of observed rainfall and MLR model predictions for ten years is given in Figure 6.6.

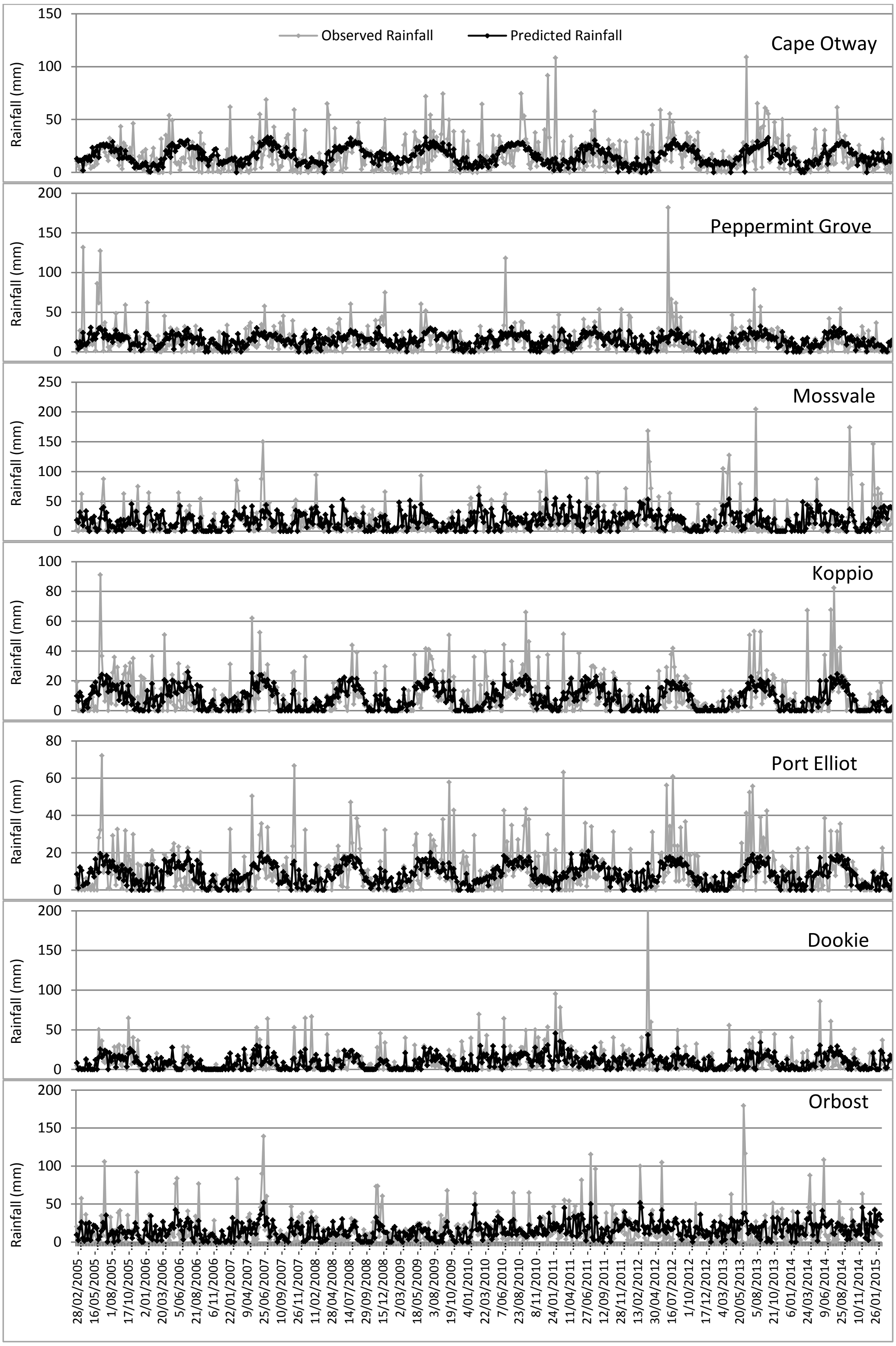

Figure 6.6: Observed rainfall vs. MLR model weekly predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, VP, Rad	21.01	11.89	0.60	0.36
Koppio	TMax, TMin, VP	11.32	6.70	0.59	0.29
Port Elliot	TMax, TMin, Rad	10.69	6.62	0.62	0.22
Dookie	TMax, TMin, Rad	12.27	6.71	0.49	0.49
Orbost	TMax, TMin, VP	17.80	10.51	0.59	0.32
Cape Otway	TMax, TMin, Rad	15.61	10.47	0.69	0.11
Peppermint Grove	TMax, TMin, VP	17.03	9.20	0.63	0.14
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	21.71	11.99	0.61	0.32
Koppio	TMax, TMin, Evap, VP, Rad	11.85	7.59	0.67	0.23
Port Elliot	TMax, TMin, Evap, VP, Rad	10.95	7.03	0.66	0.18
Dookie	TMax, TMin, Evap, VP, Rad	13.38	7.52	0.55	0.40
Orbost	TMax, TMin, Rad	19.31	11.17	0.63	0.20
Cape Otway	TMax, TMin, VP, Rad	16.82	11.56	0.76	-0.04
Peppermint Grove	TMax, TMin, Evap, VP, Rad	18.61	10.31	0.70	-0.03

Table 6.4: The SVM(RBF) model prediction performance for weekly rainfall using best and worst combinations of input variables in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	21.65	13.35	0.68	0.32
Koppio	TMax, TMin, VP	10.94	6.76	0.60	0.34
Port Elliot	TMax, TMin, Rad	10.22	6.77	0.64	0.29
Dookie	TMax, TMin, Evap, VP, Rad	13.60	7.44	0.55	0.38
Orbost	TMax, TMin, Evap, VP, Rad	17.95	11.88	0.67	0.31
Cape Otway	TMax, TMin, Evap, VP, Rad	14.51	10.22	0.67	0.23
Peppermint Grove	TMax, TMin, Rad	16.65	10.41	0.71	0.17
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	22.43	14.60	0.74	0.27
Koppio	TMax, TMin, Evap, VP, Rad	11.10	6.75	0.60	0.32
Port Elliot	TMax, TMin, VP	10.36	6.92	0.65	0.27
Dookie	TMax, TMin, Rad	13.84	7.62	0.56	0.35
Orbost	TMax, TMin, Rad	19.15	12.77	0.72	0.22
Cape Otway	TMax, TMin, Rad	15.14	10.89	0.71	0.16
Peppermint Grove	TMax, TMin, VP	16.79	10.47	0.71	0.16

Table 6.5: The MLR model prediction performance for weekly rainfall with best and worst combinations of input variables in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, VP, Rad	24.34	12.90	0.65	0.14
Koppio	TMax, TMin, Rad	10.66	6.63	0.59	0.37
Port Elliot	TMax, TMin, VP	10.01	6.63	0.62	0.32
Dookie	TMax, TMin, Evap, VP, Rad	12.77	7.06	0.52	0.45
Orbost	TMax, TMin, VP	16.64	11.23	0.63	0.41
Cape Otway	TMax, TMin, Evap, VP, Rad	13.41	8.99	0.59	0.34
Peppermint Grove	TMax, TMin, VP	16.59	10.39	0.71	0.18
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	24.91	12.85	0.65	0.10
Koppio	TMax, TMin, VP	11.09	7.08	0.63	0.32
Port Elliot	TMax, TMin, Rad	10.22	6.77	0.64	0.29
Dookie	TMax, TMin, VP, Rad	13.40	7.16	0.53	0.39
Orbost	TMax, TMin, Rad	19.25	12.75	0.72	0.21
Cape Otway	TMax, TMin, VP	15.17	10.99	0.72	0.16
Peppermint Grove	TMax, TMin, Evap, VP, Rad	16.69	10.31	0.70	0.17

Table 6.6: The ANN(0) model prediction performance for weekly rainfall with best and worst combinations of input variables in temperate zone.

Table 6.6, summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables in temperate classification zone. According to these results, the model with input variables TMax, TMin, and VP provides best predictions in Port Elliot, Orbost and Peppermint Grove; with TMax, TMin and Rad in the location Koppio; with TMax, TMin, VP and Rad in the location Mossvale and with full set of five input variables in the remaining locations Dookie and Cape Otway. The ANN(0) model provides the worst predictions with input variable TMax, TMin and Rad in Mossvale, Port Elliot and Orbost; with TMax, TMin and VP in Koppio and Cape Otway; with TMax, TMin, VP and Rad in the location Dookie and with full set of five input variables in the remaining location Peppermint Grove.

The performance measure RMSE for the ANN(0) model in predicting weekly rainfall ranges from 10.01 to 24.34, MAE from 6.63 to 12.90, MASE from 0.52 to 0.71 and CE from 0.14 to 0.45. The performance measures RMSE, and MAE indicate the ANN(0) model provide best predictions in the locations Port Elliot; MASE and CE indicate in Dookie while the performance measures RMSE, MAE and CE indicate worst predictions in Mossvale and MASE indicates in Peppermint Grove (see Figure 6.1). The graphical display of observed rainfall and ANN(0) model predictions over the period of ten years is given in Figure 6.7.

Table 6.7, summarizes the prediction performance of the ANN(1) model with best and worst combinations of input variables in temperate classification zone. Results show that the model with input variables TMax, TMin, and Rad provides best predictions in four locations (Koppio, Port Elliot, Cape Otway and Peppermint Grove);

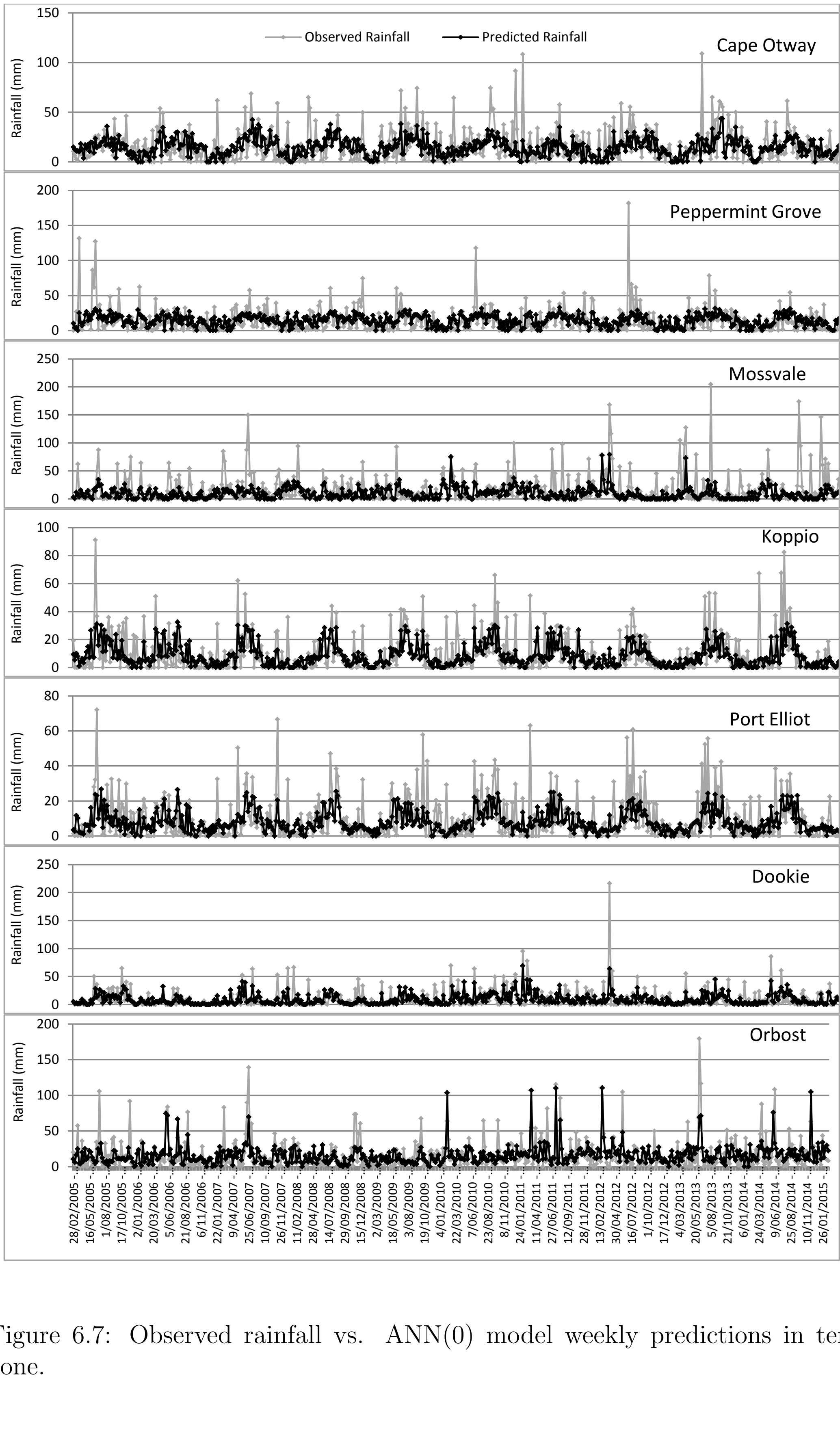

Figure 6.7: Observed rainfall vs. ANN(0) model weekly predictions in temperate zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Vp, Rad	21.26	13.58	0.69	0.34
Koppio	TMax, TMin, Rad	10.78	6.78	0.60	0.36
Port Elliot	TMax, TMin, Rad	10.03	6.63	0.62	0.31
Dookie	TMax, TMin, VP, Rad	13.01	7.09	0.52	0.43
Orbost	TMax, TMin, Evap, VP, Rad	16.04	10.75	0.60	0.45
Cape Otway	TMax, TMin, Rad	15.20	11.05	0.72	0.15
Peppermint Grove	TMax, TMin, Rad	16.58	10.01	0.68	0.18
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	23.06	13.74	0.70	0.23
Koppio	TMax, TMin, VP, Rad	13.48	9.81	0.87	0.00
Port Elliot	TMax, TMin, VP	10.23	6.88	0.65	0.29
Dookie	TMax, TMin, Evap, VP, Rad	13.40	7.19	0.53	0.39
Orbost	TMax, TMin, Rad	21.66	14.49	0.81	0.00
Cape Otway	TMax, TMin, VP, Rad	16.64	13.00	0.85	-0.02
Peppermint Grove	TMax, TMin, Vp, Rad	18.32	12.15	0.83	0.00

Table 6.7: The ANN(1) model performance in predicting weekly rainfall with best and worst combinations of input variables in temperate zone.

with TMax, TMin, VP and Rad in two locations (Mossvale and Dookie) and with full set of five input variables in the remaining location Orbost. The model provides worst predictions with input variable TMax, TMin and Rad in two locations (Mossvale and Orbost); with TMax, TMin and VP in the location Port Elliot; with TMax, TMin, VP and Rad in three locations (Koppio, Cape Otway and Peppermint Grove) and with full set of five input variables in the remaining location Dookie.

The performance measure RMSE for the ANN(1) model in predicting weekly rainfall ranges from 10.03 to 21.26, MAE from 6.63 to 13.58, MASE from 0.52 to 0.72 and CE from 0.15 to 0.45. The performance measures RMSE and MAE indicate the ANN(1) model provides best predictions in the location Port Elliot; MASE indicates in the location Dookie and CE indicates in the location Orbost while RMSE and MAE indicate worst predictions in Mossvale and MASE and CE indicate in Cape Otway (see Figure 6.1). The graphical display of observed rainfall and ANN(1) model predictions over the test period is given in Figure 6.8.

Table 6.8, summarizes the prediction performance of the k-NN model with best and worst combinations of input variables in temperate classification zone. The results presented show that the model with input variables TMax, TMin, and Rad provides best predictions in the location Port Elliot; with TMax, TMin, VP in the location Orbost; with TMax, TMin, VP and Rad in the location Peppermint Grove and with a full set of five input variables in the remaining four locations (Mossvale, Koppio, Dookie and Cape Otway). The k-NN model provides the worst predictions with input variable TMax, TMin and Rad in four locations (Mossvale, Koppio, Orbost and

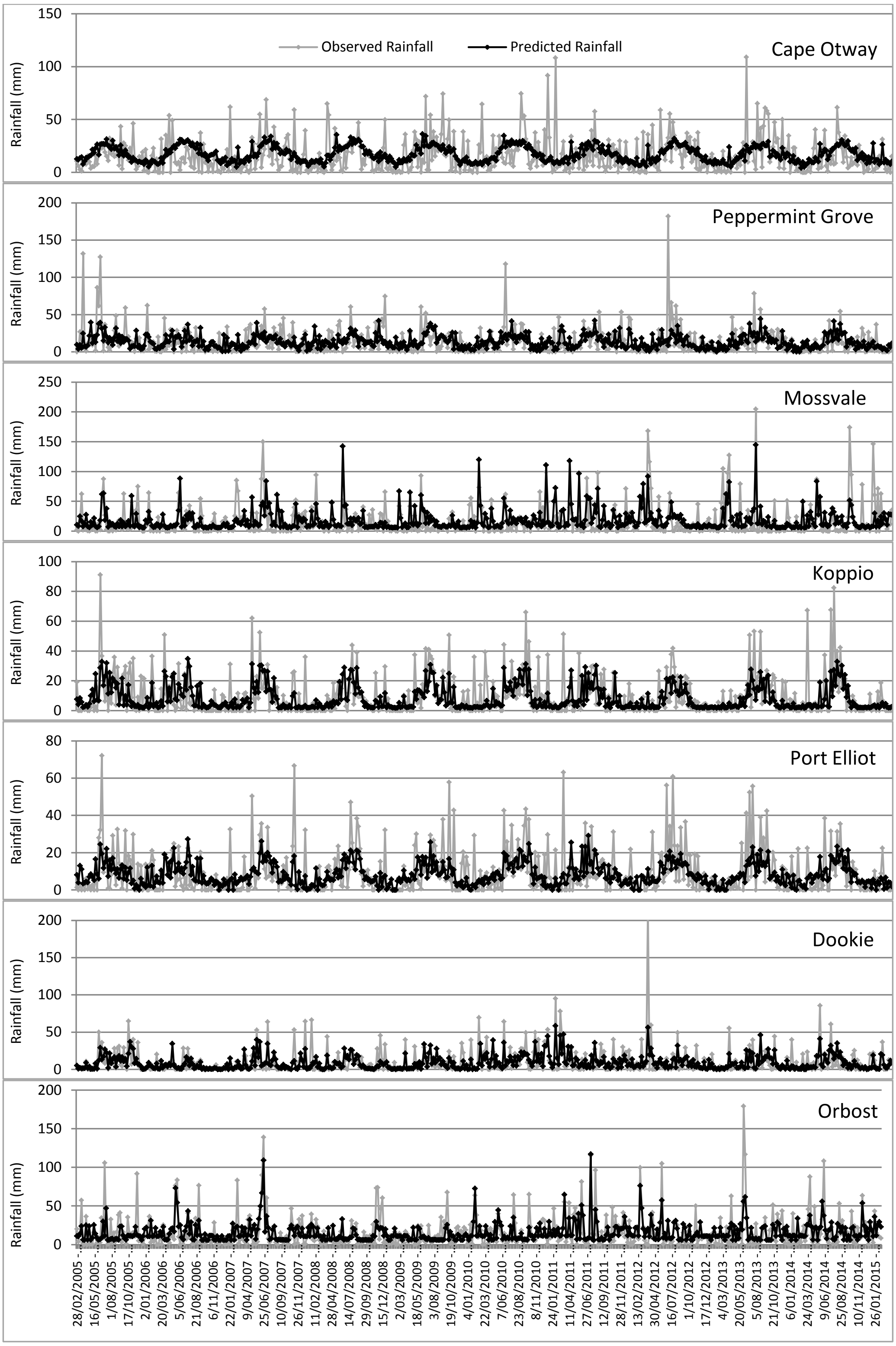

Figure 6.8: Observed rainfall vs. ANN(1) model weekly predictions in temperate zone.

Peppermint Grove); with TMax, TMin and VP in two locations (Dookie and Cape Otway) and with a full set of five input variables in the remaining location Port Elliot.

The performance measure RMSE for the k-NN model in predicting weekly rainfall ranges from 10.14 to 21.45, MAE from 6.76 to 13.19, MASE from 0.58 to 0.69 and CE from 0.20 to 0.36. The performance measures RMSE and MAE indicate the k-NN model provides best predictions in the location Port Elliot; MASE indicates in the location Dookie, and CE indicates in the location Koppio while the performance measures RMSE, and MAE indicate worst predictions in Mossvale; MASE indicates in Cape Otway and CE in Peppermint Grove (see Figure 6.1). The graphical display of observed rainfall and k-NN model predictions over the test period is given in Figure 6.9.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Mossvale	TMax, TMin, Evap, VP, Rad	21.45	13.19	0.67	0.33
Koppio	TMax, TMin, Evap, VP, Rad	10.81	6.78	0.60	0.36
Port Elliot	TMax, TMin, Rad	10.14	6.76	0.64	0.30
Dookie	TMax, TMin, Evap, VP, Rad	14.43	7.86	0.58	0.30
Orbost	TMax, TMin, VP	17.46	11.11	0.62	0.35
Cape Otway	TMax, TMin, Evap, VP, Rad	14.61	10.48	0.69	0.22
Peppermint Grove	TMax, TMin, VP, rad	16.44	9.85	0.67	0.20
Performance measures for worst input combinations					
Mossvale	TMax, TMin, Rad	22.00	13.87	0.70	0.30
Koppio	TMax, TMin, Rad	10.96	6.97	0.62	0.34
Port Elliot	TMax, TMin, Evap, VP, Rad	10.46	6.89	0.65	0.26
Dookie	TMax, TMin, VP	14.59	7.88	0.58	0.28
Orbost	TMax, TMin, Rad	18.90	12.31	0.69	0.24
Cape Otway	TMax, TMin, VP	16.00	11.79	0.77	0.06
Peppermint Grove	TMax, TMin, Rad	17.06	10.12	0.69	0.13

Table 6.8: The k-NN model performance in predicting weekly rainfall with best and worst combinations of input variables in temperate zone.

Table 6.9, summarizes the performance of all selected models in predicting weekly rainfall with best combinations of input variables in temperate classification zone. Best results among all models are highlighted in bold. According to these results, at least one performance measure indicate that the CLR(Opt) is best in three out of seven locations (Moss Vale, Koppio and Port Elliot); SVM(RBF) in four locations (Moss Vale, Dookie, Orbost and Peppermint Grove) and ANN(0) in two locations (PortElliot and Cape Otway) while CR(EM) in Moss Vale; ANN(1) in Orbost; k-NN and SVM(Linear) in Peppermint Grove.

According to the performance measure RMSE, the CLR(Opt) model outperformed other models in Mossvale and Koppio; the ANN(0) outperformed in Port Elliot and Cape Otway; ANN(1) in Orbost; SVM(RBF) in Dookie and k-NN in Peppermint

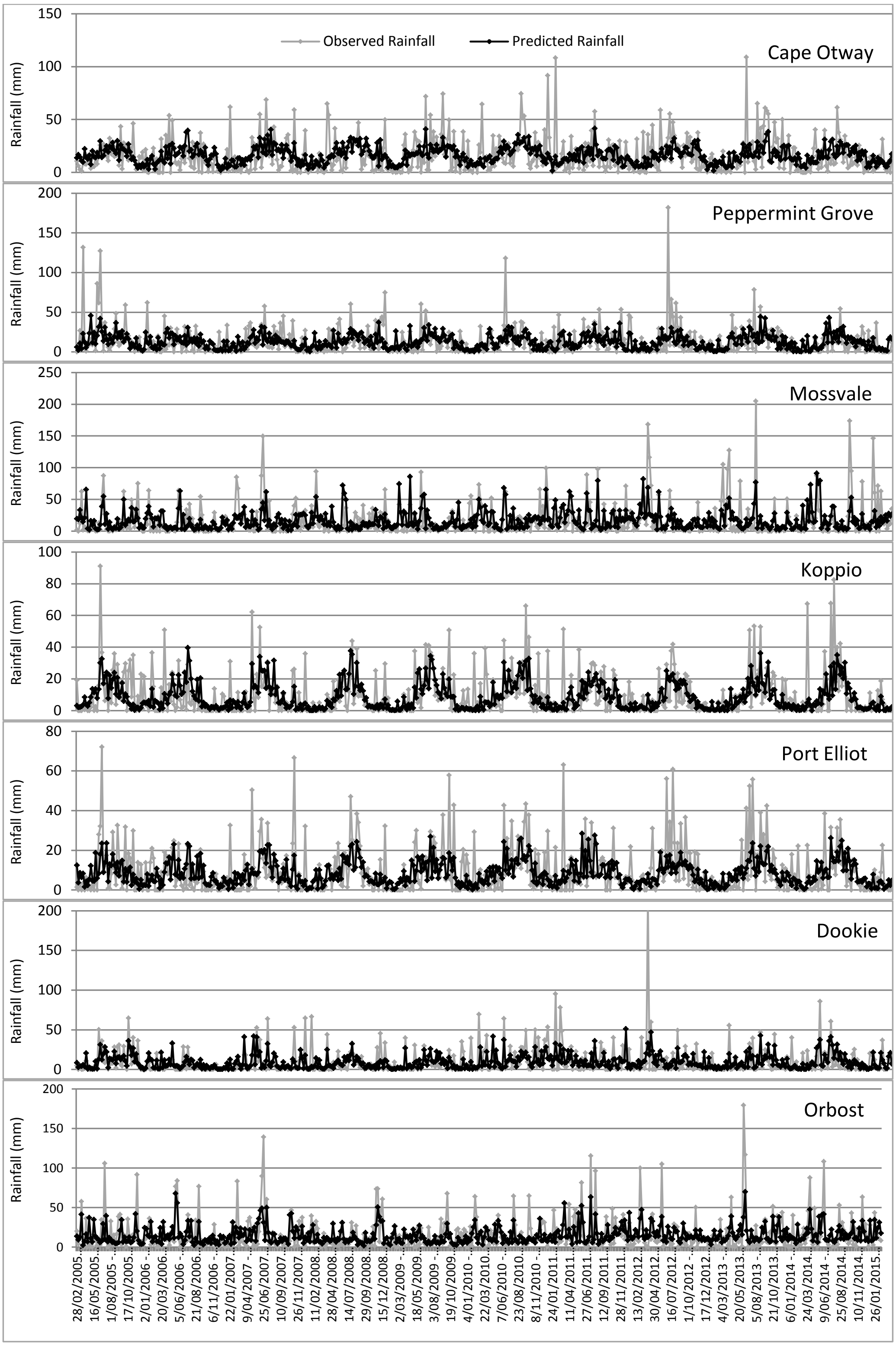

Figure 6.9: Observed rainfall vs. k-NN model weekly predictions in temperate zone.

Stations	Measures	CLR (Opt)	CR (EM)	SVM (Linear)	SVM (RBF)	MLR	ANN (0)	ANN (1)	KNN
Mossvale	RMSE	**20.04**	21.01	23.84	21.01	21.65	24.34	21.26	21.45
	MAE	12.42	**11.89**	12.64	**11.89**	13.35	12.90	13.58	13.19
	MASE	0.63	**0.60**	0.64	**0.60**	0.68	0.65	0.69	0.67
	CE	**0.42**	0.36	0.18	0.36	0.32	0.14	0.34	0.33
Koppio	RMSE	**10.55**	11.84	12.02	11.32	10.94	10.66	10.78	10.81
	MAE	**6.59**	7.12	6.89	6.70	6.76	6.63	6.78	6.78
	MASE	**0.58**	0.63	0.61	0.59	0.60	0.59	0.60	0.60
	CE	**0.39**	0.23	0.21	0.29	0.34	0.37	0.36	0.36
Port Elliot	RMSE	10.07	10.88	10.92	10.69	10.22	**10.01**	10.03	10.14
	MAE	**6.51**	7.18	6.64	6.62	6.77	6.63	6.63	6.76
	MASE	**0.61**	0.67	0.62	0.62	0.64	0.62	0.62	0.64
	CE	0.31	0.19	0.19	0.22	0.29	**0.32**	0.31	0.30
Dookie	RMSE	13.49	14.83	14.89	**12.27**	13.60	12.77	13.01	14.43
	MAE	7.43	7.41	7.16	**6.71**	7.44	7.06	7.09	7.86
	MASE	0.54	0.54	0.53	**0.49**	0.55	0.52	0.52	0.58
	CE	0.39	0.26	0.25	**0.49**	0.38	0.45	0.43	0.30
Orbost	RMSE	17.13	19.09	19.53	17.80	17.95	16.64	**16.04**	17.46
	MAE	10.86	11.88	11.06	**10.51**	11.88	11.23	10.75	11.11
	MASE	0.61	0.67	0.62	**0.59**	0.67	0.63	0.60	0.62
	CE	0.37	0.22	0.18	0.32	0.31	0.41	**0.45**	0.35
Cape Otway	RMSE	14.34	14.74	15.19	15.61	14.51	**13.41**	15.20	14.61
	MAE	9.91	10.66	9.94	10.47	10.22	**8.99**	11.05	10.48
	MASE	0.65	0.70	0.65	0.69	0.67	**0.59**	0.72	0.69
	CE	0.25	0.20	0.15	0.11	0.23	**0.34**	0.15	0.22
Peppermint Grove	RMSE	16.60	17.01	17.05	17.03	16.65	16.59	16.58	**16.44**
	MAE	10.01	10.06	9.27	**9.20**	10.41	10.39	10.01	9.85
	MASE	0.68	0.68	**0.63**	**0.63**	0.71	0.71	0.68	0.67
	CE	0.18	0.14	0.13	0.14	0.17	0.18	0.18	**0.20**

Table 6.9: Models performance for predicting weekly rainfall with best combination of input variables in temperate classification zone.

Grove. In Port Elliot, the CLR(Opt) model performance is 0.60% lower than the best one; in Cape Otway 6.49%; in Orbost 6.36%; in Dookie 9.04% and in Peppermint Grove 0.96%. In two locations (Port Elliot and Peppermint Grove), the CLR(Opt) model performance is less than 1% lower than the best models.

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 5.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8 and 6.9, show that all models follow the series patterns at all locations.

Results presented in this section demonstrate that the CLR(Opt) and ANN(0) models are superior than other models. Both CLR(Opt) and SVM(RBF) models are the most suitable models in finding the pattern and trends of the observations in temperate classification zone for weekly rainfall predictions.

6.2 Weekly rainfall predictions in grassland zone

In this section, first we present the weekly rainfall prediction results of each model in predicting weekly rainfall with best and worst combinations of input variables in grassland classification zone. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models.

Table 6.10, summarizes the prediction performance of the CLR(Opt) model with the best and worst combinations of input variables in grassland classification zone. According to these results the CLR(Opt) model provides best predictions with input variables TMax, TMin and VP in three out of eight locations (Alexandria, Annuello and Ningaloo); with TMax, TMin and VP in two locations (Blinman and Dowerine) and with full set of five meteorological variables in the remaining three locations (Warren, Richmond and Newry). The model provides worst predictions with input variables TMax, TMin and Rad in all eight locations.

The performance measure RMSE for the CLR(Opt) model in predicting weekly rainfall ranges from 8.51 to 28.49, MAE from 4.28 to 11.92, MASE from 0.53 to 0.63 and CE from 0.30 to 0.59. The performance measures RMSE indicates that the model provides best predictions in Dowerine; MAE indicates in Ningaloo; MASE indicates in Warren and CE indicates in Richmond while the performance RMSE, MAE and MASE indicate worst prediction in Newry and CE indicate in Dowerine (see Figure 6.10). The graphical display of observed rainfall and CLR(Opt) model predictions over the test period is given in Figure 6.11.

Table 6.11, summarizes the prediction performance of the CR(EM) model with best and worst combinations of input variables in grassland classification zone. According to these results, the CR(EM) model provides best predictions with input variables TMax, TMin and Rad in one location (Warren); with input variables TMax, TMin, VP in three locations (Newry, Alexandria and Blinman); with TMax, TMin, VP and Rad in two locations (Annuello and Ningaloo) and with full set of five meteorological variables in two locations (Richmond and Dowerine). The model provides worst predictions with input variables TMax, TMin and Rad in all locations except Warren. In Warren, the model provides worst predictions with full set of input meteorological parameters.

The performance measure RMSE for the CR(EM) model in predicting weekly rainfall ranges from 9.39 to 37.08, MAE from 4.53 to 16.50, MASE from 0.49 to 0.88 and CE from 0.15 to 0.59. The performance measures RMSE indicates the model provides best predictions with lowest prediction error in Dowerine; MAE indicates in the location Ningaloo; MASE indicates in the location Warren and CE indicates in the location Richmond while all four performance measures indicate worst predictions

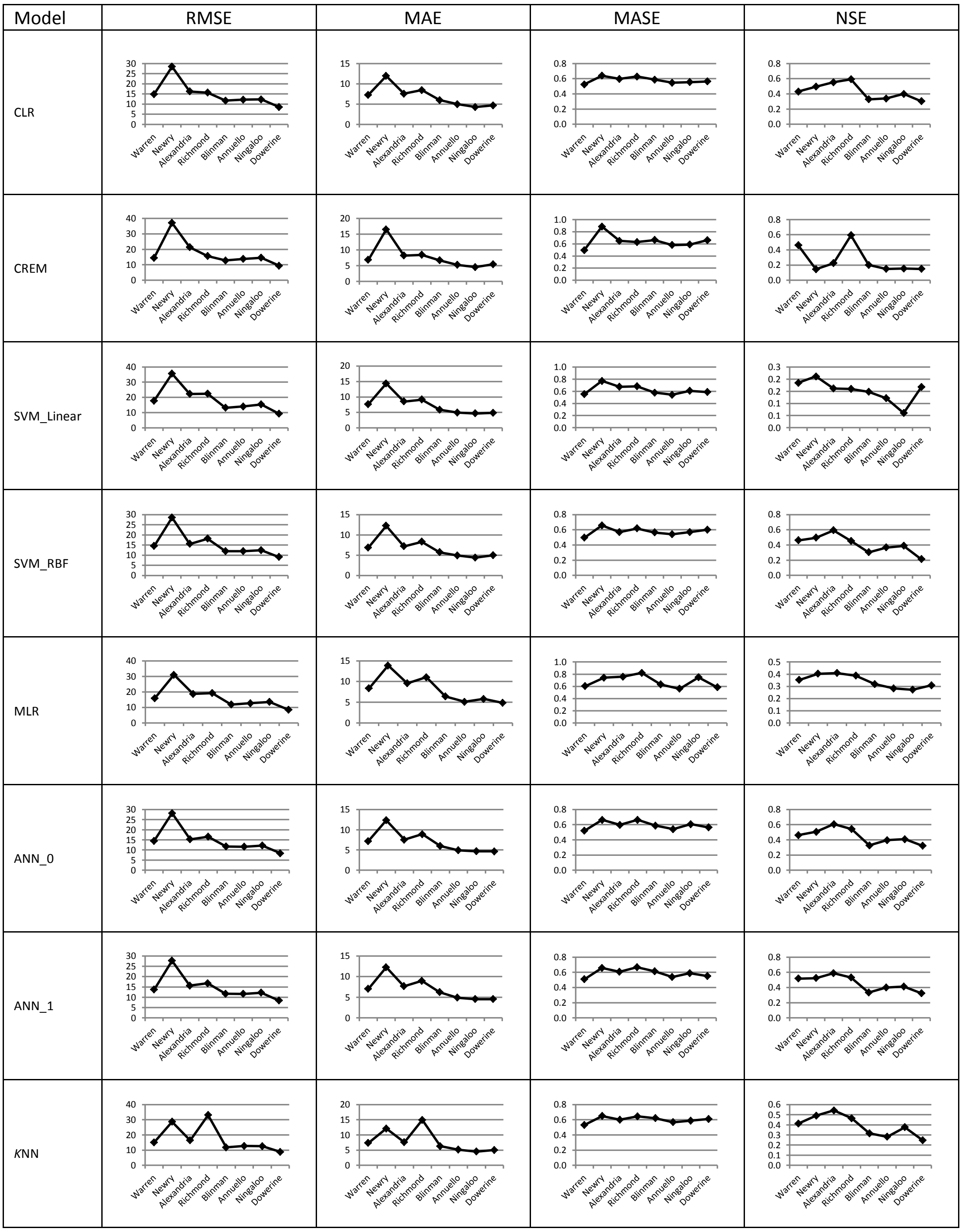

Figure 6.10: Graphical display of models performance measures in predicting weekly rainfall in grassland zone.

in Newry (see Figure 6.10). The graphical display of observed rainfall and CR(EM) model predictions over the test period are given in Figure 6.12.

Table 6.12, summarizes the prediction performance of the SVM(Linear) model

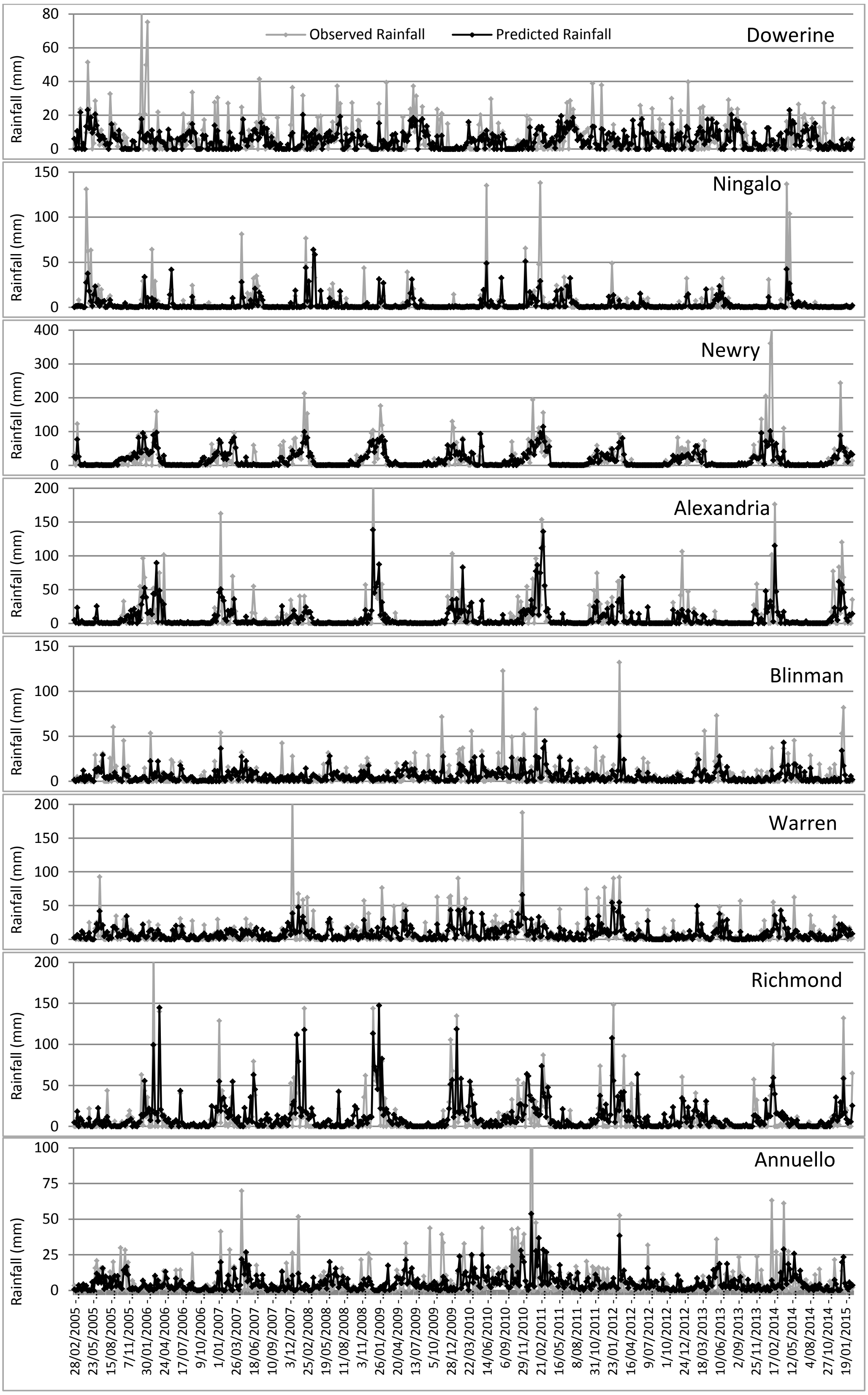

Figure 6.11: Observed rainfall vs. CLR(Opt) model weekly predictions in grassland classification zone.

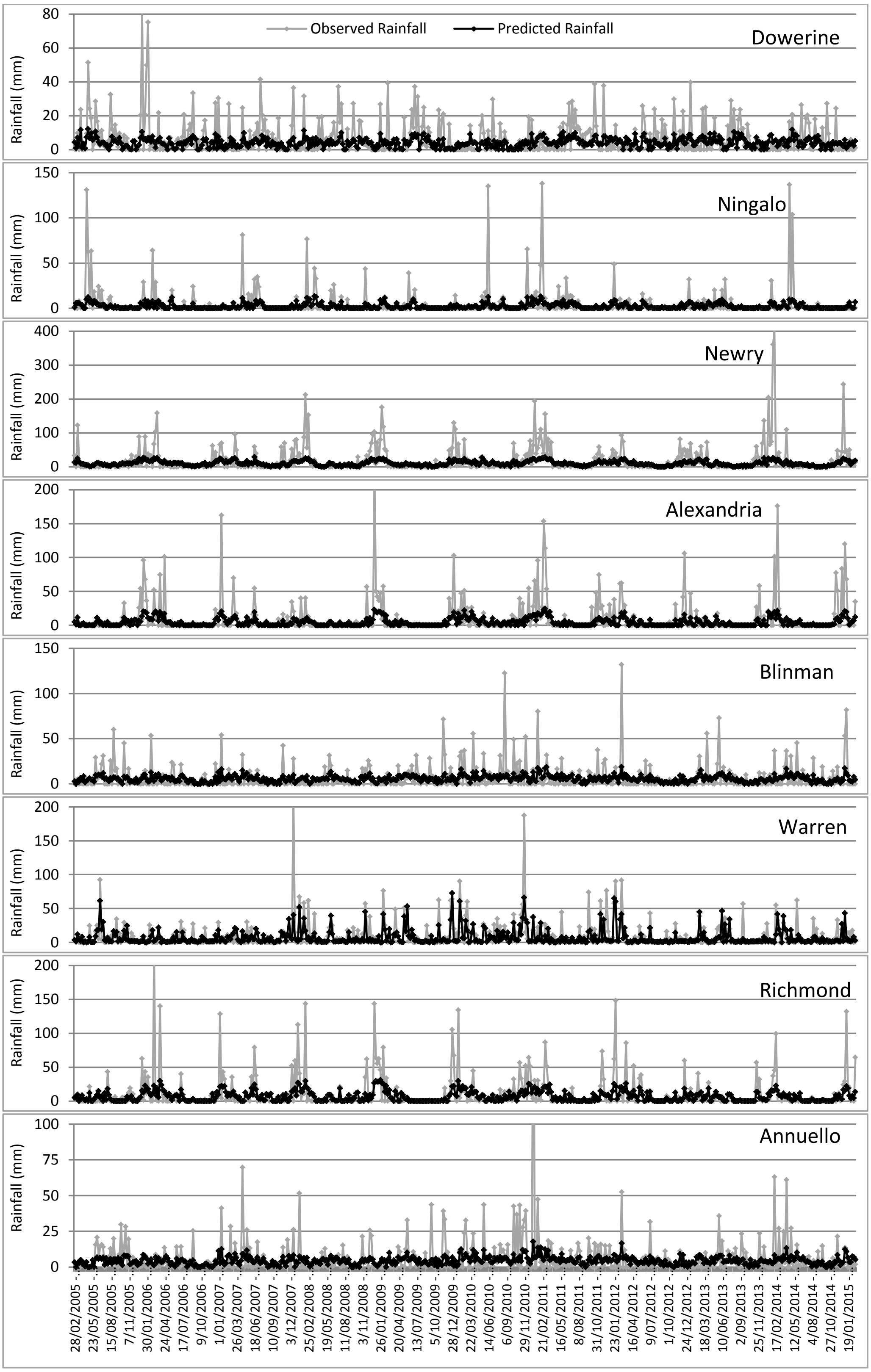

Figure 6.12: Observed rainfall vs. CR(EM) model weekly predictions in grassland classification zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Evap, VP, Rad	14.91	7.27	0.53	0.43
Newry	TMax, TMin, Evap, VP, Rad	28.49	11.92	0.64	0.50
Alexandria	TMax, TMin, VP	16.25	7.55	0.60	0.55
Richmond	TMax, TMin, Evap, VP, Rad	15.63	8.44	0.63	0.59
Blinman	TMax, TMin, VP, Rad	11.72	5.97	0.59	0.33
Annuello	TMax, TMin, Vp	12.18	4.98	0.55	0.34
Ningaloo	TMax, TMin, VP	12.31	4.28	0.55	0.40
Dowerine	TMax, TMin, VP, Rad	8.51	4.66	0.56	0.30
Performance measures for worst input combinations					
Warren	TMax, TMin, Rad	15.33	7.48	0.54	0.40
Newry	TMax, TMin, Rad	29.71	12.71	0.68	0.45
Alexandria	TMax, TMin, Rad	17.84	8.63	0.68	0.46
Richmond	TMax, TMin, Rad	18.20	9.01	0.67	0.45
Blinman	TMax, TMin, Rad	12.12	6.21	0.61	0.28
Annuello	TMax, TMin, Rad	13.15	4.95	0.54	0.23
Ningaloo	TMax, TMin, Rad	13.48	4.68	0.61	0.28
Dowerine	TMax, TMin, Rad	8.76	4.81	0.58	0.26

Table 6.10: The CLR(Opt) model performance for weekly rainfall predictions with best and worst combinations of input variables in grassland classification zone.

with best and worst combinations of input variables in grassland classification zone. The results show that the SVM(Linear) model provides best predictions with input variables TMax, TMin, VP and Rad in two locations (Annuello and Dowerine); with full set of five meteorological variables in the remaining six locations (Warren, Newry, Alexandria, Richmond, Blinman and Ningaloo). The SVM(Linear) model provides worst predictions with input variables TMax, TMin and Rad in all locations except Ningaloo. In Ningaloo, the model provides worst predictions with the input variables TMax, TMin and VP.

The performance measure RMSE for the SVM(Linear) model in predicting weekly rainfall ranges from 9.29 to 35.63, MAE from 4.68 to 14.34, MASE from 0.54 to 0.77 and CE from 0.0.06 to 0.21. The performance measures RMSE indicates the model has lowest prediction error in the location Dowerine; MAE indicates in the location Ningaloo; MASE indicates in the location Annuello and CE indicates in the location Newry while performance measures RMSE, MAE and MASE indicates highest prediction error in the location Newry and CE in the location Ningaloo (see Figure 6.10). The graphical display of observed rainfall and SVM(Linear) model predictions are given in Figure 6.13.

Table 6.13, summarizes the prediction performance of the SVM(RBF) model with best and worst combinations of input variables in grassland classification zone. According to these results the SVM(RBF) model provides best predictions with input

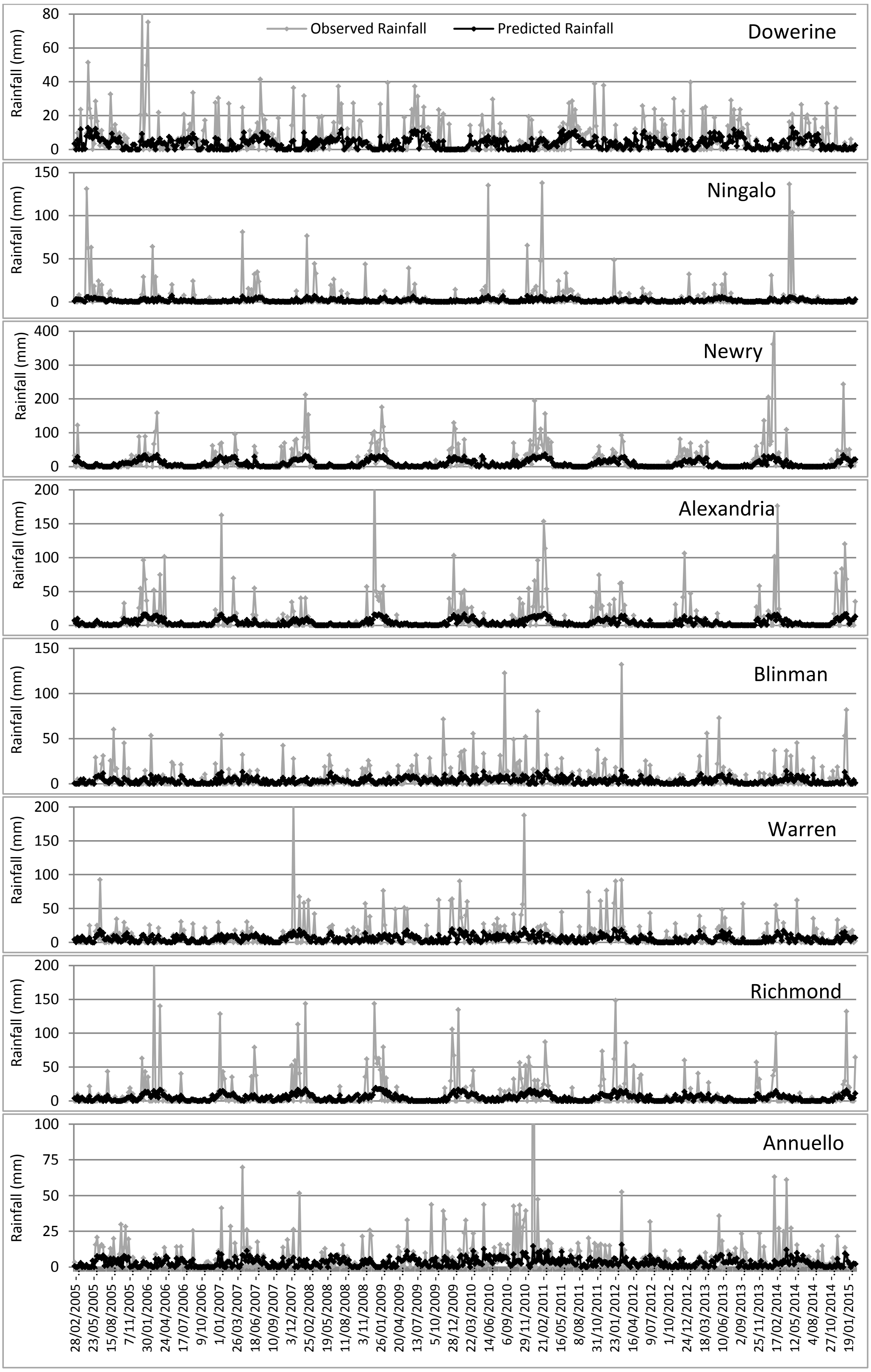

Figure 6.13: Observed rainfall vs. SVM(Linear) model weekly predictions in grassland classification zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Rad	14.50	6.84	0.49	0.46
Newry	TMax, TMin, VP	37.08	16.50	0.88	0.15
Alexandria	TMax, TMin, VP	21.38	8.19	0.65	0.23
Richmond	TMax, TMin, Evap, VP, Rad	15.63	8.44	0.63	0.59
Blinman	TMax, TMin, VP	12.78	6.72	0.66	0.20
Annuello	TMax, TMin, VP, Rad	13.82	5.28	0.58	0.15
Ningaloo	TMax, TMin, VP, Rad	14.61	4.53	0.59	0.15
Dowerine	TMax, TMin, Evap, VP, Rad	9.39	5.45	0.66	0.15
Performance measures for worst input combinations					
Warren	TMax, TMin, Evap, VP, Rad	15.75	7.44	0.54	0.36
Newry	TMax, TMin, Rad	37.27	17.17	0.92	0.14
Alexandria	TMax, TMin, Rad	21.55	8.36	0.66	0.21
Richmond	TMax, TMin, Rad	18.20	9.01	0.67	0.45
Blinman	TMax, TMin, Rad	13.39	6.86	0.68	0.12
Annuello	TMax, TMin, Rad	14.05	5.59	0.61	0.12
Ningaloo	TMax, TMin, Rad	14.70	4.64	0.60	0.14
Dowerine	TMax, TMin, Rad	9.52	5.56	0.67	0.13

Table 6.11: The CR(EM) model performance for weekly rainfall predictions with best and worst combinations of input variables in grassland classification zone.

variables TMax, TMin and Rad in one location (Warren); with TMax, Tmin and VP in three locations (Alexandria, Blinman and Annuello); with TMax, TMin, VP and Rad in Richmond and with full set of five meteorological variables in the remaining three locations (Newry, Ningaloo and Dowerine) while the model provides worst predictions with input variables TMax, TMin and Rad in two locations (Newry and Richmond); with TMax, TMin and VP in two locations (Ningaloo and Dowerine); with TMax, TMin, VP and Rad in Blinman and with full set of variables in three locations (Warren, Alexandria and Annuello).

The performance measure RMSE for the SVM(RBF) model in predicting weekly rainfall ranges from 9.04 to 28.52, MAE from 4.39 to 12.24, MASE from 0.49 to 0.66 and CE from 0.21 to 0.59. The performance measures RMSE indicates the SVM(RBF) model provides best predictions in the location Dowerine; MAE indicates in the location Ningaloo; MASE indicates in the location Warren and CE indicates in the location Alexandria while performance measures RMSE, MAE and MASE indicates worst predictions in the location Newry; CE in the location Dowerine (see Figure 6.10). The graphical display of observed rainfall and SVM(RBF) model predictions over the test period is given in Figure 6.14.

Table 6.14, summarizes the prediction performance of the MLR model with best and worst combinations of input variables in grassland classification zone. The results show that the MLR model provides best predictions with input variables TMax,

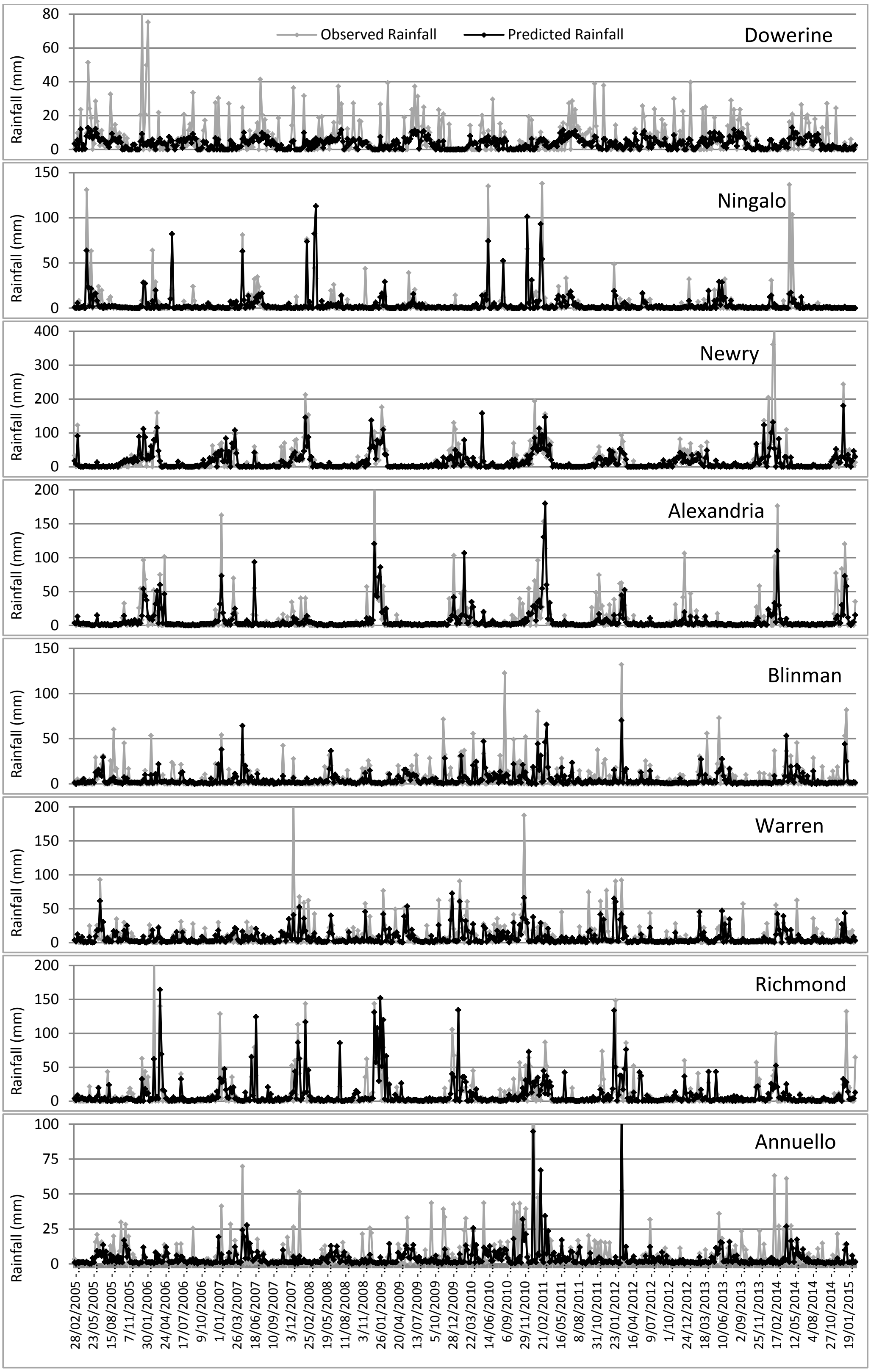

Figure 6.14: Observed rainfall vs. SVM(RBF) model weekly predictions in grassland classification zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Evap, VP, Rad	17.83	7.64	0.55	0.18
Newry	TMax, TMin, Evap, VP, Rad	35.63	14.34	0.77	0.21
Alexandria	TMax, TMin, Evap, VP, Rad	22.25	8.51	0.67	0.16
Richmond	TMax, TMin, Evap, VP, Rad	22.46	9.15	0.68	0.16
Blinman	TMax, TMin, Evap, VP, Rad	13.21	5.86	0.58	0.15
Annuello	TMax, TMin, VP, Rad	14.05	4.93	0.54	0.12
Ningaloo	TMax, TMin, Evap, VP, Rad	15.39	4.68	0.61	0.06
Dowerine	TMax, TMin, VP, Rad	9.29	4.86	0.59	0.17
Performance measures for worst input combinations					
Warren	TMax, TMin, Rad	18.14	7.88	0.57	0.16
Newry	TMax, TMin, Rad	37.48	15.42	0.83	0.13
Alexandria	TMax, TMin, Rad	23.12	8.84	0.70	0.09
Richmond	TMax, TMin, Rad	23.53	9.53	0.71	0.08
Blinman	TMax, TMin, Rad	13.49	6.01	0.59	0.11
Annuello	TMax, TMin, Rad	14.26	5.01	0.55	0.10
Ningaloo	TMax, TMin, VP	15.51	4.70	0.61	0.05
Dowerine	TMax, TMin, Rad	9.54	4.94	0.60	0.12

Table 6.12: The SVM(Linear) model performance for weekly rainfall prediction with best and worst combinations of input variables in grassland zone.

TMin and VP in the location Alexandria; with TMax, TMin, VP and Rad in two locations (Annuello and Ningaloo) and with full set of five meteorological variables in the remaining five locations (Warren, Newry, Richmond, Blinman and Dowerine) while the model provides worst predictions with input variables TMax, TMin and Rad in all eight locations.

The performance measure RMSE for the MLR model in predicting weekly rainfall ranges from 9.04 to 28.52, MAE from 4.39 to 12.24, MASE from 0.49 to 0.66 and CE from 0.21 to 0.59. The performance measures RMSE and MAE indicate that the MLR model provides best predictions with lowest prediction error in the location Dowerine; MASE indicates in the location Annuello and CE indicates in the location Alexandria while the performance measures RMSE, MAE indicates worst predictions in the location Newry; MASE in the location Richmond and CE in the location Ningaloo (see Figure 6.10). The graphical display of observed rainfall and MLR model predictions over the test period is given in Figure 6.15.

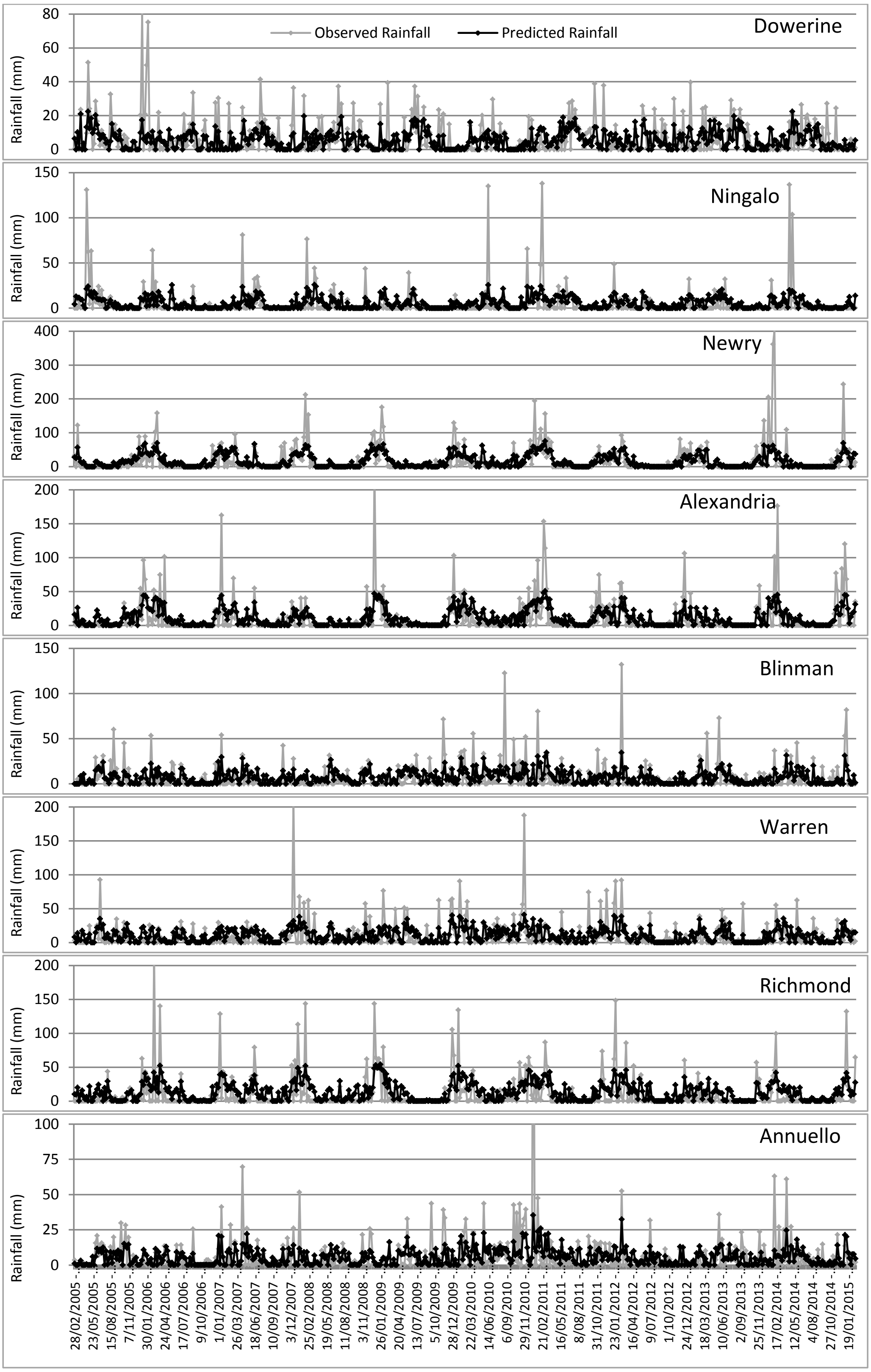

Figure 6.15: Observed rainfall vs. MLR model weekly predictions in grassland classification zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Rad	14.50	6.84	0.49	0.46
Newry	TMax, TMin, Evap, VP, Rad	28.52	12.24	0.66	0.49
Alexandria	TMax, TMin, VP	15.51	7.18	0.57	0.59
Richmond	TMax, TMin, VP, Rad	18.13	8.29	0.62	0.45
Blinman	TMax, TMin, VP	11.94	5.72	0.56	0.30
Annuello	TMax, TMin, VP	11.92	4.92	0.54	0.37
Ningaloo	TMax, TMin, Evap, VP, Rad	12.43	4.39	0.57	0.39
Dowerine	TMax, TMin, Evap, VP, Rad	9.04	4.97	0.60	0.21
Performance measures for worst input combinations					
Warren	TMax, TMin, Evap, VP, Rad	15.75	7.44	0.54	0.36
Newry	TMax, TMin, Rad	31.12	12.91	0.69	0.40
Alexandria	TMax, TMin, Evap, VP, Rad	18.69	8.46	0.67	0.41
Richmond	TMax, TMin, Rad	19.67	8.54	0.64	0.36
Blinman	TMax, TMin, VP, Rad	12.48	5.70	0.56	0.24
Annuello	TMax, TMin, Evap, VP, Rad	12.25	5.32	0.58	0.33
Ningaloo	TMax, TMin, VP	12.98	4.58	0.59	0.33
Dowerine	TMax, TMin, VP	9.55	4.86	0.59	0.12

Table 6.13: The SVM(RBF) model performance for weekly rainfall prediction with best and worst combinations of input variables in grassland classification zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Evap, VP, Rad	15.89	8.36	0.60	0.35
Newry	TMax, TMin, Evap, VP, Rad	30.98	13.85	0.74	0.40
Alexandria	TMax, TMin, VP	18.68	9.58	0.76	0.41
Richmond	TMax, TMin, Evap, VP, Rad	19.18	11.03	0.82	0.39
Blinman	TMax, TMin, Evap, VP, Rad	11.81	6.40	0.63	0.32
Annuello	TMax, TMin, VP, Rad	12.69	5.10	0.56	0.28
Ningaloo	TMax, TMin, VP, Rad	13.54	5.78	0.75	0.27
Dowerine	TMax, TMin, Evap, VP, Rad	8.47	4.83	0.58	0.31
Performance measures for worst input combinations					
Warren	TMax, TMin, Rad	16.33	8.69	0.63	0.32
Newry	TMax, TMin, Rad	32.36	15.50	0.83	0.35
Alexandria	TMax, TMin, Rad	19.28	10.03	0.79	0.37
Richmond	TMax, TMin, Rad	20.08	11.62	0.86	0.33
Blinman	TMax, TMin, Rad	12.41	6.97	0.69	0.25
Annuello	TMax, TMin, Rad	13.18	5.29	0.58	0.23
Ningaloo	TMax, TMin, Rad	13.68	5.82	0.75	0.26
Dowerine	TMax, TMin, Rad	8.72	4.82	0.58	0.27

Table 6.14: The MLR model performance for weekly rainfall predictions with best and worst combinations of input variables in grassland classification zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Evap, VP, Rad	14.49	7.19	0.52	0.46
Newry	TMax, TMin, Evap, VP, Rad	28.18	12.36	0.66	0.51
Alexandria	TMax, TMin, VP, Rad	15.24	7.54	0.60	0.61
Richmond	TMax, TMin, Evap, VP, Rad	16.58	8.90	0.66	0.54
Blinman	TMax, TMin, VP, Rad	11.74	5.97	0.59	0.33
Annuello	TMax, TMin, VP	11.64	4.92	0.54	0.40
Ningaloo	TMax, TMin, VP	12.22	4.68	0.61	0.41
Dowerine	TMax, TMin, VP, Rad	8.39	4.66	0.56	0.32
Performance measures for worst input combinations					
Warren	TMax, TMin, VP	15.97	8.44	0.61	0.35
Newry	TMax, TMin, Rad	32.36	15.50	0.83	0.35
Alexandria	TMax, TMin, Evap, VP, Rad	18.69	9.55	0.76	0.41
Richmond	TMax, TMin, Rad	19.68	9.76	0.73	0.36
Blinman	TMax, TMin, Rad	12.13	6.57	0.65	0.28
Annuello	TMax, TMin, Evap, VP, Rad	12.70	5.05	0.55	0.28
Ningaloo	TMax, TMin, Evap, VP, Rad	13.55	5.83	0.76	0.27
Dowerine	TMax, TMin, Rad	8.75	4.83	0.58	0.26

Table 6.15: The ANN(0) model performance for weekly rainfall predictions with best and worst combinations of input variables in grassland zone.

Table 6.15, summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables in grassland classification zone. According to these results the ANN(0) model provides the best predictions with input variables TMax, TMin and VP in two locations (Annuello and Ningaloo); with TMax, Tmin, VP and Rad in three locations (Alexandria, Blinman and Dowerine) and with full set of five meteorological variables in the remaining three locations (Warren, Newry and Richmond). The model provides worst predictions with input variables TMax, TMin and Rad in four locations (Newry, Richmond, Blinman and Dowerine); with TMax, TMin and VP in the location Warren and with full set of variables in three locations (Alexandria, Annuello and Ningaloo).

The performance measure RMSE for the ANN(0) model in predicting weekly rainfall ranges from 8.39 to 28.18, MAE from 4.66 to 12.36, MASE from 0.52 to 0.66 and CE from 0.32 to 0.61.. The performance measures RMSE and MAE indicate that the ANN(0) model provides best predictions in the location Dowerine; MASE indicates in the location Warren and CE indicates in the location Alexandria while the performance measures RMSE, MAE and MASE indicate worst predictions in the location Newry and CE in the location Dowerine (see Figure 6.10). The graphical display of observed rainfall and ANN(0) model predictions over the test period is given in Figure 6.16.

Table 6.16, summarizes the prediction performance of the ANN(1) model with best

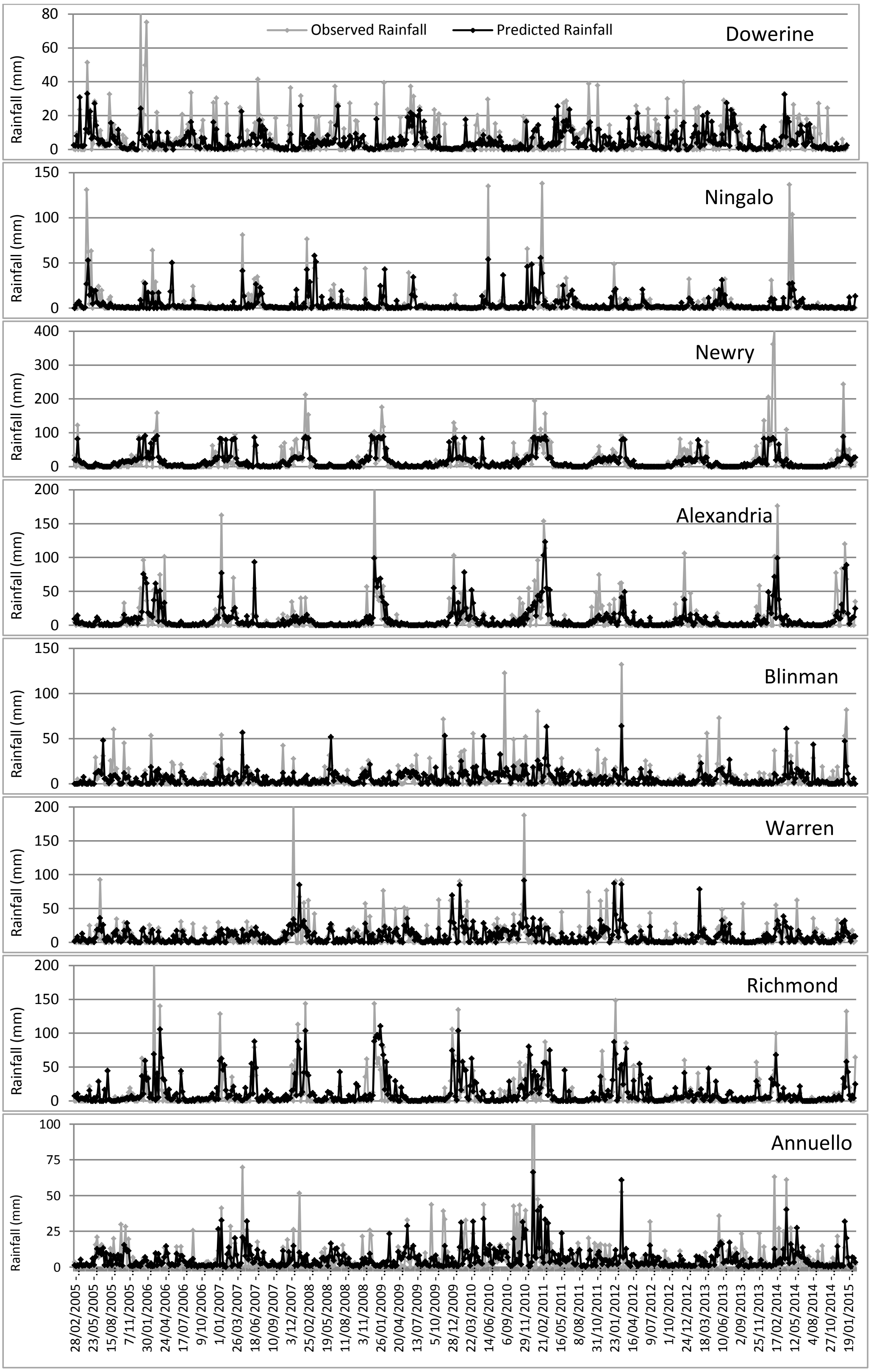

Figure 6.16: Observed rainfall vs. ANN(0) model weekly predictions in grassland classification zone.

and worst combinations of input variables in grassland classification zone. According to these results the ANN(1) model provides the best predictions with input variables TMax, TMin and VP in two locations (Annuello and Blinman); with TMax, Tmin, VP and Rad in the location Dowerine and with full set of five meteorological variables in the remaining five locations (Warren, Newry, Alexandria, Richmond, Ningaloo). The model provides worst predictions with input variables TMax, TMin and Rad in five locations (Warren, Newry, Alexandria, Blinman and Ningaloo); with TMax, TMin and VP in the location Dowerine and with TMax, TMin, VP and Rad in two locations (Annuello and Richmond).

The performance measure RMSE for the ANN(1) model in predicting weekly rainfall ranges from 8.37 to 27.68, MAE from 4.53 to 12.25, MASE from 0.51 to 0.67 and CE from 0.32 to 0.59. The performance measure RMSE indicates that the ANN(1) model provides best predictions in the location Dowerine; MAE indicates in the location Ningaloo; MASE indicates in the location Warren and CE indicates in the location Alexandria while the performance measures RMSE and MAE indicates worst predictions in the location Newry; MASE indicates in the location Richmond and CE in the location Dowerine (see Figure 6.10). The graphical display of observed rainfall and ANN(1) model predictions fover the test period is given in Figure 6.17.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, Evap, VP, Rad	13.72	7.03	0.51	0.52
Newry	TMax, TMin, Evap, VP, Rad	27.68	12.25	0.66	0.52
Alexandria	TMax, TMin, Evap, VP, Rad	15.60	7.67	0.61	0.59
Richmond	TMax, TMin, Evap, VP, Rad	16.79	8.96	0.67	0.53
Blinman	TMax, TMin, VP	11.68	6.22	0.61	0.33
Annuello	TMax, TMin, VP	11.61	4.89	0.54	0.40
Ningaloo	TMax, TMin, Evap, VP, Rad	12.18	4.53	0.59	0.41
Dowerine	TMax, TMin, VP, Rad	8.37	4.55	0.55	0.32
Performance measures for worst input combinations					
Warren	TMax, TMin, Rad	14.72	7.75	0.56	0.44
Newry	TMax, TMin, Rad	32.01	16.90	0.91	0.36
Alexandria	TMax, TMin, Rad	16.60	7.74	0.61	0.53
Richmond	TMax, TMin, VP, Rad	19.11	9.90	0.74	0.39
Blinman	TMax, TMin, Rad	12.65	6.52	0.64	0.22
Annuello	TMax, TMin, VP, Rad	14.99	7.38	0.81	0.00
Ningaloo	TMax, TMin, Rad	13.64	5.64	0.73	0.26
Dowerine	TMax, TMin, VP	9.29	5.50	0.66	0.17

Table 6.16: The ANN(1) model performance for weekly rainfall predictions with best and worst combinations of input variables in grassland zone.

Table 6.17, summarizes the performance of the k-NN model for weekly rainfall predictions with best and worst combinations of input variables in grassland classifi-

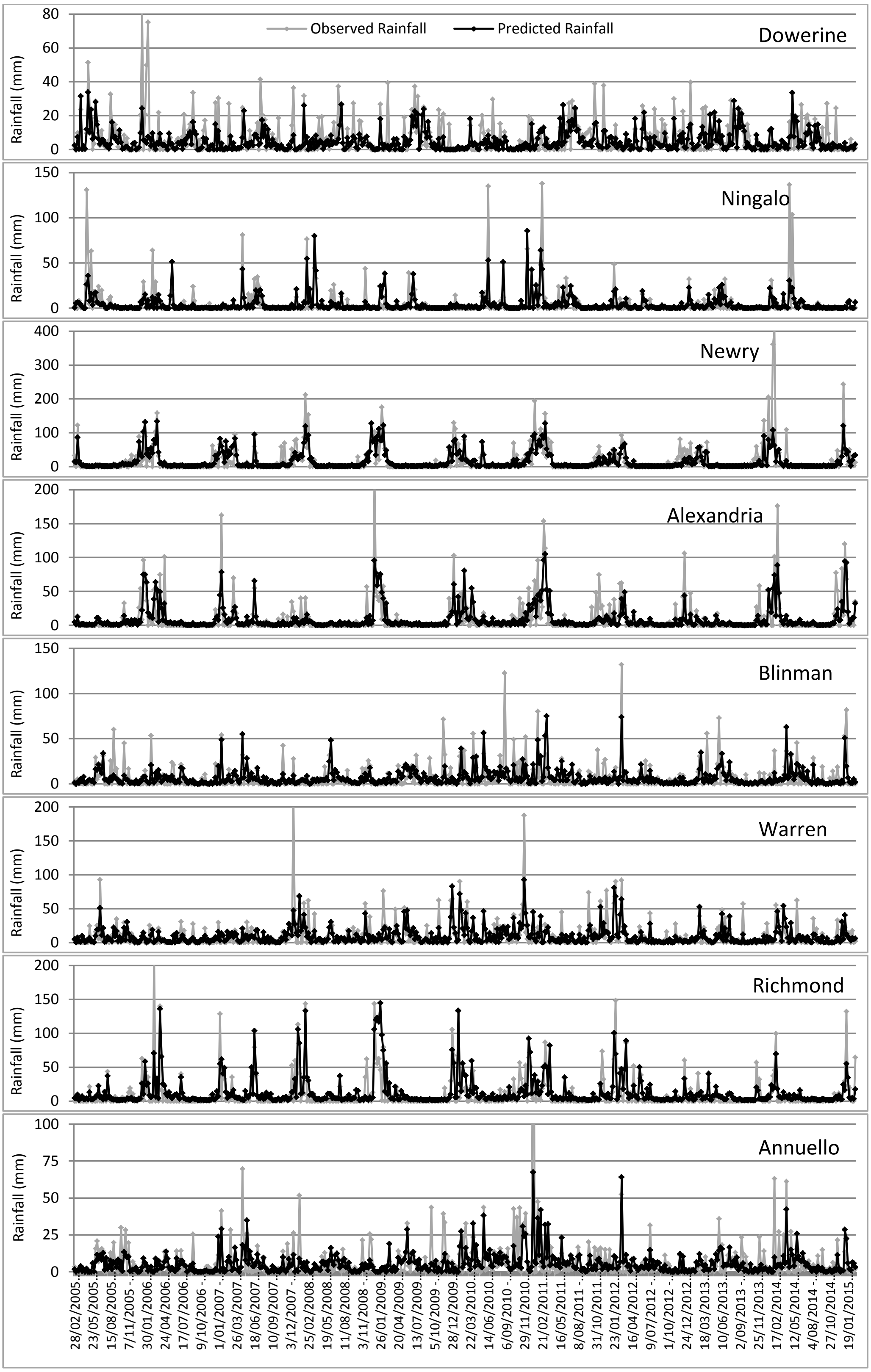

Figure 6.17: Observed rainfall vs. ANN(1) model weekly predictions in grassland classification zone.

cation zone.The results presented show that the k-NN model provides best predictions with input variables TMax, TMin and VP in six locations (Warren, Alexandria, Blinman, Annuello, Ningaloo and Dowerine); with TMax, Tmin and Rad in the location Richmond and with full set of five meteorological variables in the location Newry. The k-NN model provides the worst predictions with input variables TMax, TMin and Rad in all locations except Richmond. In Richmond, the model provides worst predictions with full set of five input variables.

The performance measure RMSE for the k-NN model in predicting weekly rainfall ranges from 8.83 to 33.11, MAE from 4.53 to 14.93, MASE from 0.53 to 0.65 and CE from 0.25 to 0.54. The performance measures RMSE indicates the k-NN model provides best predictions in the location Dowerine; MAE indicates in the location Ningaloo; MASE indicates in the location Warren and CE indicates in the location Alexandria while the performance measures RMSE and MAE indicate worst predictions in the location Richmond; MASE indicates in the location Newry and CE in the location Dowerine (see Figure 6.10). The graphical display of observed rainfall and k-NN model predictions over the test period is given in Figure 6.18.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Warren	TMax, TMin, VP	15.13	7.35	0.53	0.41
Newry	TMax, TMin, Evap, VP, Rad	28.65	12.10	0.65	0.49
Alexandria	TMax, TMin, VP	16.44	7.58	0.60	0.54
Richmond	TMax, TMin, Rad	33.11	14.93	0.64	0.46
Blinman	TMax, TMin, VP	11.83	6.30	0.62	0.32
Annuello	TMax, TMin, VP	12.71	5.16	0.57	0.28
Ningaloo	TMax, TMin, VP	12.55	4.53	0.59	0.38
Dowerine	TMax, TMin, VP	8.83	5.04	0.61	0.25
Performance measures for worst input combinations					
Warren	TMax, TMin, Rad	15.86	8.05	0.58	0.36
Newry	TMax, TMin, Rad	30.90	13.13	0.70	0.41
Alexandria	TMax, TMin, Rad	18.33	8.28	0.66	0.43
Richmond	TMax, TMin, Evap, VP, Rad	33.90	15.21	0.66	0.44
Blinman	TMax, TMin, Rad	12.46	6.52	0.64	0.24
Annuello	TMax, TMin, Rad	13.15	5.18	0.57	0.23
Ningaloo	TMax, TMin, Rad	13.56	4.87	0.63	0.27
Dowerine	TMax, TMin, Rad	9.06	5.01	0.61	0.21

Table 6.17: The k-NN model performance for weekly rainfall predictions with best and worst combinations of input variables in grassland zone.

Table 6.18, summarizes the performance of all eight models in predicting weekly rainfall with best combinations of input variables in grassland classification zone. Best results among all models are highlighted in bold. According to these results, at least one performance measure indicate that the CLR(Opt) is best in four out of eight

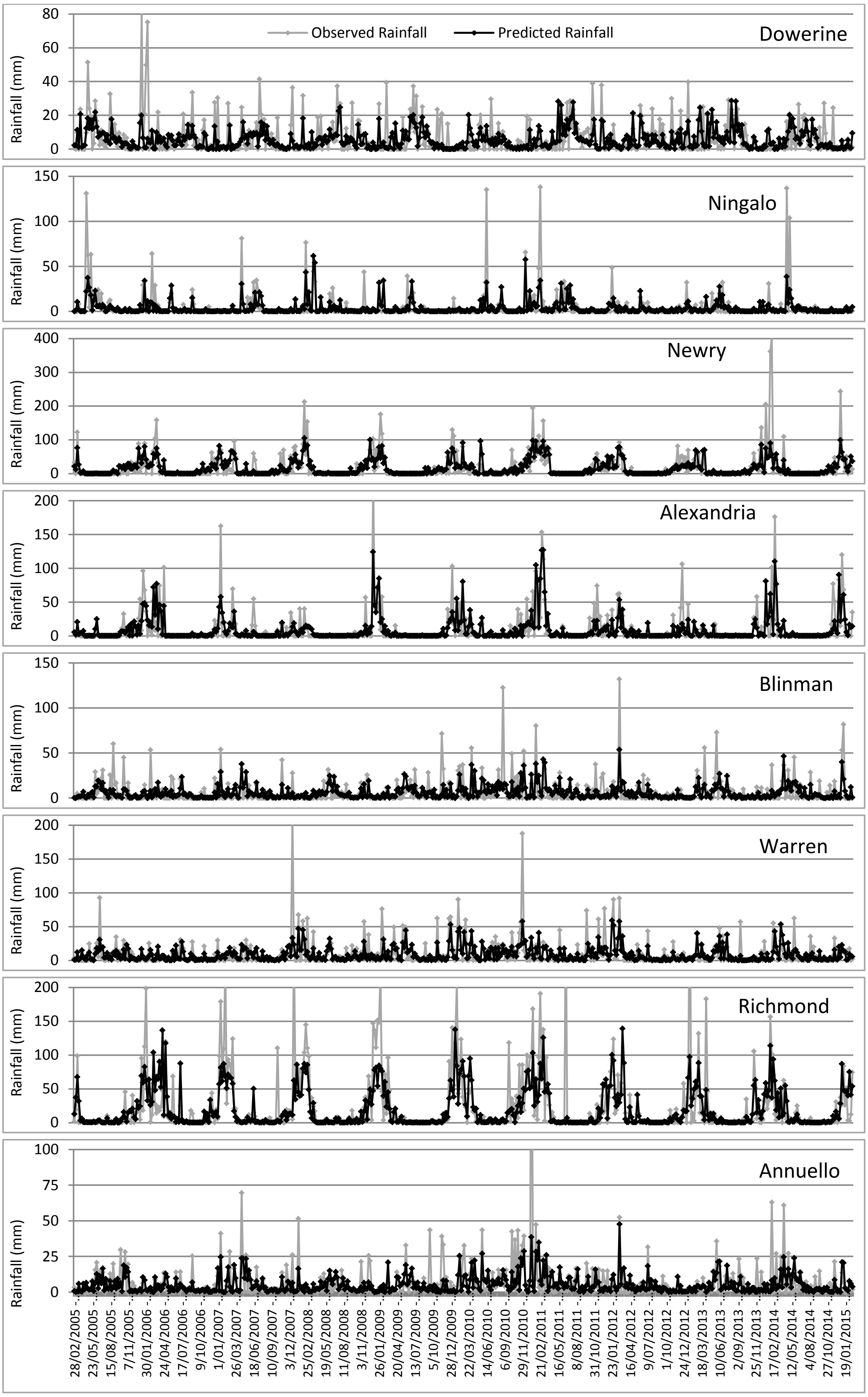

Figure 6.18: Observed rainfall vs. k-NN model weekly predictions in grassland classification zone.

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	MLR	ANN (0)	ANN (1)	KNN
Warren	RMSE	14.91	14.50	17.83	14.50	15.89	14.49	**13.72**	15.13
	MAE	7.27	**6.84**	7.64	**6.84**	8.36	7.19	7.03	7.35
	MASE	0.53	**0.49**	0.55	**0.49**	0.60	0.52	0.51	0.53
	CE	0.43	0.46	0.18	0.46	0.35	0.46	**0.52**	0.41
Newry	RMSE	28.49	37.08	35.63	28.52	30.98	28.18	**27.68**	28.65
	MAE	**11.92**	16.50	14.34	12.24	13.85	12.36	12.25	12.10
	MASE	**0.64**	0.88	0.77	0.66	0.74	0.66	0.66	0.65
	CE	0.50	0.15	0.21	0.49	0.40	0.51	**0.52**	0.49
Alexandria	RMSE	16.25	21.38	22.25	15.51	18.68	**15.24**	15.60	16.44
	MAE	7.55	8.19	8.51	**7.18**	9.58	7.54	7.67	7.58
	MASE	0.60	0.65	0.67	**0.57**	0.76	0.60	0.61	0.60
	CE	0.55	0.23	0.16	0.59	0.41	**0.61**	0.59	0.54
Richmond	RMSE	**15.63**	**15.63**	22.46	18.13	19.18	16.58	16.79	33.11
	MAE	8.44	8.44	9.15	**8.29**	11.03	8.90	8.96	14.93
	MASE	0.63	0.63	0.68	**0.62**	0.82	0.66	0.67	0.64
	CE	**0.59**	**0.59**	0.16	0.45	0.39	0.54	0.53	0.46
Blinman	RMSE	11.72	12.78	13.21	11.94	11.81	11.74	**11.68**	11.83
	MAE	5.97	6.72	5.86	**5.72**	6.40	5.97	6.22	6.30
	MASE	0.59	0.66	0.58	**0.56**	0.63	0.59	0.61	0.62
	CE	**0.33**	0.20	0.15	0.30	0.32	**0.33**	**0.33**	0.32
Annuello	RMSE	12.18	13.82	14.05	11.92	12.69	11.64	**11.61**	12.71
	MAE	4.98	5.28	4.93	4.92	5.10	4.92	**4.89**	5.16
	MASE	0.55	0.58	**0.54**	**0.54**	0.56	**0.54**	**0.54**	0.57
	CE	0.34	0.15	0.12	0.37	0.28	**0.40**	**0.40**	0.28
Ningaloo	RMSE	12.31	14.61	15.39	12.43	13.54	12.22	**12.18**	12.55
	MAE	**4.28**	4.53	4.68	4.39	5.78	4.68	4.53	4.53
	MASE	**0.55**	0.59	0.61	0.57	0.75	0.61	0.59	0.59
	CE	0.40	0.15	0.06	0.39	0.27	0.41	**0.41**	0.38
Dowerine	RMSE	8.51	9.39	9.29	9.04	8.47	8.39	**8.37**	8.83
	MAE	4.66	5.45	4.86	4.97	4.83	4.66	**4.55**	5.04
	MASE	0.56	0.66	0.59	0.60	0.58	0.56	**0.55**	0.61
	CE	0.30	0.15	0.17	0.21	0.31	**0.32**	**0.32**	0.25

Table 6.18: Models performance for weekly rainfall predictions in grassland classification zone.

locations; ANN(1) in six locations; ANN(0) in three locations; SVM(RBF) in five locations; SVM(Linear) in one location and CR(EM) in two locations. The MLR and k-NN models are not best in any location.

According to the performance measure RMSE, the CLR(Opt) and CR(EM) models are best in Richmond (Both have same results). The ANN(1) model outperformed other models in six locations(Warren, Newry, Blinman, Annuello, Ningaloo and Dowerine) and ANN(0) model in Alexandria. In these seven locations, the CLR(Opt) model performance is 7.98% lower in Warren; 2.84% in Newry; 6.22% in Alexandria; 0.34% in Blinman; 4.68% in Annuello; 1.06% in Ningaloo and 1.65% in Dowerine.

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 5.2, 6.12, 6.13, 6.14, 6.15, 6.16, 6.17 and 6.18, show that all models follow the series patterns at all locations in grassland classification zone.

In summary, based on the performance measure RMSE, the ANN(1) model is the best model for weekly rainfall predictions in grassland classification zone. The CLR(Opt) and ANN(0) models found to be the second best models in ranking.

6.3 Weekly rainfall predictions in desert zone

In this section, first we present the weekly rainfall prediction results for each model in predicting weekly rainfall with best and worst combinations of input variables in desert classification zone. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models.

Table 6.19, summarizes the prediction performance of the CLR(Opt) model with best and worst combinations of input variables in desert classification zone. according to these results the CLR(Opt) provides best predictions with input variables TMax, TMin, VP and Rad in the location Boulia and with full set of five input meteorological variables in the remaining four locations (Wilcannia, Henbury, Marree and Wiluna). The CLR(Opt) model provides worst predictions in all five locations with input variables TMax, TMin and Rad.

The performance measure RMSE for the CLR(Opt) model in predicting weekly rainfall in desert zone ranges from 6.87 to 41.01, MAE from 3.51 to 29.83, MASE from 0.52 to 1.48 and CE from -0.07 to 0.49. The performance measures RMSE and MAE indicate the model provides best predictions with in the location Marree; MASE indicates in the location Wilcannia and CE indicates in the location Boulia while all four performance measures indicate worst prediction in Wiluna (see Figure 6.19). The graphical display of observed rainfall and CLR(Opt) model predictions over the test period is given in Figure 6.20.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	12.13	4.82	0.52	0.46
Henbury	TMax, TMin, Evap, VP, Rad	10.07	4.72	0.64	0.47
Boulia	TMax, TMin, VP, Rad	8.87	4.22	0.68	0.49
Marree	TMax, TMin, Evap, VP, Rad	6.87	3.51	0.70	0.37
Wiluna	TMax, TMin, Evap, VP, Rad	41.01	29.83	1.48	-0.07
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, Rad	13.27	5.17	0.56	0.35
Henbury	TMax, TMin, Rad	10.60	4.75	0.64	0.41
Boulia	TMax, TMin, Rad	9.36	4.28	0.69	0.43
Marree	TMax, TMin, Rad	7.25	3.42	0.68	0.30
Wiluna	TMax, TMin, Rad	43.30	32.21	1.59	-0.19

Table 6.19: The CLR(Opt) model performance for weekly rainfall prediction with best and worst combinations of input variables in desert zone.

Table 6.20, summarizes the prediction performance of the CR(EM) model with best and worst combinations of input variables in desert classification zone. Results presented in this table show that the model provides best predictions with input variables TMax, TMin and VP in the location Wilcannia; with TMax, TMin, VP and

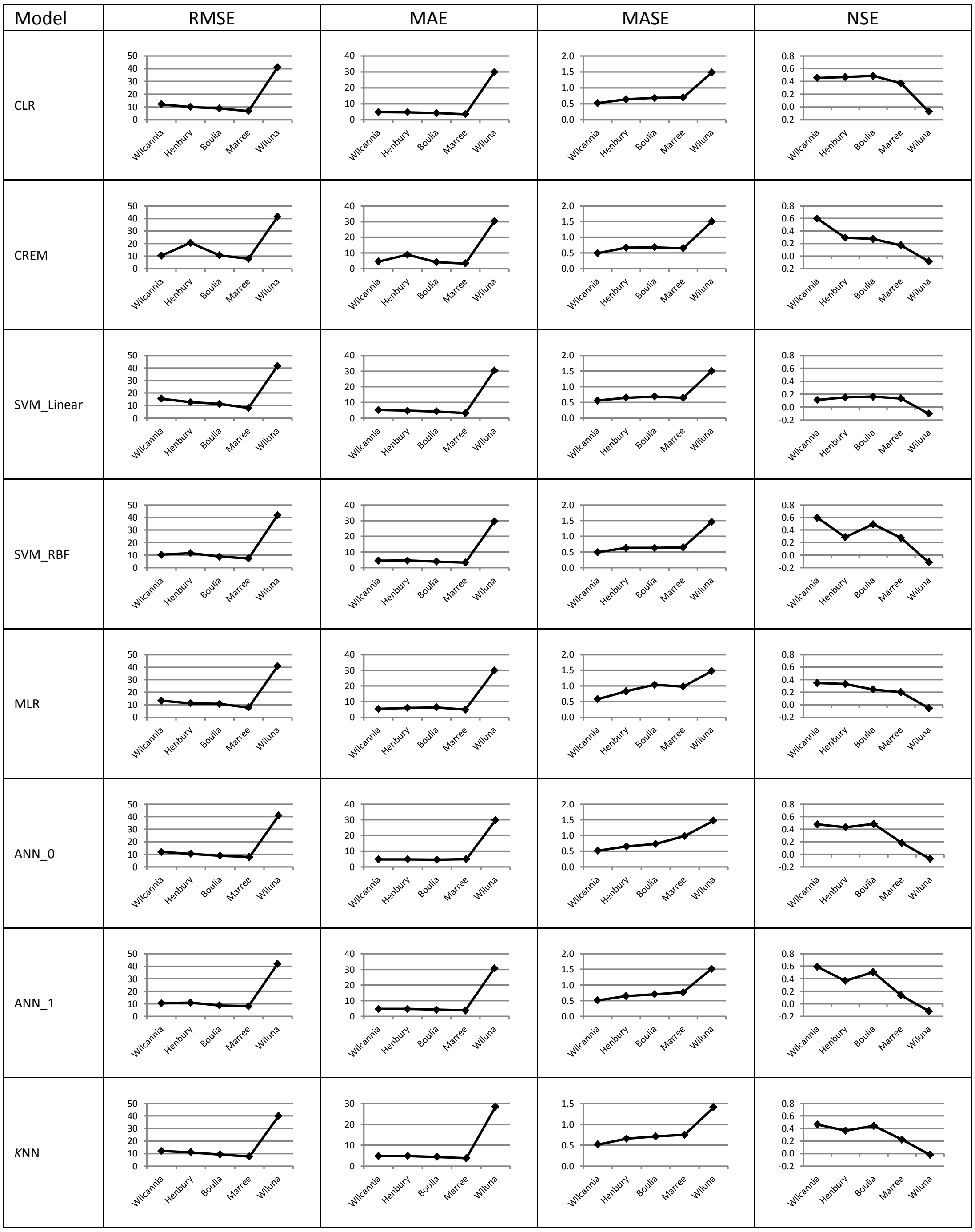

Figure 6.19: Graphical display of models performance measures in predicting weekly rainfall in desert zone.

Rad in two locations (Boulia and Marree) and with full set of five input meteorological variables in the remaining two locations (Henbury and Wiluna). The CR(EM) model provides worst predictions with input variables TMax, TMin and Rad in three

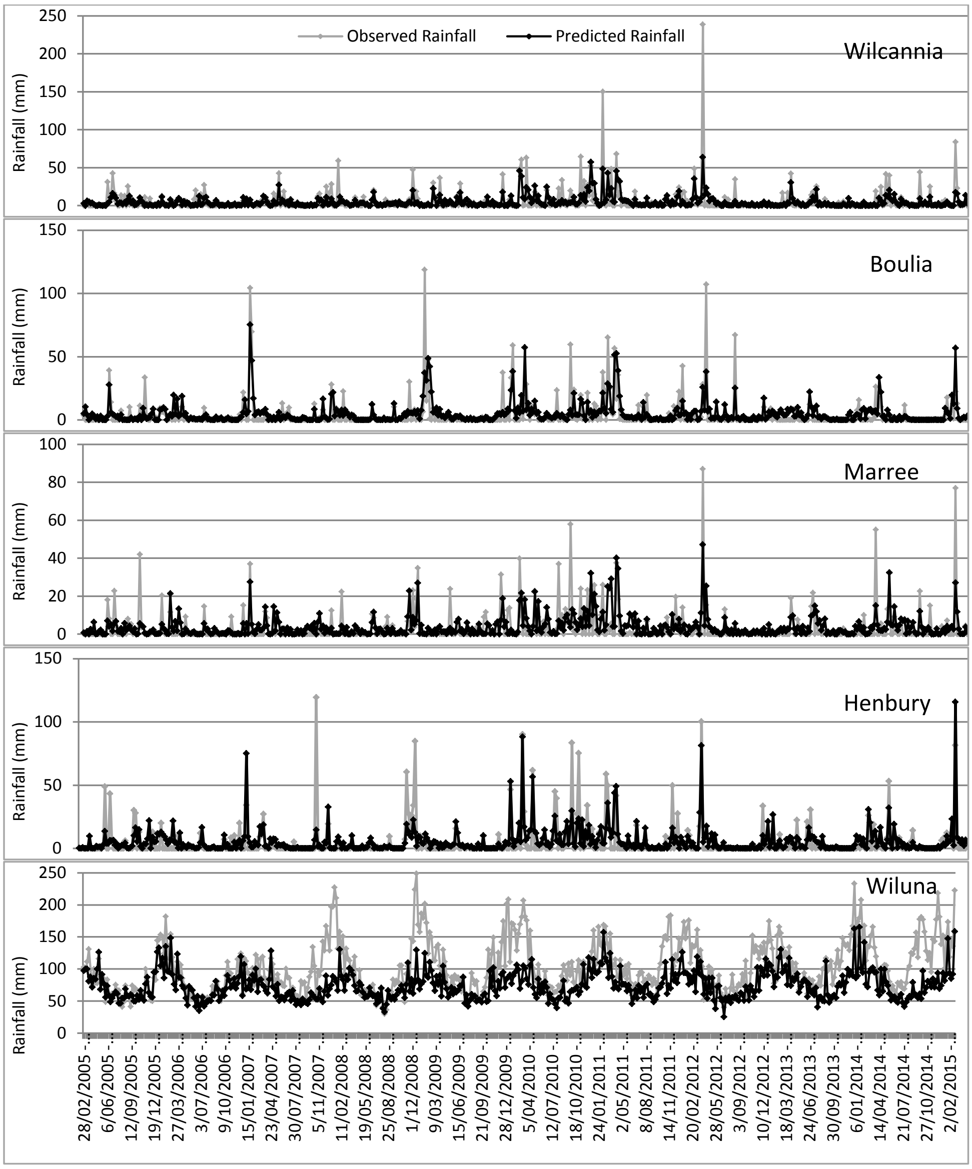

Figure 6.20: Observed rainfall vs. CLR(Opt) model weekly predictions in desert classification zone.

locations (Henbury, Boulia and Marree); with TMax, TMin and VP in the location Wiluna and with TMax, TMin, VP and Rad in the location Wilcannia.

The performance measure RMSE for the CR(EM) model in predicting weekly rainfall ranges from 7.88 to 41.30, MAE from 3.27 to 30.26, MASE from 0.49 to 1.50 and CE from -0.09 to 0.60. The performance measures RMSE and MAE indicate the model provides best predictions in the location Marree and MASE and CE indicate in the location Wilcannia while all four performance measures indicate that the model provides worst predictions in the location Wiluna (see Figure 6.19). The graphical

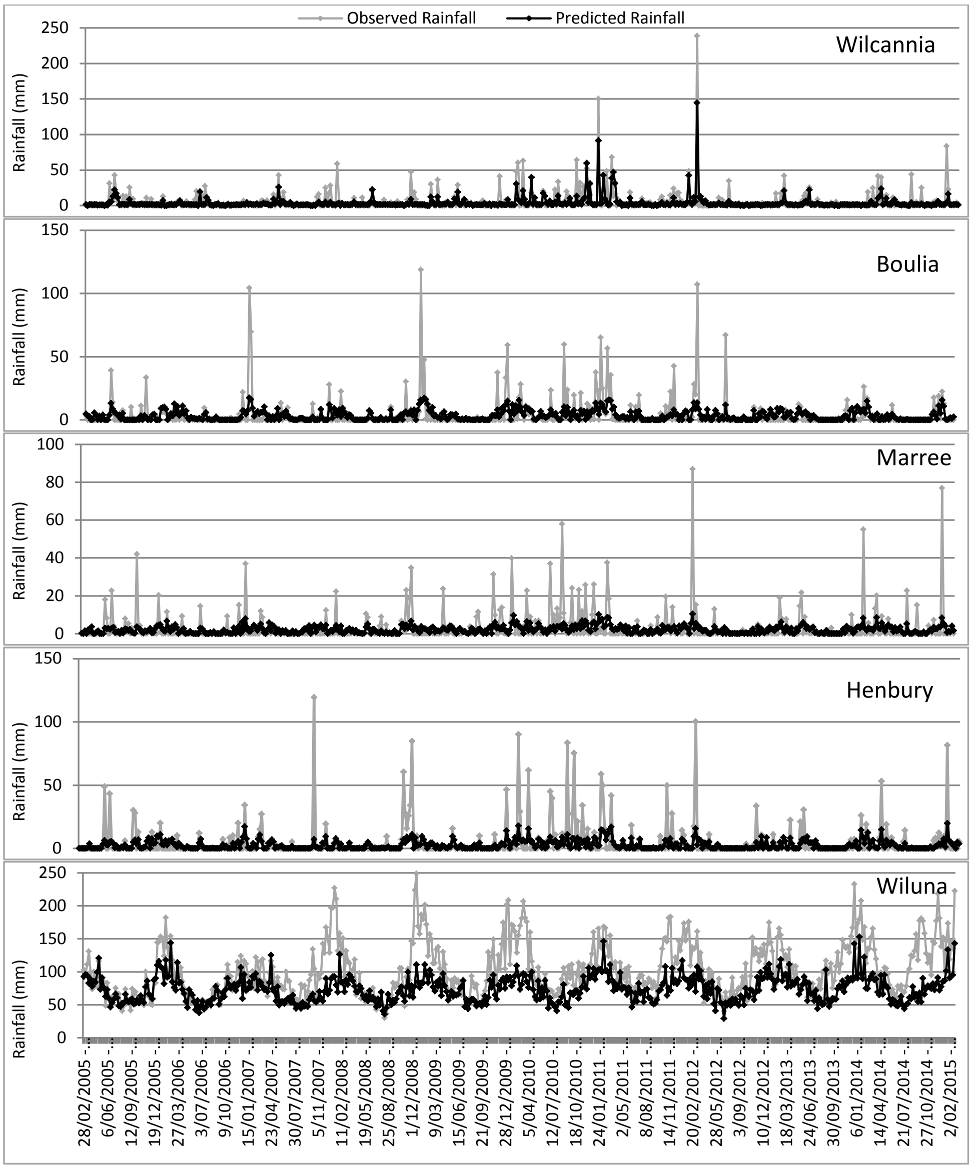

Figure 6.21: Observed rainfall vs. CR(EM) model weekly predictions in desert classification zone.

display of observed rainfall and CR(EM) model predictions over the test period is given in Figure 6.21.

Table 6.21, summarizes the prediction performance of the SVM(Linear) model with best and worst combinations of input variables in desert classification zone. according to these results the SVM(Linear) model provides best prediction with input variables TMax, TMin, VP and Rad in the location Henbury and with full set of five input meteorological variables in the remaining four locations (Wilcannia, Boulia, Marree and Wiluna). The SVM(Linear) model provides the worst predictions with input

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, VP	10.43	4.55	0.49	0.60
Henbury	TMax, TMin, Evap, VP, Rad	20.61	8.95	0.67	0.29
Boulia	TMax, TMin, VP, Rad	10.55	4.18	0.68	0.27
Marree	TMax, TMin, VP, Rad	7.88	3.27	0.65	0.17
Wiluna	TMax, TMin, Evap, VP, Rad	41.30	30.26	1.50	-0.09
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, VP, Rad	12.43	5.14	0.55	0.43
Henbury	TMax, TMin, Rad	21.14	9.33	0.69	0.26
Boulia	TMax, TMin, Rad	11.06	4.60	0.75	0.20
Marree	TMax, TMin, Rad	8.08	3.41	0.68	0.13
Wiluna	TMax, TMin, VP	42.80	31.45	1.56	-0.17

Table 6.20: The CR(EM) model prediction performance for weekly rainfall with best and worst combinations of input variables in desert classification zone.

variables TMax, TMin and Rad in four locations (Wilcannia, Henbury, Boulia and Marree) and with TMax, TMin and VP in the remaining fifth location Wiluna.

The performance measure RMSE for the SVM(Linear) model in predicting weekly rainfall ranges from 8.06 to 41.61, MAE from 3.24 to 30.34, MASE from 0.56 to 1.50 and CE from -0.10 to 0.16. The performance measures RMSE and MAE indicate the model provides best predictions in the location Marree; MASE indicates in the location Wilcannia and CE indicates in the location Boulia while all four performance measures indicate that the model provide worst predictions in the location Wiluna (see Figure 6.19). The graphical display of observed rainfall and SVM(Linear) model predictions over the test period is given in Figure 6.22.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	15.50	5.22	0.56	0.11
Henbury	TMax, TMin, VP, Rad	12.73	4.76	0.65	0.15
Boulia	TMax, TMin, Evap, VP, Rad	11.33	4.23	0.69	0.16
Marree	TMax, TMin, Evap, VP, Rad	8.06	3.24	0.64	0.13
Wiluna	TMax, TMin, Evap, VP, Rad	41.61	30.34	1.50	-0.10
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, Rad	15.94	5.37	0.58	0.06
Henbury	TMax, TMin, Rad	13.26	4.87	0.66	0.08
Boulia	TMax, TMin, Rad	12.03	4.30	0.70	0.06
Marree	TMax, TMin, Rad	8.38	3.24	0.64	0.06
Wiluna	TMax, TMin, VP	43.78	32.24	1.60	-0.22

Table 6.21: The SVM(Linear) model performance for weekly rainfall predictions with best and worst combinations of input variables in desert zone.

Table 6.22, summarizes the performance of the SVM(RBF) model in predicting weekly rainfall with best and worst combinations of input variables in desert classifica-

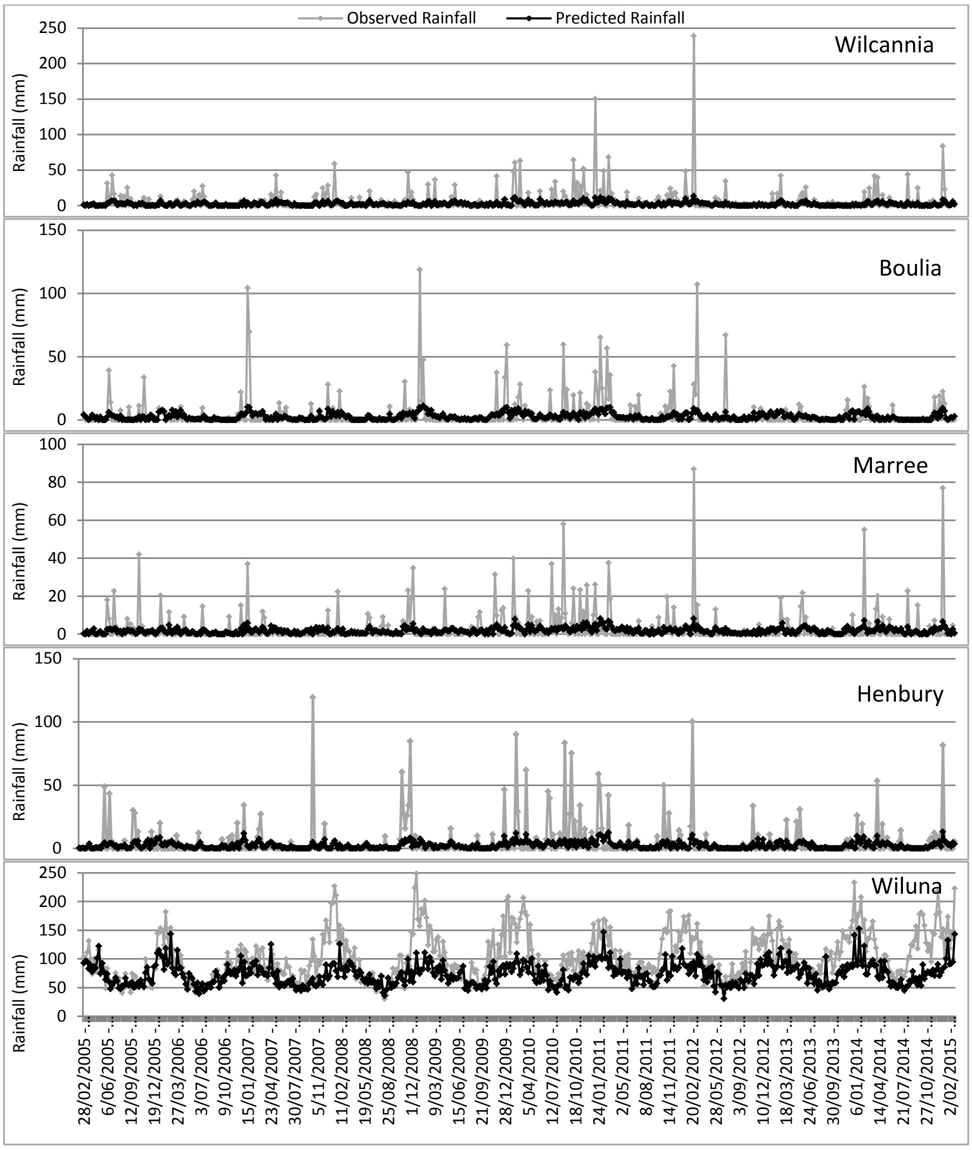

Figure 6.22: Observed rainfall vs. SVM(Linear) model weekly predictions in desert classification zone.

tion zone. according to these results the SVM(RBF) model provides best predictions with input variables TMax, TMin and VP in three locations (Wilcannia, Henbury and Boulia); with TMax, TMin and Rad in the location Marree and with TMax, TMin, VP and Rad in the location Wiluna. The SVM(RBF) model provides the worst predictions with input variables TMax, TMin, VP and Rad in the location Wilcannia and with full set of input variables in the remaining four locations (Henbury, Boulia, Marree and Wiluna).

The performance measure RMSE for the SVM(RBF) model in predicting weekly

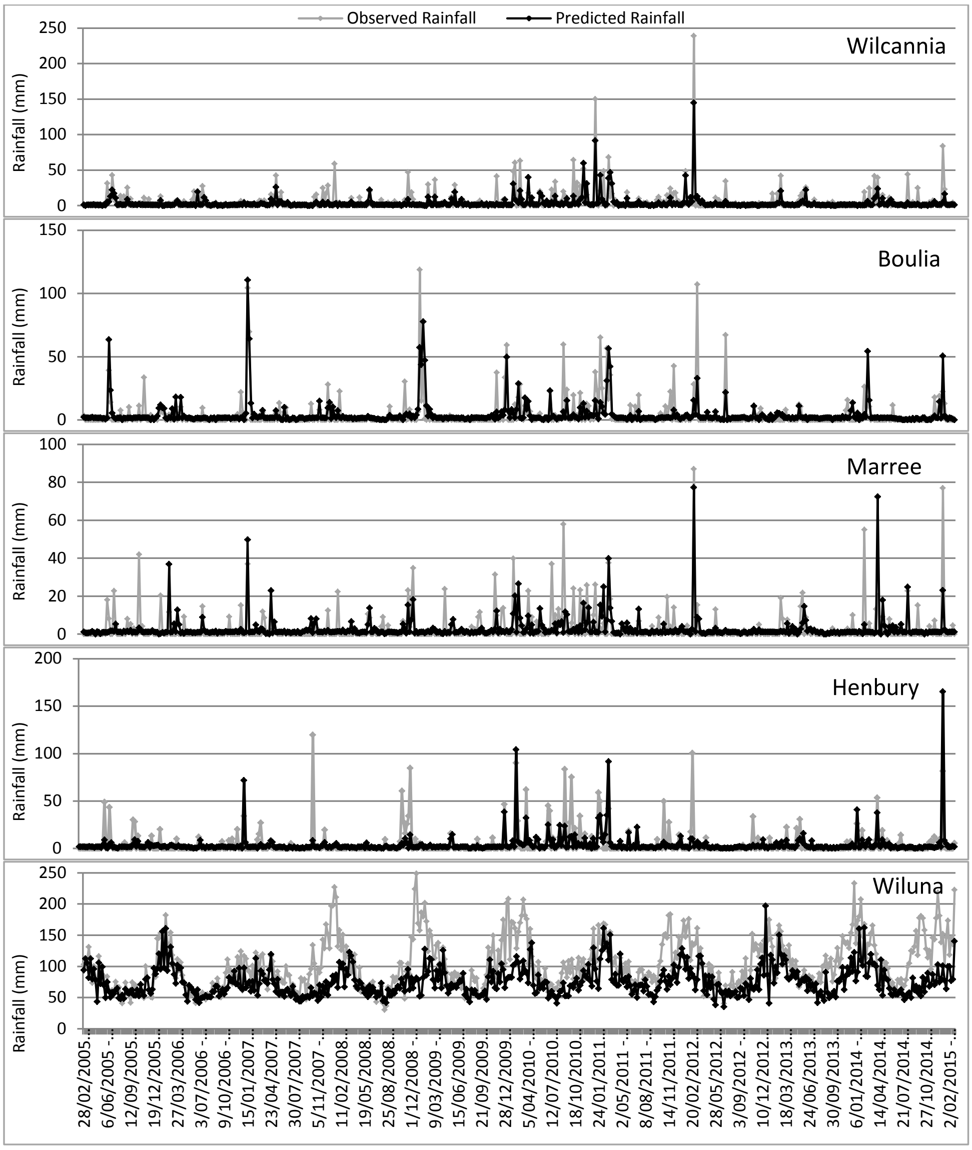

Figure 6.23: Observed rainfall vs. SVM(RBF) model weekly rainfall predictions in desert classification zone.

rainfall ranges from 7.37 to 41.84, MAE from 3.28 to 29.51, MASE from 0.49 to 1.46 and CE from -0.12 to 0.60. The performance measures RMSE and MAE indicate the SVM(RBF) model provides best predictions with in the location Marree; MASE and CE indicate in the location Wilcannia while all four performance measures indicate the model provides worst predictions in the location Wiluna (see Figure 6.19). The graphical display of observed rainfall and SVM(RBF) model predictions over the test period is given in Figure 6.23.

Table 6.23, summarizes the prediction performance of the MLR model with best

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, VP	10.43	4.55	0.49	0.60
Henbury	TMax, TMin, VP	11.67	4.64	0.63	0.29
Boulia	TMax, TMin, VP	8.82	3.91	0.63	0.49
Marree	TMax, TMin, Rad	7.37	3.28	0.65	0.27
Wiluna	TMax, TMin, VP, Rad	41.84	29.51	1.46	-0.12
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, VP, Rad	12.43	5.14	0.55	0.43
Henbury	TMax, TMin, Evap, VP, Rad	12.57	4.94	0.67	0.17
Boulia	TMax, TMin, Evap, VP, Rad	13.85	6.61	1.07	-0.25
Marree	TMax, TMin, Evap, VP, Rad	10.04	4.67	0.93	-0.35
Wiluna	TMax, TMin, Evap, VP, Rad	43.66	30.81	1.53	-0.21

Table 6.22: The SVM(RBF) model performance for weekly rainfall predictions with best and worst combinations of input variables in desert zone.

and worst combinations of input variables in desert classification zone. The results presented in this table show that the MLR model provides the best predictions with input variables TMax, TMin and VP in three locations (Henbury, Boulia and Marree); with TMax, TMin, VP and Rad in the location Wilcannia and with full set of five input meteorological parameters in the location Wiluna. The MLR model provides the worst predictions with input variables TMax, TMin and Rad in three locations (Wilcannia, Henbury and Marree); with TMax, TMin and VP in the location Wiluna and with full set of five meteorological variables in the remaining location Boulia.

The performance measure RMSE for the MLR model in predicting weekly rainfall ranges from 7.74 to 40.72, MAE from 4.93 to 29.81, MASE from 0.58 to 1.48 and CE from -0.06 to 0.35. The performance measures RMSE and MAE indicate the MLR model provides best predictions in the location Marree and MASE and CE indicate in the location Wilcannia while all four performance measures indicate that the model provides worst predictions in the location Wiluna (see Figure 6.19). The graphical display of observed rainfall and the MLR model predictions over the test period is given in Figure 6.24.

Table 6.24, summarizes the prediction performance of the ANN(0) model with best and worst combinations of input variables in desert classification zone. These results show that the ANN(0) model provides best predictions with input variables TMax, TMin and VP in two locations (Wilcannia and Boulia); with TMax, TMin, VP and Rad in the location Wiluna and with full set of five input variables in the remaining two locations (Henbury and Marree). The ANN(0) model provides worst predictions with input variables TMax, TMin, Evap, VP and Rad in the location Wilcannia; with TMax, TMin and VP in two locations (Henbury and Wiluna) and with TMax, TMin and Rad in the remaining two locations (Boulia and Marree).

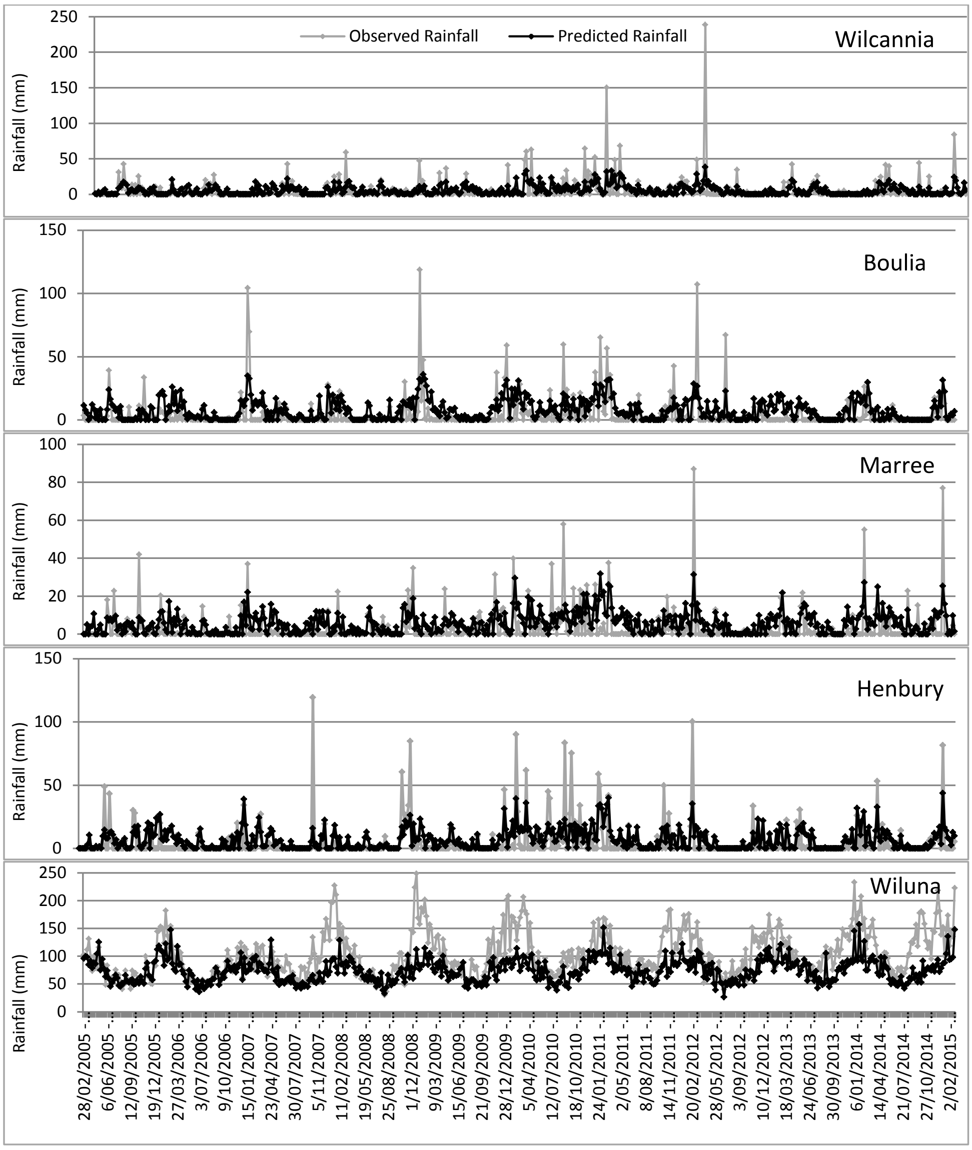

Figure 6.24: Observed rainfall vs. MLR model weekly predictions in desert classification zone.

The performance measure RMSE for the ANN(0) model in predicting weekly rainfall ranges from 7.83 to 41.03, MAE from 4.53 to 29.81, MASE from 0.52 to 1.48 and CE from -0.07 to 0.49. The performance measures RMSE indicates the ANN(0) model provides best predictions in the location Marree; MAE indicates in the location Boulia; MASE indicates in the location Wilcannia and CE indicates in the location Boulia while all four performance measures indicate that the model provides worst predictions in the location Wiluna (see Figure 6.19). The graphical display of observed rainfall and ANN(0) model predictions over the test period is given in Figure 6.25.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, VP, Rad	13.29	5.39	0.58	0.35
Henbury	TMax, TMin, VP	11.29	6.10	0.83	0.33
Boulia	TMax, TMin, VP	10.77	6.39	1.04	0.24
Marree	TMax, TMin, VP	7.74	4.93	0.98	0.20
Wiluna	TMax, TMin, Evap, VP, Rad	40.72	29.81	1.48	-0.06
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, Rad	14.09	5.73	0.62	0.27
Henbury	TMax, TMin, Rad	11.82	6.50	0.88	0.27
Boulia	TMax, TMin, Evap, VP, Rad	11.13	6.95	1.13	0.19
Marree	TMax, TMin, Rad	7.84	4.57	0.91	0.18
Wiluna	TMax, TMin, VP	42.65	31.41	1.56	-0.16

Table 6.23: The MLR model performance for weekly rainfall predictions with best and worst combinations of input variables in desert zone.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, VP	11.89	4.81	0.52	0.48
Henbury	TMax, TMin, Evap, VP, Rad	10.41	4.80	0.65	0.43
Boulia	TMax, TMin, VP	8.88	4.53	0.74	0.49
Marree	TMax, TMin, Evap, VP, Rad	7.83	4.97	0.99	0.18
Wiluna	TMax, TMin, VP, Rad	41.03	29.81	1.48	-0.07
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, Evap, Evap, VP, Rad	13.13	5.60	0.60	0.36
Henbury	TMax, TMin, VP	11.82	5.54	0.75	0.27
Boulia	TMax, TMin, Rad	10.78	5.28	0.86	0.24
Marree	TMax, TMin, Rad	9.17	4.16	0.83	-0.12
Wiluna	TMax, TMin, VP	42.54	31.15	1.54	-0.15

Table 6.24: The ANN(0) model performance for weekly rainfall predictions with best and worst combinations of input variables in desert zone.

Table 6.25, summarizes the performance of the ANN(1) model in predicting weekly rainfall with best and worst combinations of input variables in desert classification zone. according to these results the ANN(1) model provides best predictions with input variables TMax, TMin and VP in two locations (Henbury and Boulia); with TMax, TMin and Rad in the location Wiluna and with TMax, TMin, VP and Rad in the remaining two locations (Wilcannia and Marree). The ANN(1) model provides worst predictions with input variables TMax, TMin and VP in the location Wilcannia; with TMax, TMin and Rad in two locations (Henbury and Boulia) and with a full set of five input variables in the remaining two locations (Marree and Wiluna).

The performance measure RMSE for the ANN(1) model in predicting weekly rainfall ranges from 8.04 to 41.87, MAE from 3.87 to 30.68, MASE from 0.51 to 1.52 and CE from -0.12 to 0.59. The performance measures RMSE and MAE indicate the

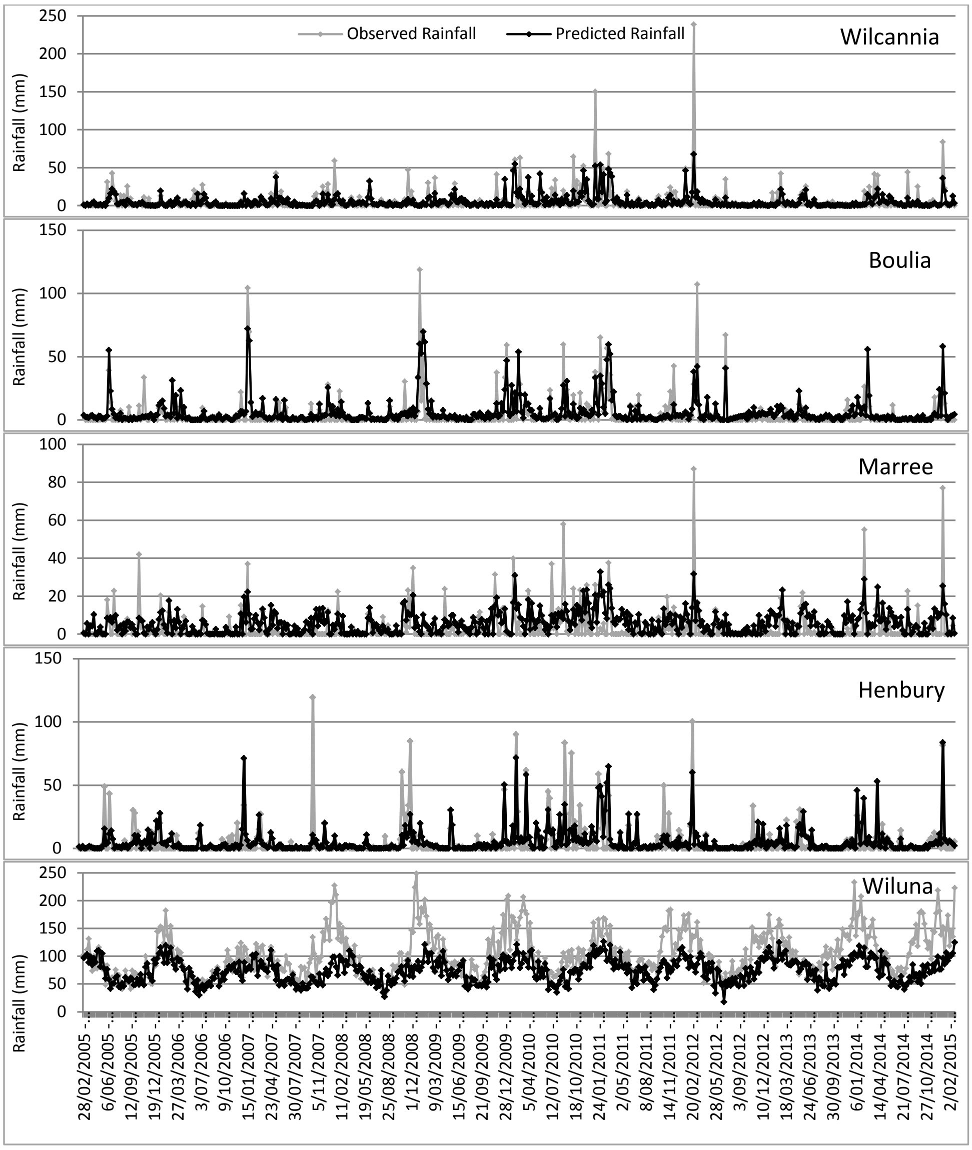

Figure 6.25: Observed rainfall vs. ANN(0) model weekly predictions in desert classification zone.

ANN(1) model provides best predictions in the location Marree and MASE and CE in the location Wilcannia while all four performance measures indicate the model provides worst predictions in the location Wiluna (see Figure 6.19). The graphical display of observed rainfall and ANN(1) model predictions over the test period is given in Figure 6.26.

Table 6.26, summarizes the prediction performance of the k-NN model with best and worst combinations of input variables in desert classification zone. According to these results, the k-NN model provides best predictions with input variables TMax,

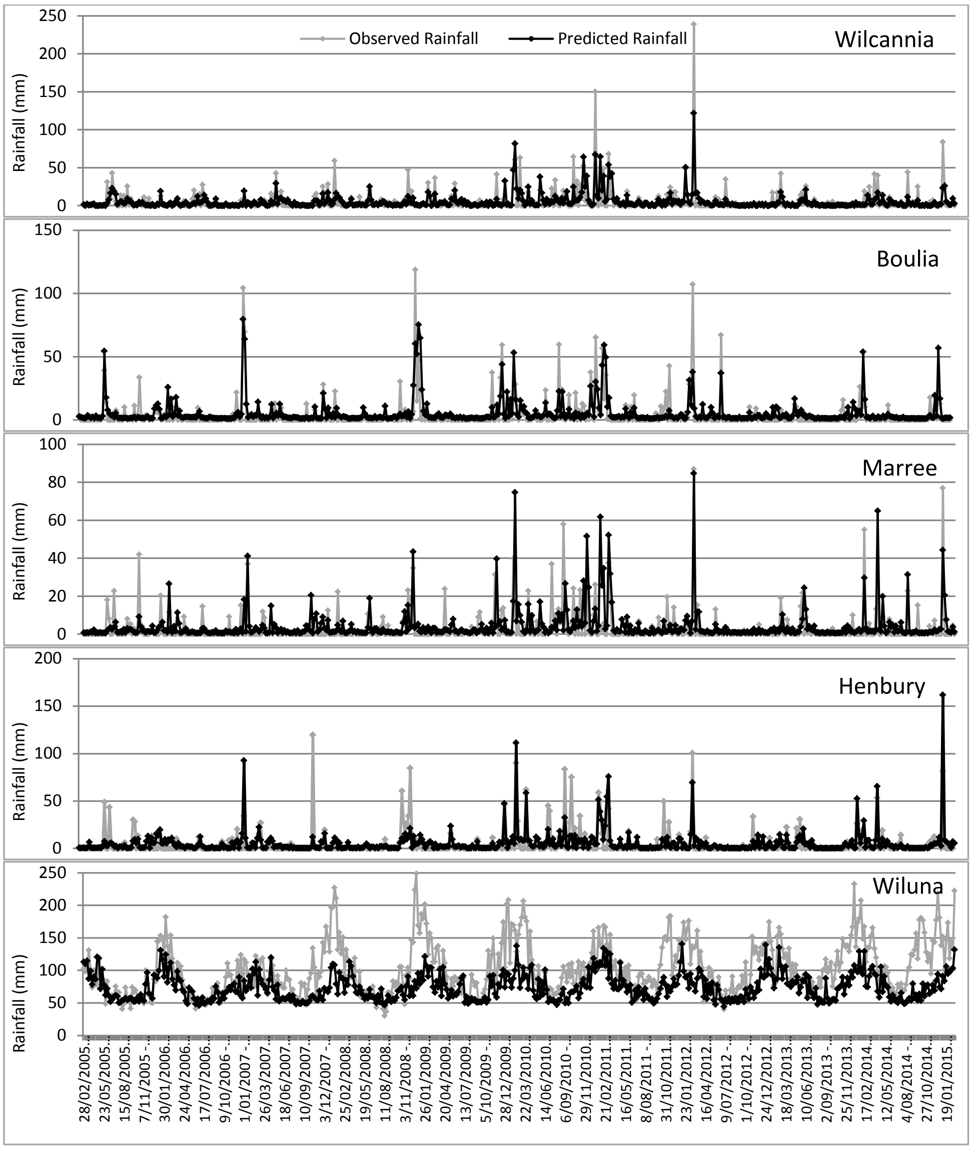

Figure 6.26: Observed rainfall vs. ANN(1) model weekly predictions in desert classification zone.

TMin and Rad provides the in the location Marree; with TMax, TMin, VP and Rad in the location Boulia and with a full set of five input variables in the remaining three locations (Wilcannia, Henbury and Wiluna). The k-NN model provides worst predictions with input variables TMax, TMin and Rad in four locations (Wilcannia, Henbury, Boulia and Wiluna) and with TMax, TMin and VP in the remaining fifth location Marree.

The performance measure RMSE for the k-NN model in predicting weekly rainfall ranges from 7.63 to 40.03, MAE from 3.78 to 28.42, MASE from 0.51 to 1.41 and CE

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, VP, Rad	10.52	4.75	0.51	0.59
Henbury	TMax, TMin, VP	10.98	4.77	0.65	0.37
Boulia	TMax, TMin, VP	8.70	4.32	0.70	0.51
Marree	TMax, TMin, VP, Rad	8.04	3.87	0.77	0.14
Wiluna	TMax, TMin, Rad	41.87	30.68	1.52	-0.12
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, VP	14.06	5.35	0.57	0.27
Henbury	TMax, TMin, Rad	13.39	6.86	0.93	0.06
Boulia	TMax, TMin, Rad	12.94	6.60	1.07	-0.09
Marree	TMax, TMin, Evap, VP, Rad	9.59	5.01	0.99	-0.23
Wiluna	TMax, TMin, Evap, VP, Rad	48.55	35.45	1.76	-0.50

Table 6.25: The ANN(1) model performance for weekly rainfall predictions with best and worst combinations of input variables in desert zone.

from -0.02 to 0.46. The performance measures RMSE and MAE indicate the k-NN model provides best predictions in the location Marree, and MASE and CE indicate in the location Wilcannia while all four performance measures indicate that the model provides worst predictions in the location Wiluna (see Figure 6.19). The graphical display of observed rainfall and k-NN model predictions over the test period is given in Figure 6.27.

Stations	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations					
Wilcannia	TMax, TMin, Evap, VP, Rad	12.03	4.79	0.51	0.46
Henbury	TMax, TMin, Evap, VP, Rad	11.01	4.84	0.66	0.36
Boulia	TMax, TMin, VP, Rad	9.24	4.37	0.71	0.44
Marree	TMax, TMin, Rad	7.63	3.78	0.75	0.22
Wiluna	TMax, TMin, Evap, VP, Rad	40.03	28.42	1.41	-0.02
Performance measures for worst input combinations					
Wilcannia	TMax, TMin, Rad	13.40	5.14	0.55	0.34
Henbury	TMax, TMin, Rad	11.36	5.16	0.70	0.32
Boulia	TMax, TMin, Rad	10.05	4.90	0.79	0.34
Marree	TMax, TMin, VP	8.72	4.35	0.86	-0.02
Wiluna	TMax, TMin, Rad	42.39	30.48	1.51	-0.14

Table 6.26: The k-NN model performance for weekly rainfall predictions with best and worst combinations of input variables in desert zone.

Table 6.27, summarizes the performance of all eight models in predicting weekly rainfall with best combinations of input variables in desert classification zone. Best results among all models are highlighted in bold. According to these results, at least one performance measure indicate that the CLR(Opt) is best in two out of five locations (Henbury and Marree); SVM(RBF) in three locations (Wilcannia, Henbury and Boulia); CR(EM) in Wilcannia, SVM(Linear) in Marree, ANN(1) in Boulia and

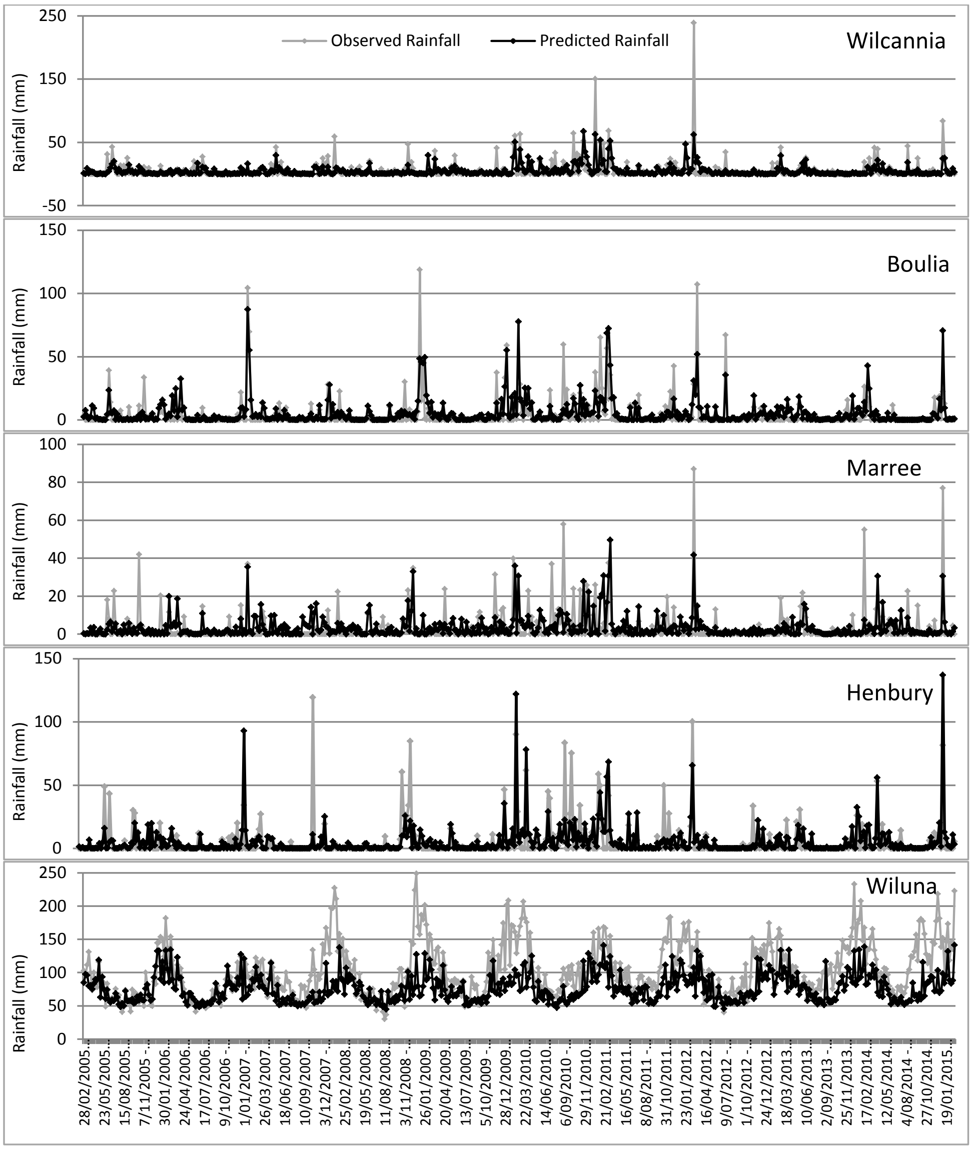

Figure 6.27: Observed rainfall vs. k-NN model weekly predictions in desert classification zone.

k-NN in Wiluna while MLR and ANN(0) modesl are not best in any location.

According to the performance measures RMSE, the CLR(Opt) model outperformed other models in two locations (Henbury and Marree); CR(EM) and SVM(RBF) both models in Wilcannia (Both have Same results); ANN(1) model in Boulia and k-NN model in Wiluna. In Wilcannia, the CLR(Opt) model performance is 14.01% lower than the outperformed model. Similarly, in Boulia 1.92% and in Wiluna 2.39% lower than the outperformed model.

A visual comparison of model predictions with the actual observations in temperate

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	MLR	ANN (0)	ANN (1)	KNN
Wilcannia	RMSE	12.13	**10.43**	15.50	**10.43**	13.29	11.89	10.52	12.03
	MAE	4.82	**4.55**	5.22	**4.55**	5.39	4.81	4.75	4.79
	MASE	0.52	**0.49**	0.56	**0.49**	0.58	0.52	0.51	0.51
	CE	0.46	**0.60**	0.11	**0.60**	0.35	0.48	0.59	0.46
Henbury	RMSE	**10.07**	20.61	12.73	11.67	11.29	10.41	10.98	11.01
	MAE	4.72	8.95	4.76	**4.64**	6.10	4.80	4.77	4.84
	MASE	0.64	0.67	0.65	**0.63**	0.83	0.65	0.65	0.66
	CE	**0.47**	0.29	0.15	0.29	0.33	0.43	0.37	0.36
Boulia	RMSE	8.87	10.55	11.33	8.82	10.77	8.88	**8.70**	9.24
	MAE	4.22	4.18	4.23	**3.91**	6.39	4.53	4.32	4.37
	MASE	0.68	0.68	0.69	**0.63**	1.04	0.74	0.70	0.71
	CE	0.49	0.27	0.16	0.49	0.24	0.49	**0.51**	0.44
Marree	RMSE	**6.87**	7.88	8.06	7.37	7.74	7.83	8.04	7.63
	MAE	3.51	3.27	**3.24**	3.28	4.93	4.97	3.87	3.78
	MASE	0.70	0.65	**0.64**	0.65	0.98	0.99	0.77	0.75
	CE	**0.37**	0.17	0.13	0.27	0.20	0.18	0.14	0.22
Wiluna	RMSE	41.01	41.30	41.61	41.84	40.72	41.03	41.87	**40.03**
	MAE	29.83	30.26	30.34	29.51	29.81	29.81	30.68	**28.42**
	MASE	1.48	1.50	1.50	1.46	1.48	1.48	1.52	**1.41**
	CE	-0.07	-0.09	-0.10	-0.12	-0.06	-0.07	-0.12	**-0.02**

Table 6.27: Models performance for weekly rainfall predictions in desert classification zone.

classification zone, given in Figures 5.2, 6.21, 6.22, 6.23, 6.24, 6.25, 6.26 and 6.27, show that all models follow the series patterns in all locations of desert classification zone.

In summary , based on our primary performance measure RMSE, the CLR(Opt) model is the the most suitable models in finding the pattern and trends of the observations compared to other models at most locations of desert classification zone.

6.4 Weekly rainfall predictions in tropical and subtropical zones

In this section, first we present the weekly rainfall prediction results for each model in predicting weekly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. Then we summarize the performance of all models with best combination of input variables. Finally we compare the CLR(Opt) model performance with other models.

Table 6.28, summarizes the performance of the CLR(Opt) model for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical classification zones. According to these results the CLR(Opt) model provides the best predictions with input variables TMax, TMin and Rad in Palmerville and Yamba and with full set of five input variables in the remaining locations Katherine and Fairymead. The model provides worst predictions with input variables TMax, TMin and Rad in the location Katherine; with TMax, TMin and VP in two locations Yamba and Fairymead and with TMax, TMin, VP and Rad in the location Palmerville.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP, Rad	27.58	13.71	0.62	0.58
Palmerville	Tropical	TMax, TMin, Rad	32.37	14.37	0.62	0.49
Yamba	Subtropical	TMax, TMin, Rad	36.99	21.77	0.59	0.37
Fairymead	Subtropical	TMax, TMin, Evap, VP, Rad	45.18	17.43	0.57	0.29
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	30.38	14.93	0.68	0.49
Palmerville	Tropical	TMax, TMin, VP, Rad	34.20	16.82	0.73	0.43
Yamba	Subtropical	TMax, TMin, VP	40.35	22.15	0.60	0.25
Fairymead	Subtropical	TMax, TMin, VP	46.94	17.22	0.56	0.23

Table 6.28: The CLR(Opt) model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP	36.71	17.24	0.78	0.26
Palmerville	Tropical	TMax, TMin, VP, Rad	38.62	17.77	0.77	0.27
Yamba	Subtropical	TMax, TMin, Rad	38.57	20.73	0.56	0.32
Fairymead	Subtropical	TMax, TMin, Rad	50.17	18.68	0.61	0.13
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	37.65	18.94	0.86	0.22
Palmerville	Tropical	TMax, TMin, Rad	39.28	16.51	0.71	0.25
Yamba	Subtropical	TMax, TMin, Evap, VP, Rad	43.26	24.11	0.65	0.14
Fairymead	Subtropical	TMax, TMin, VP	50.46	18.68	0.61	0.12

Table 6.29: The CR(EM) model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

The performance measure RMSE for the CLR(Opt) model in predicting weekly rainfall ranges from 27.58 to 45.18, MAE from 13.71 to 21.77, MASE from 0.57 to 0.62 and CE from 0.29 to 0.58. The performance measures RMSE, MAE and CE indicates that the CLR(Opt) model provides best predictions in the location Katherine and MASE indicates in the location Fairymead while the performance measures RMSE and CE indicates worst predictions in the location Fairymead; MAE in the location Yamba and MASE in the location Katherine (see Figure 6.28). The graphical display of observed rainfall and CLR(Opt) model predictions over the test period is given in Figure 6.29.

Table 6.29, summarizes the performance of the CR(EM) model in predicting weekly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. These results show that the CR(EM) model provides best predictions with input variables TMax, TMin and Rad in two locations (Yamba and Fairymead); with TMax, TMin and VP in the location Katherine and TMax, TMin, VP and Rad in the location Palmerville. The model provides worst predictions with input variables TMax, TMin and Rad in two locations (Katherine and Palmerville); with TMax, TMin and VP in the location Fairymead and with TMax, TMin, Evap, VP and Rad in the location Yamba.

The performance measure RMSE for the CR(EM) model in predicting weekly rainfall ranges from 36.71 to 50.17, MAE from 17.24 to 18.68, MASE from 0.56 to 0.78 and CE from 0.13 to 0.32. The performance measures RMSE and MAE indicate that the CR(EM) model provides best predictions in the location Katherine and MASE and CE indicates in the location Yamba while the performance measure RMSE and CE indicate worst predictions in the location Fairymead; MAE in the location Yamba and MASE in the location Katherine (see Figure 6.28). The graphical display of observed rainfall and CR(EM) model predictions over the test period is given in Figure 6.30.

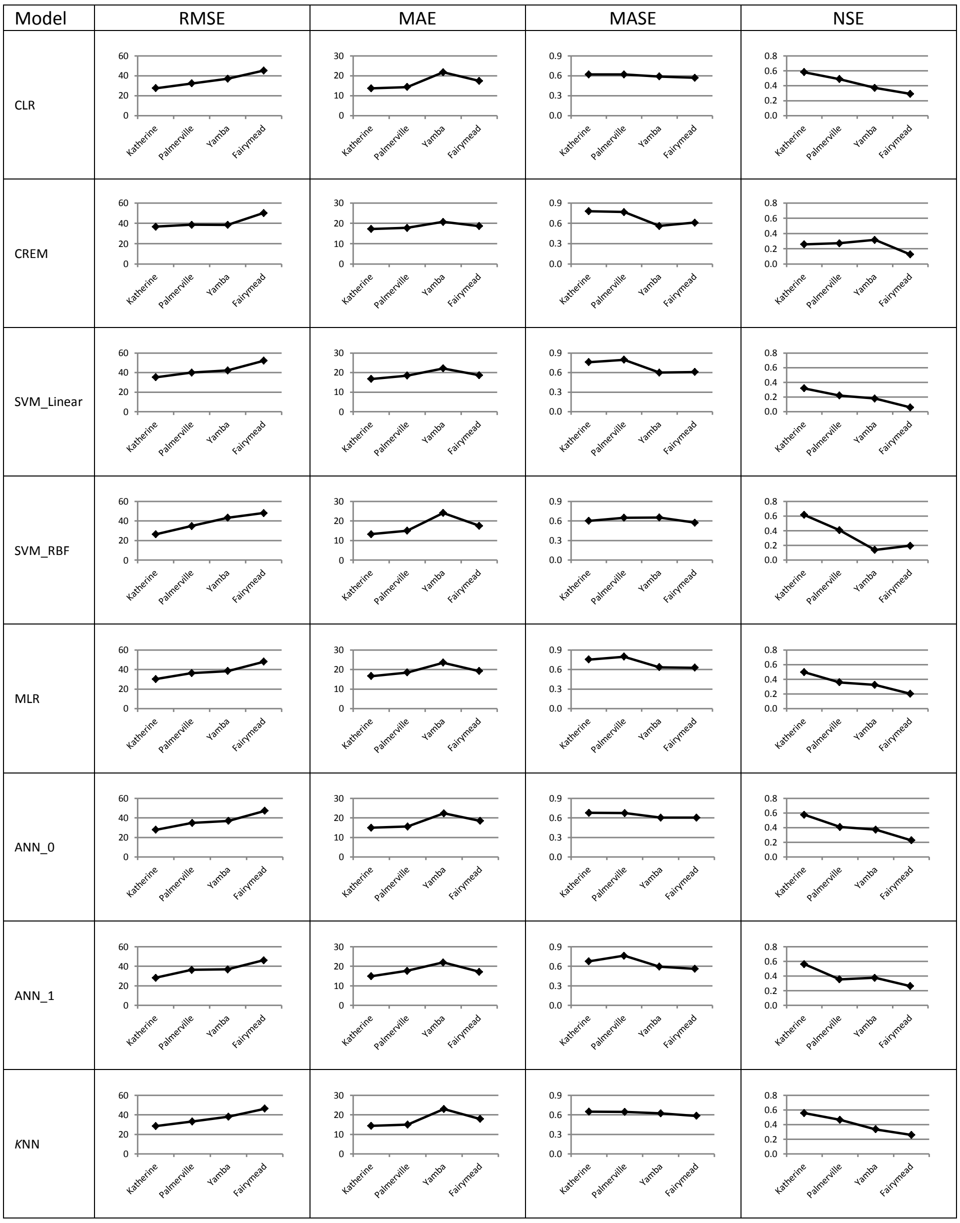

Figure 6.28: Illustration of models performance measures in predicting weekly rainfall in tropical and subtropical zones.

Table 6.30, summarizes the performance of the SVM(Linear) model in predicting weekly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. According to these results the SVM(Linear) model provides best predictions with input variables TMax, TMin and VP in the location Katherine and with a full set of five input variables in the remaining three locations (Palmerville, Yamba and Fairymead). The model provides worst predictions with input variables TMax, TMin and Rad in three locations (Katherine, Palmerville and

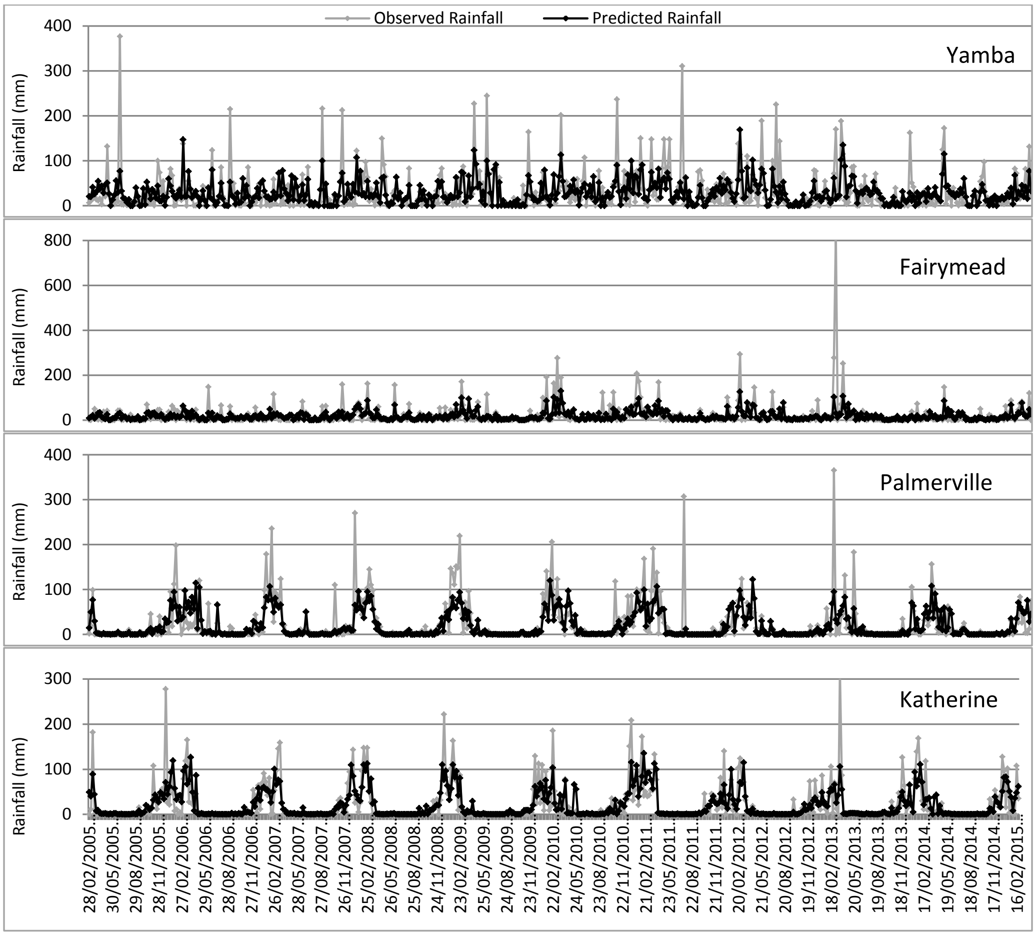

Figure 6.29: Observed rainfall vs. CLR(Opt) model weekly predictions in tropical and subtropical classification zones.

Fairymead) and with TMax, TMin and VP in the location Yamba.

The performance measure RMSE for the SVM(Linear) model in predicting weekly rainfall ranges from 35.18 to 52.08, MAE from 16.77 to 22.14, MASE from 0.60 to 0.80 and CE from 0.06 to 0.32. The performance measures RMSE, MAE and CE indicate that the SVM(Linear) model provides best predictions in the location Katherine and MASE indicates in the location Yamba while the performance measure RMSE and CE indicate worst predictions in the location Fairymead; MAE in the location Yamba and MASE in the location Palmerville (see Figure 6.28). The graphical display of observed rainfall and SVM(Linear) model predictions over the test period is given in Figure 6.31.

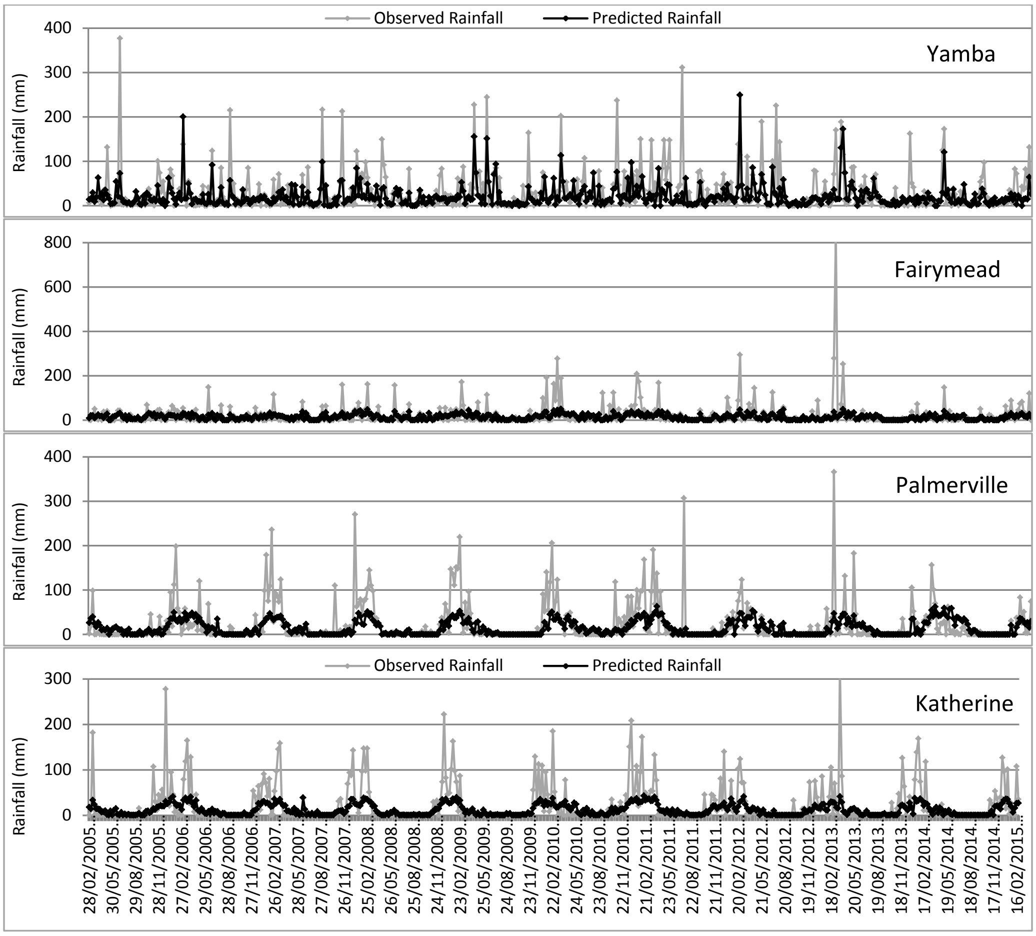

Figure 6.30: Observed rainfall vs. CR(EM) model weekly predictions in tropical and subtropical classification zones.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP	35.18	16.77	0.76	0.32
Palmerville	Tropical	TMax, TMin, Evap,VP, Rad	39.99	18.47	0.80	0.22
Yamba	Subtropical	TMax, TMin, Evap,VP, Rad	42.22	22.14	0.60	0.18
Fairymead	Subtropical	TMax, TMin, Evap,VP, Rad	52.08	18.64	0.61	0.06
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	37.32	17.35	0.79	0.23
Palmerville	Tropical	TMax, TMin, Rad	42.20	18.33	0.79	0.13
Yamba	Subtropical	TMax, TMin, VP	44.59	22.96	0.62	0.09
Fairymead	Subtropical	TMax, TMin, Rad	52.54	18.99	0.62	0.04

Table 6.30: The SVM(Linear) model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

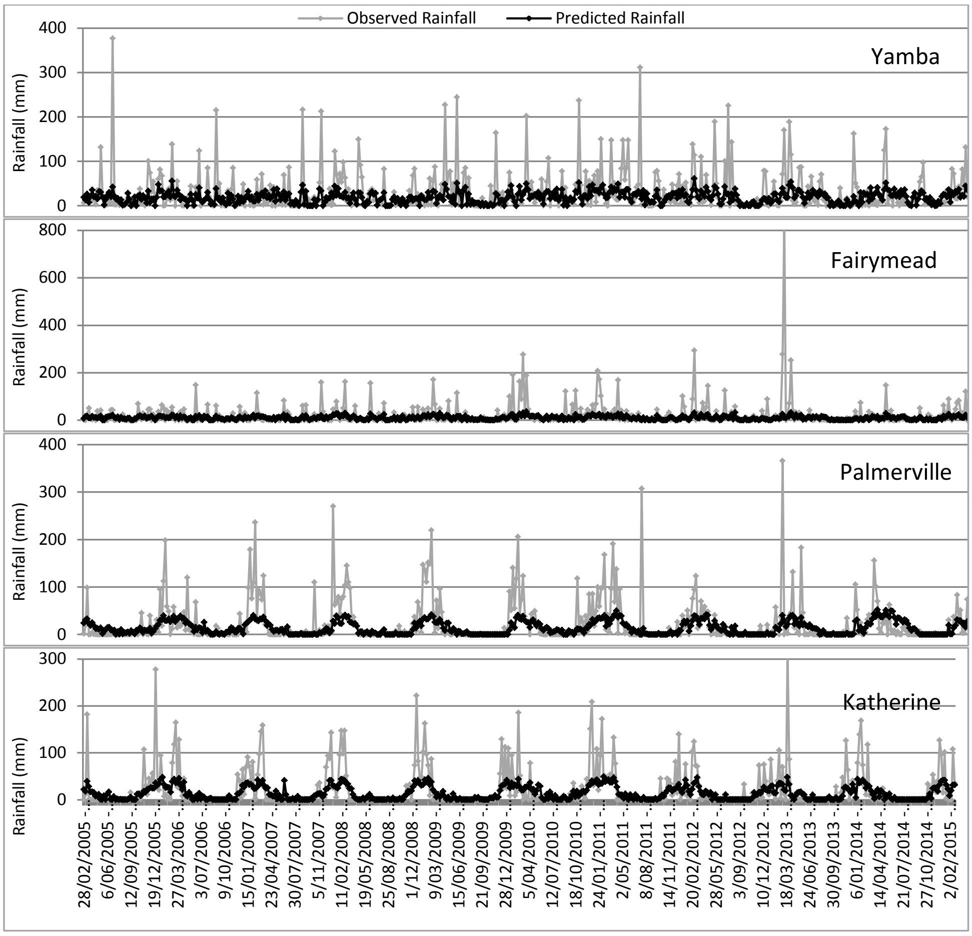

Figure 6.31: Observed rainfall vs. SVM(Linear) model weekly predictions in tropical and subtropical classification zones.

Table 6.31, summarizes the performance of the SVM(RBF) model in predicting weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical classification zones. According to the results, the SVM(RBF) model provides best predictions with input variables TMax, TMin and VP in two locations (Katherine and Fairymead) and with TMax, TMin and Rad in the remaining two locations (Palmerville and Yamba). The model provides worst predictions with full set of input variables in all four locations.

The performance measure RMSE for the SVM(RBF) model in predicting weekly rainfall ranges from 26.39 to 47.64, MAE from 13.27 to 20.73, MASE from 0.55 to 0.60 and CE from 0.21 to 0.62. The performance measures RMSE, MAE and CE indicate that the SVM(RBF) model provides best predictions in the location Katherine and MASE indicates in the location Fairymead while the performance measures RMSE

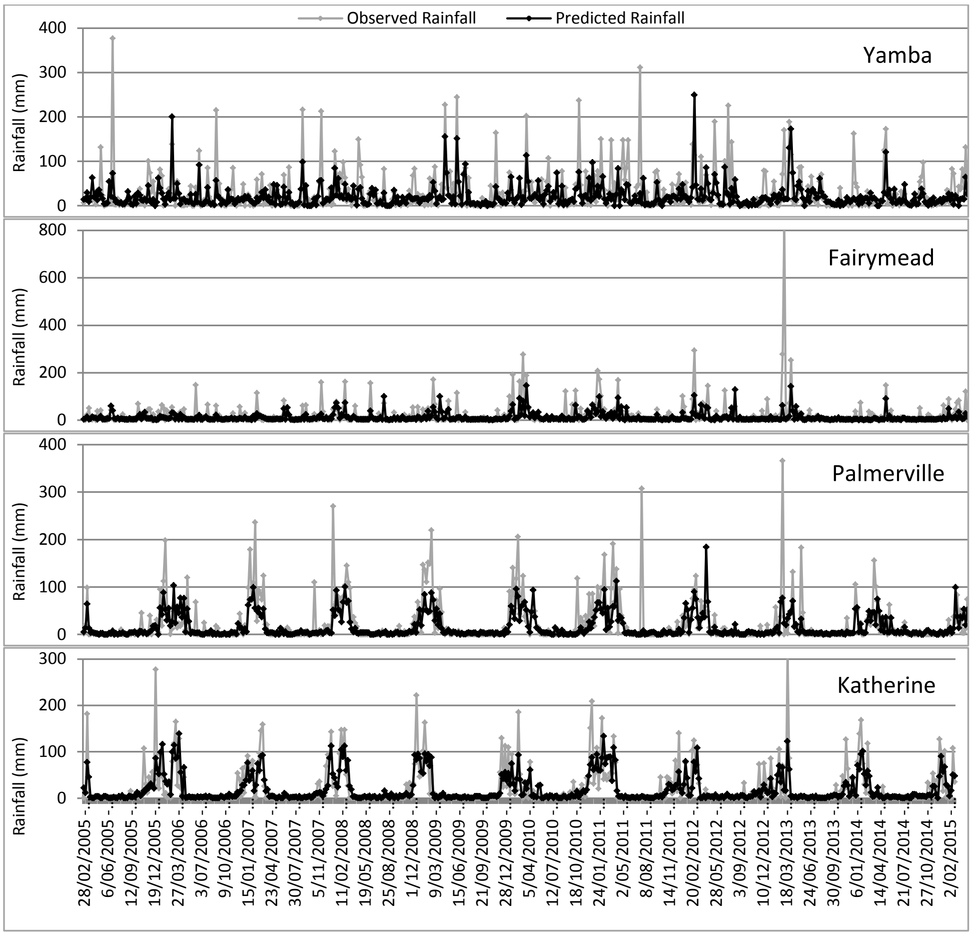

Figure 6.32: Observed rainfall vs. SVM(RBF) model weekly predictions in tropical and subtropical classification zones.

and CE indicate worst predictions in the location Fairymead; MAE indicates in the location Yamba and MASE in the location Katherine (see Figure 6.28). The graphical display of observed rainfall and SVM(RBF) model predictions over the test period is given in Figure 6.32.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP	26.39	13.27	0.60	0.62
Palmerville	Tropical	TMax, TMin, Rad	34.86	15.04	0.65	0.41
Yamba	Subtropical	TMax, TMin, Rad	38.57	20.73	0.56	0.32
Fairymead	Subtropical	TMax, TMin, VP	47.64	16.77	0.55	0.21
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP, Rad	30.64	14.81	0.67	0.48
Palmerville	Tropical	TMax, TMin, Evap, VP, Rad	48.44	21.87	0.94	-0.15
Yamba	Subtropical	TMax, TMin, Evap, VP, Rad	43.26	24.11	0.65	0.14
Fairymead	Subtropical	TMax, TMin, Evap, VP, Rad	48.14	17.56	0.57	0.20

Table 6.31: The SVM(RBF) model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

Table 6.32, summarizes the performance of the MLR model in predicting weekly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. Results presented in this table show that the MLR model provides best predictions with input variables TMax, TMin and Rad in the location Palmerville and with a full set of five variables in the remaining three locations (Katherine, Yamba and Fairymead). The model provides worst predictions with TMax, TMin and Rad in locations Katherine and Fairymead; with TMax, TMin and VP in Yamba and with a full set of parameters in Palmerville.

The performance measure RMSE for the MLR model in predicting weekly rainfall ranges from 30.25 to 47.94, MAE from 16.64 to 23.47, MASE from 0.63 to 0.80 and CE from 0.20 to 0.50. The performance measures RMSE, MAE and CE indicate that the MLR model provides best predictions in the location Katherine and MASE indicates in the location Fairymead while the performance measure RMSE and CE indicate the model provides worst predictions in the location Fairymead; MAE in the location Yamba and MASE in the location Palmerville (see Figure 6.28). The graphical display of observed rainfall and MLR model predictions over the test period is given in Figure 6.33.

Table 6.33, summarizes the performance of the ANN(0) model in predicting weekly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. According to these results ANN(0) model provides best predictions with input variables TMax, TMin and Rad in the location Palmerville; with TMax, TMin, VP and Rad in the location Yamba and with a full set of five input variables in the remaining two locations (Katherine and Fairymead). The model provides worst predictions with TMax, TMin and Rad in the location Katherine; with TMax, TMin and VP in two locations (Yamba and Fairymead) and with a full set of input variables in the remaining location Palmerville.

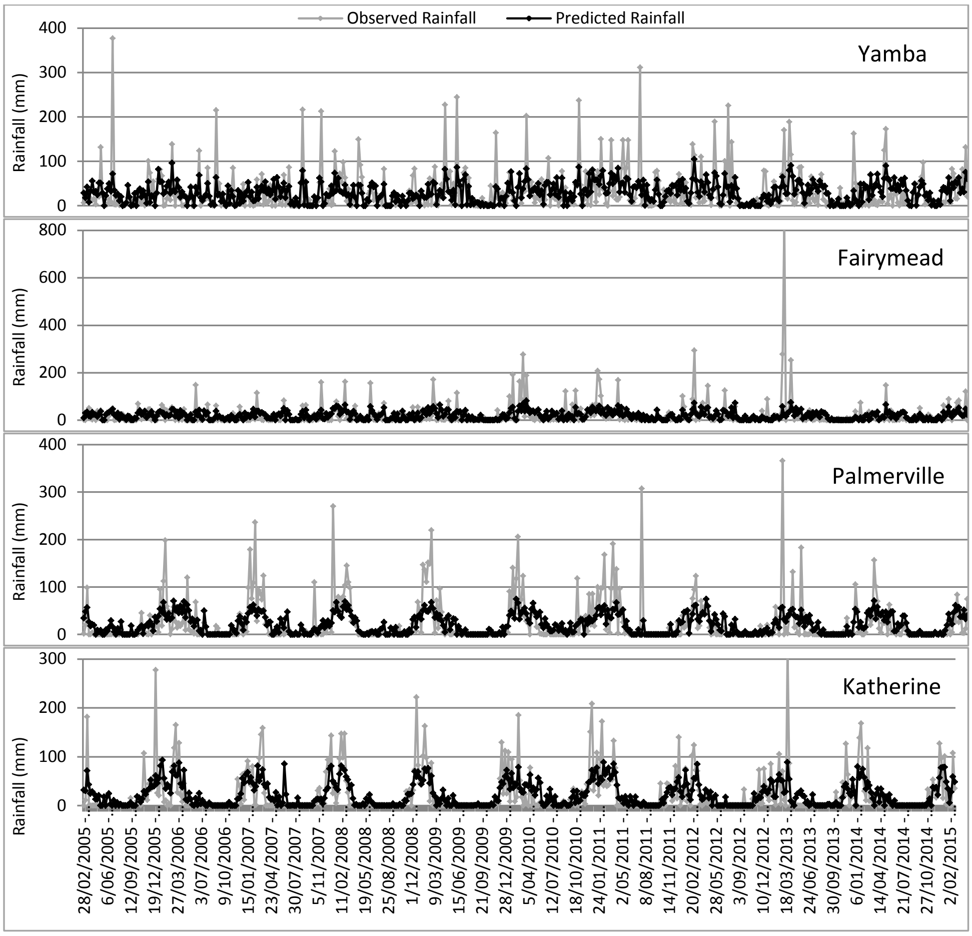

Figure 6.33: Observed rainfall vs. MLR model weekly predictions in tropical and subtropical classification zones.

The performance measure RMSE for the ANN(0) model in predicting weekly rainfall ranges from 27.77 to 47.18, MAE from 14.93 to 22.32, MASE from 0.60 to 0.68 and CE from 0.23 to 0.58. The performance measures RMSE, MAE and CE indicate that the ANN(0) model provides best predictions in the location Katherine, and MASE indicates in the location Fairymead while the performance measure RMSE and CE indicate worst predictions in the location Fairymead; MAE in the location Yamba and MASE in the location Katherine (see Figure 6.28). The graphical display of observed rainfall and ANN(0) model predictions over the test period is given in Figure 6.34.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP, Rad	30.25	16.64	0.75	0.50
Palmerville	Tropical	TMax, TMin, Rad	36.27	18.44	0.80	0.36
Yamba	Subtropical	TMax, TMin, Evap, VP, Rad	38.35	23.47	0.64	0.32
Fairymead	Subtropical	TMax, TMin, Evap, VP, Rad	47.94	19.19	0.63	0.20
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	31.97	17.29	0.78	0.44
Palmerville	Tropical	TMax, TMin, Evap, VP, Rad	36.88	20.71	0.89	0.34
Yamba	Subtropical	TMax, TMin, VP	40.29	23.26	0.63	0.25
Fairymead	Subtropical	TMax, TMin, Rad	48.46	20.03	0.65	0.18

Table 6.32: The MLR model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

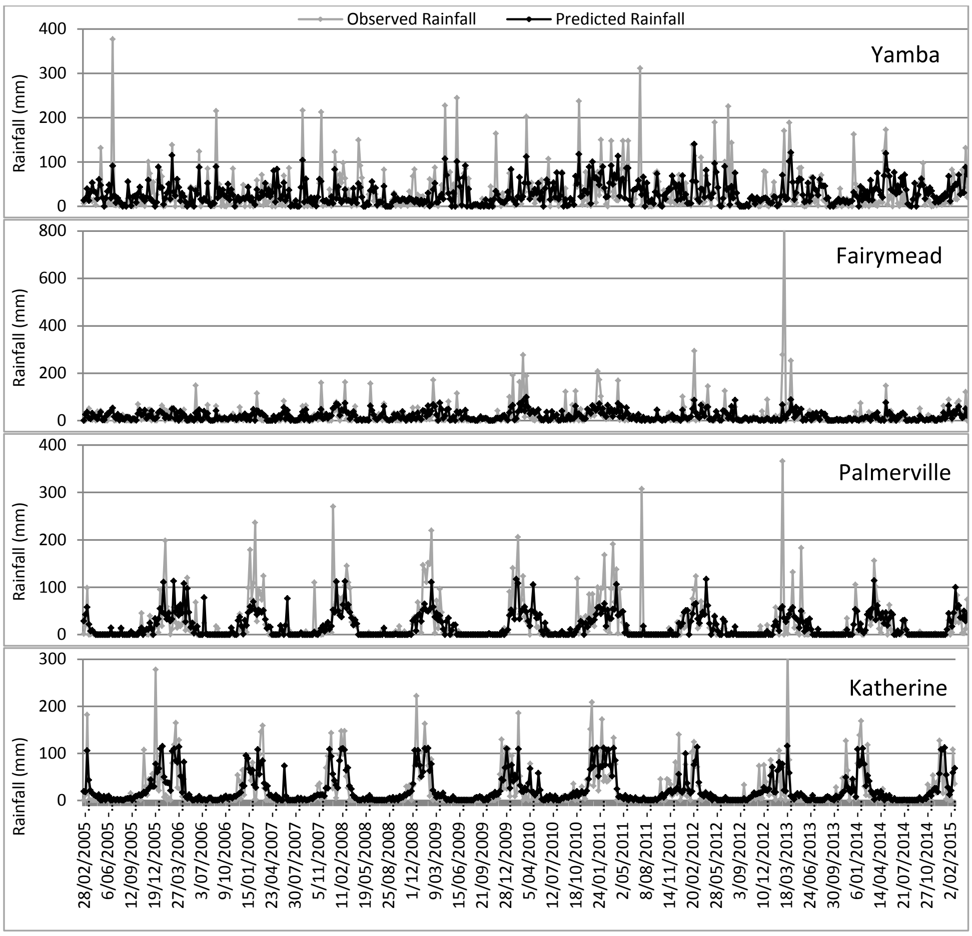

Figure 6.34: Observed rainfall vs. ANN(0) model weekly predictions in tropical and subtropical classification zones.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, Evap, VP, Rad	27.77	14.93	0.68	0.58
Palmerville	Tropical	TMax, TMin, Rad	34.82	15.59	0.67	0.41
Yamba	Subtropical	TMax, TMin, VP, Rad	36.91	22.32	0.60	0.37
Fairymead	Subtropical	TMax, TMin, Evap, VP, Rad	47.18	18.43	0.60	0.23
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	31.97	17.29	0.78	0.44
Palmerville	Tropical	TMax, TMin, Evap, VP, Rad	41.44	22.51	0.97	0.16
Yamba	Subtropical	TMax, TMin, VP	40.38	23.24	0.63	0.25
Fairymead	Subtropical	TMax, TMin, VP	48.40	18.85	0.62	0.19

Table 6.33: The ANN(0) model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

Table 6.34, summarizes the prediction performance of the ANN(1) model with best and worst combinations of input variables in tropical and subtropical classification zones.According to these results the ANN(1) model provides best predictions with input variables TMax, TMin and VP in two locations (Katherine and Fairymead) and with TMax, TMin, VP and Rad in the remaining two locations (Palmerville and Yamba). The model provides worst predictions with TMax, TMin and Rad in the three locations (Katherine, Palmerville and Fairymead) and with a full set of input variables in location Yamba.

The performance measure RMSE for the ANN(1) model in preding weekly rainfall ranges from 28.18 to 46.10, MAE from 14.90 to 21.98, MASE from 0.56 to 0.76 and CE from 0.26 to 0.56. The performance measures RMSE, MAE and CE indicate that the ANN(1) model provides best predictions in the location Katherine and MASE indicates in the location Fairymead while the performance measure RMSE and CE indicate worst predictions in the location Fairymead; MAE indicates in the location Yamba and MASE in the location Palmerville (see Figure 6.28). The graphical display of observed rainfall and ANN(1) model predictions over the test period is given in Figure 6.35.

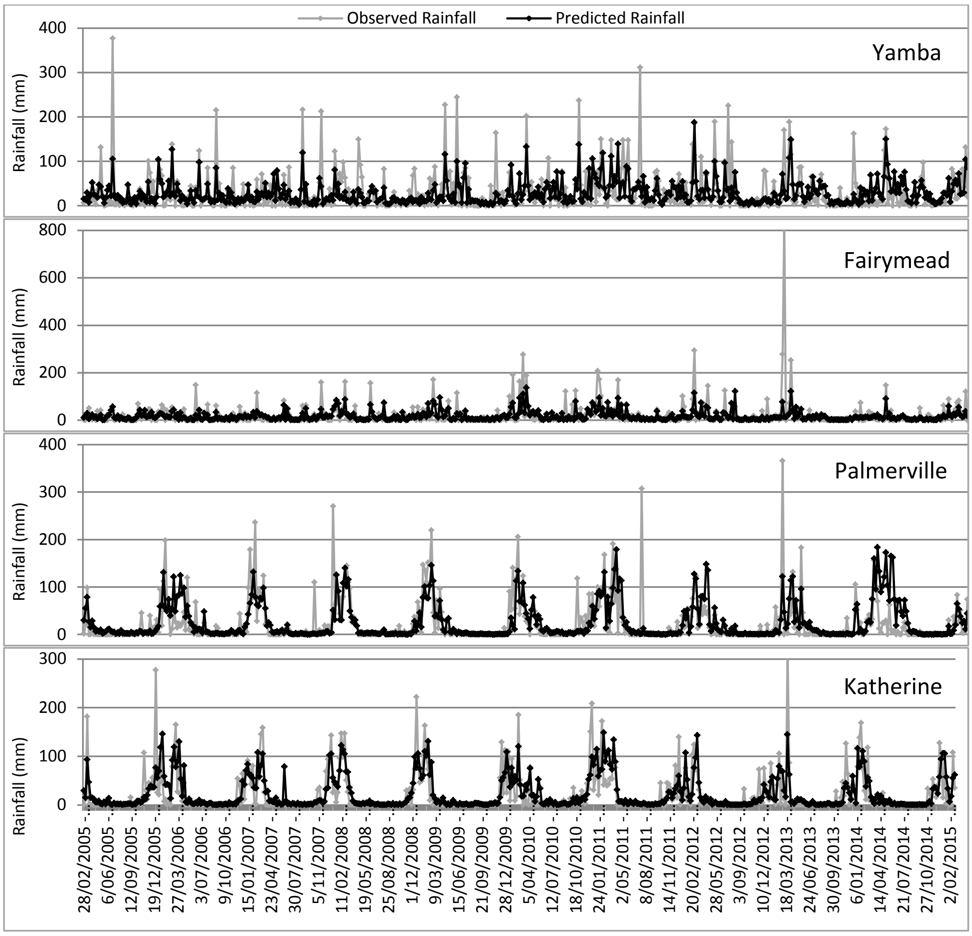

Figure 6.35: Observed rainfall vs. ANN(1) model weekly predictions in tropical and subtropical classification zones.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP	28.18	14.90	0.67	0.56
Palmerville	Tropical	TMax, TMin, VP, Rad	36.38	17.66	0.76	0.35
Yamba	Subtropical	TMax, TMin, VP, Rad	36.83	21.98	0.59	0.38
Fairymead	Subtropical	TMax, TMin, VP	46.10	17.18	0.56	0.26
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	42.63	28.70	1.30	0.00
Palmerville	Tropical	TMax, TMin, Rad	45.36	28.17	1.22	0.00
Yamba	Subtropical	TMax, TMin, EVap, VP, Rad	41.20	26.06	0.71	0.22
Fairymead	Subtropical	TMax, TMin, Rad	46.60	17.89	0.58	0.25

Table 6.34: The ANN(1) model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

Table 6.35, summarizes the performance of the k-NN model in predicting weekly rainfall with best and worst combinations of input variables in tropical and subtropical classification zones. These results shoe that k-NNprovides best predictions with input variables TMax, TMin, VP and Rad in two locations (Katherine and Fairymead); with TMax, TMin and Rad in Palmerville and with a full set of input variables in the remaining location Yamba. The k-NN provides worst predictions with TMax, TMin, and Rad in two locations (Katherine and Fairymead); with TMax, TMin, and VP in Yamba and with a full set of input variables in the location Palmerville.

The performance measure RMSE for the k-NN model in predicting weekly rainfall ranges from 28.41 to 46.22, MAE from 14.31 to 17.85, MASE from 0.58 to 0.65 and CE from 0.26 to 0.56. The performance measures RMSE, MAE and CE indicate that the k-NN model provides best predictions in the location Katherine and MASE indicates in the location Fairymead while the performance measure RMSE, MAE and CE indicate the model provides worst predictions in the location Fairymead and MASE in the location Katherine (see Figure 6.28). The graphical display of observed rainfall and k-NN model predictions over the test period is given in Figure 6.36.

Stations	Zone	Combination	RMSE	MAE	MASE	CE
Performance measures for best input combinations						
Katherine	Tropical	TMax, TMin, VP, Rad	28.41	14.31	0.65	0.56
Palmerville	Tropical	TMax, TMin, Rad	33.11	14.93	0.64	0.46
Yamba	Subtropical	TMax, TMin, Evap, VP, Rad	38.01	22.96	0.62	0.34
Fairymead	Subtropical	TMax, TMin, VP, Rad	46.22	17.85	0.58	0.26
Performance measures for worst input combinations						
Katherine	Tropical	TMax, TMin, Rad	31.00	15.08	0.68	0.47
Palmerville	Tropical	TMax, TMin, Evap, VP, Rad	33.90	15.21	0.66	0.44
Yamba	Subtropical	TMax, TMin, VP	40.28	22.49	0.61	0.25
Fairymead	Subtropical	TMax, TMin, Rad	47.18	19.34	0.63	0.23

Table 6.35: The k-NN model performance for weekly rainfall predictions with best and worst combinations of input variables in tropical and subtropical zones.

Table 6.36, summarizes the performance of all eight models in predicting weekly rainfall with best combinations of input variables in tropical and subtropical zones. Results presented in this table show that at least one performance measure indicate that the CLR(Opt) model outperformed other models in three out of four locations (Palmerville, Yamba and Fairymead); MLR model in two locations (Yamba and Fairymead); CR(EM) and ANN(1) in Yamba and SVM(RBF) in Katherine. The SVM(Linear), ANN(0) and k-NN models are not best in any location.

According to performance measure RMSE, the CLR(Opt) model outperformed other models in Palmerville and Fairymead; the SVM(RBF) model in Katherine and the MLR model in Yamba. In the location Katherine, the CLR(Opt) model perfor-

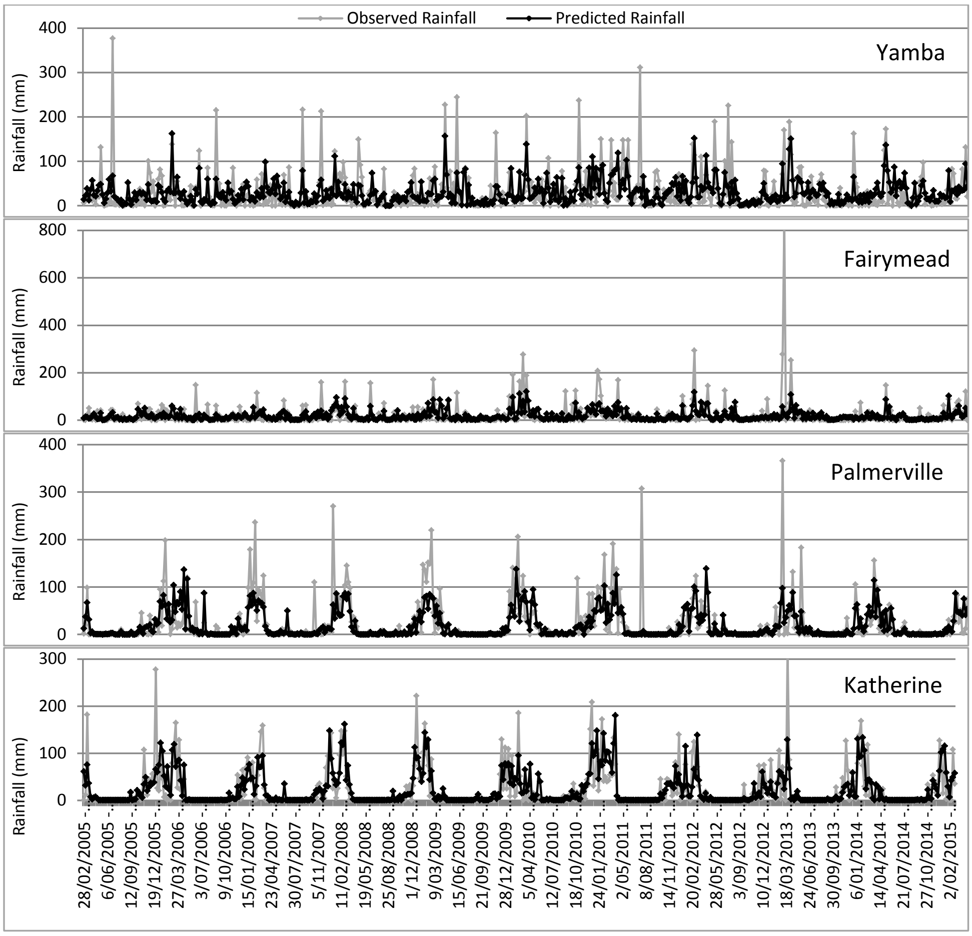

Figure 6.36: Observed rainfall vs. k-NN model weekly predictions in tropical and subtropical classification zones.

Stations	Measures	CLR	CR (EM)	SVM (Linear)	SVM (RBF)	ANN (0)	ANN (1)	MLR	KNN
Katherine (Tropical)	RMSE	27.58	36.71	35.18	**26.39**	30.25	27.77	28.18	28.41
	MAE	13.71	17.24	16.77	**13.27**	16.64	14.93	14.90	14.31
	MASE	0.62	0.78	0.76	**0.60**	0.75	0.68	0.67	0.65
	CE	0.58	0.26	0.32	**0.62**	0.50	0.58	0.56	0.56
Palmerville (Tropical)	RMSE	**32.37**	38.62	39.99	34.86	36.27	34.82	36.38	33.11
	MAE	**14.37**	17.77	18.47	15.04	18.44	15.59	17.66	14.93
	MASE	**0.62**	0.77	0.80	0.65	0.80	0.67	0.76	0.64
	CE	**0.49**	0.27	0.22	0.41	0.36	0.41	0.35	0.46
Yamba (Subtropical)	RMSE	36.99	38.57	42.22	43.26	38.35	36.91	**36.83**	38.01
	MAE	21.77	**20.73**	22.14	24.11	23.47	22.32	21.98	22.96
	MASE	**0.59**	0.56	0.60	0.65	0.64	0.60	**0.59**	0.62
	CE	**0.37**	0.32	0.18	0.14	0.32	**0.37**	0.38	0.34
Fairymead (Subtropical)	RMSE	**45.18**	50.17	52.08	48.14	47.94	47.18	46.10	46.22
	MAE	17.43	18.68	18.64	17.56	19.19	18.43	**17.18**	17.85
	MASE	0.57	0.61	0.61	0.57	0.63	0.60	**0.56**	0.58
	CE	**0.29**	0.13	0.06	0.20	0.20	0.23	0.26	0.26

Table 6.36: Models performance of all eight linear and non-linear models for weekly rainfall predictions in tropical and subtropical zones using best combination of input variables.

mance is 4.31% lower and in Yamba 0.43% than the outperformed models. Results demonstrate that the CLR(Opt) model is superior than SVM(Linear), ANN(0) and k-NN models in predicting weekly rainfall in tropical and subtropical classification zones.

A visual comparison of model predictions with the actual observations in temperate classification zone, given in Figures 6.29, 6.30, 6.31, 6.32, 6.33, 6.34, 6.35 and 6.36, show that all models follow the series patterns in all locations of tropical and subtropical classification zones.

In summary, based on the primary performance measure RMSE, the CLR(Opt) model is the most suitable model in finding the patterns and trends of the observations compared to other models at most sites in tropical and subtropical classification zones.

6.5 Summary of chapter

In this chapter, we reported the results of the proposed model and seven other models for weekly rainfall predictions over ten years. In the development of our models, twenty-four geographically diverse weather stations were used: seven from the temperate zone, eight from the grassland, five from the desert, two from the tropical and two from the subtropical zones. All the selected prediction models were developed for each weather station using training sets and evaluated by using test sets. The prediction performance of models was evaluated by comparing observed and predicted rainfall using performance measures RMSE, MAE, MASE and CE. The computational results of the proposed model were compared with those obtained from CR(EM), SVM(Linear), SVM(RBF), MLR, ANN(0), ANN(1) and k-NN models.

The results reported in this chapter show that in temperate classification zone, the CLR(Opt) and ANN(0) models are most capable of finding the patterns and trends of the observations. In grassland classification zone, ANN(1) model is the most suitable model for monthly rainfall predictions compared to other models. In the desert, tropical and subtropical classification zones, the CLR(Opt) is the best in predicting weekly rainfall in most locations.

The prediction performance of all models varied considerably with changes of geographic regions. In tropical and subtropical zones, predictions have more deviation from the actual rainfall observation. This may be because of higher rainfall variability and extreme values.

The results also confirmed that no single input variable provides the best predictions. Adding more input variables improved the performance of models in most locations used in this study. Results also found that these five meteorological variables are suitable predictors for weekly rainfall prediction in Australia.

Chapter 7

Conclusions and Future Research

The aims of this thesis were to develop optimization based clusterwise linear regression method and prediction methods based on it to improve the accuracy of rainfall prediction in Australia. This study also investigated the influence of meteorological parameters (both individually and in combination) and geographic regions on the performance of models in predicting monthly and weekly rainfall in Australia.

Accurate rainfall prediction is a serious concern in many countries; especially in Australia where the climate is highly variable and water supply mainly depends on rainfall. Any change in the probability of rainfall (heavy rainfall or drought) has important implications for future resource planning, management and investment.

Predicting rainfall is a complex process, needing continual improvement. There are many approaches to modeling rainfall. Many studies predicted rainfall using only historical data as input and some used climate indices as input variables. There have been comparatively few studies where rainfall was predicted using meteorological variables as input.

In this thesis, we have developed a new optimization based clusterwise linear regression method. This method simultaneously divides the data into k clusters and fits the regression functions within each cluster. It is an incremental algorithm that starts with one linear function and gradually adds one function at each iteration until the satisfactory approximation is achieved.

Predictions can be computed from each cluster using prediction methods, however finding the one with the lowest error is a challenge. We introduced and applied new prediction methods for both monthly and weekly rainfall predictions. We assessed the prediction performance of the proposed CLR(Opt) model with several multivariate models including clusterwise linear regression based on maximum likelihood estimation. This research also investigated the influence of meteorological variables and geographic regions on the performance of models in predicting monthly and weekly rainfall in Australia.

This thesis has presented results that showed:

- Among all eight models, the CLR(Opt) is the most accurate for weekly rainfall predictions in tropical and subtropical climate zones and for both monthly and weekly rainfall predictions in the desert climate zone. In temperate zone, CLR(Opt) and SVM(RBF) models give the best predictions for monthly rainfall and CLR(Opt) and ANN(0) for weekly rainfall. In tropical and subtropical climate zones, the CLR(Opt) and k-NN models are the most accurate models for monthly rainfall predictions. The ANN(1) model is best in grassland zone.

- The CLR(Opt) is superior to the CLR(EM), SVM(Linear), ANN(0) and MLR models in most locations used in this study for both weekly and monthly rainfall predictions.

- Two linear models, SVM(linear) and MLR, in general, are not accurate models for rainfall prediction.

- All eight models at all locations, with a very few exceptions, fail to predict extreme rainfalls.

- The prediction performance of all models varies considerably both within and across climate zones. Predictions have the lowest deviation from the actual observations in desert climate zone and the highest in tropical and subtropical zones.

- No single meteorological parameter as an input variable in the models provides best predictions. Adding input variables increased the prediction performance of models in most locations used in this study. Results also found that these five meteorological variables are suitable predictors for both monthly and weekly rainfall predictions in Australia.

Rainfall is a very complex climate variable. It is controlled by physical processes involving random fluctuations. The relationship between rainfall and climate or meteorological variables is highly nonlinear. Results confirm that data-driven modeling presents a powerful approach for rainfall prediction. Models which are able to capture nonlinearities are the most suitable for such predictions. Our results on the CLR(Opt), SVM(RBF), k-NN and ANN(1) models confirm this conclusion. Our results also confirm that the CLR(Opt) model is an efficient method for rainfall predictions and a good alternative to existing mainstream models. However, results from this research also show that mainstream models are not always successful for rainfall predictions especially, for extreme rainfall predictions.

The results obtained in this research are useful to understand the relationship between rainfall and meteorological variables in different climate zones in Australia and may lead to further development in the modeling and predictions of weekly and monthly rainfall. Future work should continue to develop more accurate prediction methods based on clusterwise linear regression. Furthermore, hybrids of clusterwise linear regression and other models can be developed. Such models should be able, in particular, to predict extreme rainfall events which are the real challenge for all existing models.

Bibliography

[1] D. B. Rubin A. P. Dempster, N. M. Laird. Maximum Likelihood from Incomplete Data via the *EM* Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.

[2] John Abbot and Jennifer Marohasy. Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences*, 29(4):717–730, 2012.

[3] John Abbot and Jennifer Marohasy. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmospheric Research*, 138:166–178, 2014.

[4] Nachiketa Acharya, Nitin Anand Shrivastava, BK Panigrahi, and UC Mohanty. Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. *Climate dynamics*, 43(5-6):1303–1310, 2013.

[5] Hafzullah Aksoy and Ahmad Dahamsheh. Artificial neural network models for forecasting monthly precipitation in Jordan. *Stochastic Environmental Research and Risk Assessment*, 23(7):917–931, 2009.

[6] YourHome Australian Government. http://www.yourhome.gov.au/water, 2013.

[7] Adil M Bagirov. Modified global k-means algorithm for minimum sum-of-squares clustering problems. *Pattern Recognition*, 41(10):3192–3199, 2008.

[8] Adil M Bagirov. Modified global k-means algorithm for minimum sum-of-squares clustering problems. *Pattern Recognition*, 41(10):3192–3199, 2008.

[9] Adil M Bagirov, Julien Ugon, and Dean Webb. Fast modified global k-means algorithm for incremental cluster construction. *Pattern recognition*, 44(4):866–876, 2011.

[10] Adil M Bagirov and John Yearwood. A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems. *European Journal of Operational Research*, 170(2):578–596, 2006.

[11] A.M. Bagirov, B. Karasozen, and M. Sezer. Discrete gradient method: Derivative-free method for nonsmooth optimization. *Journal of Optimization Theory and Applications*, 137(2):317–334, 2008.

[12] AM Bagirov, AM Rubinov, NV Soukhoroukova, and J Yearwood. Unsupervised and supervised data classification via nonsmooth and global optimization. *Top*, 11(1):1–75, 2003.

[13] A.M. Bagirov and J. Ugon. Piecewise partially separable functions and a derivative-free algorithm for large scale nonsmooth optimization. *Journal of Global Optimization*, 35(2):163–195, 2006.

[14] A.M. Bagirov, J. Ugon, and H. Mirzayeva. Nonsmooth nonconvex optimization approach to clusterwise linear regression problems. *European Journal of Operational Research*, 229(1):132–142, 2013.

[15] A.M. Bagirov, J. Ugon, and H.G. Mirzayeva. An algorithm for clusterwise linear regression based on smoothing techniques. *Optimization Letters*, 9(2):375–390, 2015.

[16] A.M. Bagirov, J. Ugon, and H.G. Mirzayeva. Nonsmooth optimization algorithm for solving clusterwise linear regression problems. *Journal of Optimization Theory and Applications*, 164(3):755–780, 2015.

[17] A Brath, A Montanari, and E Toth. Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrology and Earth System Sciences Discussions*, 6(4):627–639, 2002.

[18] R.A. Carbonneau, G. Caporossi, and P. Hansen. Globally optimal clusterwise regression by mixed logical-quadratic programming. *European Journal of Operational Research*, 212:213–222, 2011.

[19] Surajit Chattopadhyay and Goutami Chattopadhyay. Comparative study among different neural net learning algorithms applied to rainfall time series. *Meteorological applications*, 15(2):273–280, 2008.

[20] Surajit Chattopadhyay and Goutami Chattopadhyay. Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. *Comptes Rendus Geoscience*, 342(2):100–107, 2010.

[21] FHS Chiew, WJ Young, W Cai, and J Teng. Current drought and future hydroclimate projections in Southeast Australia and implications for water resources management. *Stochastic Environmental Research and Risk Assessment*, 25(4):601–612, 2011.

[22] R. Collobert and S. Bengio. Support vector machines for large-scale regression problems, 2000.

[23] Corinna Cortes and Vladimir Vapnik. Support-Vector Networks. *Machine learning*, 20(3):273–297, 1995.

[24] Ravinesh C Deo and Mehmet Şahin. Application of the artificial neural network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in Eastern Australia. *Atmospheric Research*, 161:65–81, 2015.

[25] Wayne S DeSarbo, Richard L Oliver, and Arvind Rangaswamy. A simulated annealing methodology for clusterwise linear regression. *Psychometrika*, 54(4):707–736, 1989.

[26] W.S. DeSarbo and W.L. Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(2):249–282, 1988.

[27] A El-Shafie, Muhammad Mukhlisin, Ali A Najah, and Mohd Raihan Taha. Performance of artificial neural network and regression techniques for rainfall-runoff prediction. *International Journal of Physical Sciences*, 6(8):1997–2003, 2011.

[28] A El-Shafie, A Noureldin, MR Taha, and A Hussain. Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia. *Hydrology and Earth System Sciences Discussions*, 8(4):6489–6532, 2011.

[29] Alireza Eskandarin, Hadi Nazarpour, Mehdi Teimouri, and Mirkhalegh Z Ahmadi. Comparison of neural network and k-nearest neighbor methods in daily flow forecasting. *Journal of Applied Sciences*, 10:1006–1010, 2010.

[30] Jamileh Farajzadeh, Ahmad Fakheri Fard, and Saeed Lotfi. Modeling of monthly rainfall and runoff of Urmia lake basin using feed-forward neural network and time series analysis model. *Water Resources and Industry*, 7:38–48, 2014.

[31] Susana Faria and Gilda Soromenho. Fitting mixtures of linear regressions. *Journal of Statistical Computation and Simulation*, 80(2):201–225, 2010.

[32] World Business Council for Sustainable Development. Water Facts and Trends. http://wbcsdpublications.org/project/water-facts-and-trends/, 2005.

[33] S Gaffney, P Smyth, S Chaudhuri, and D Madigan. Trajectory clustering using mixtures of regression models. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, pages 63–72.

[34] Giorgio Galeati. A comparison of parametric and non-parametric methods for runoff forecasting. *Hydrological sciences journal*, 35(1):79–94, 1990.

[35] L.A. Garcìa-Escudero, A. Gordaliza, A. Mayo-Iscar, and R. San Martin. Robust clusterwise linear regression through trimming. *Computational Statistics and Data Analysis*, 54:3057–3069, 2010.

[36] Luis Angel García-Escudero, Alfonso Gordaliza, Agustín Mayo-Iscar, and R San Martín. Robust clusterwise linear regression through trimming. *Computational Statistics & Data Analysis*, 54(12):3057–3069, 2010.

[37] Ross. Garnaut. Projecting Australian Climate Change, The Garnaut Climate Change Review. http://www.garnautreview.org.au/chp5.htm, 2008.

[38] Australian Government. Australian weather and seasons, 2013.

[39] Bettina Grün, Friedrich Leisch, et al. Flexmix version 2: finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28(4):1–35, 2008.

[40] C Hennig. Models and methods for clusterwise linear regression. In *Classification in the Information Age*, pages 179–187. Springer, 1999.

[41] Christian Hennig. Identifiablity of models for clusterwise linear regression. *Journal of Classification*, 17(2):273–296, 2000.

[42] David Hosmer. Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics-Theory and Methods*, 3(10):995–1006, 1974.

[43] R.J. Hyndman and A.B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

[44] Mohammad Karamouz, Saman Razavi, and Shahab Araghinejad. Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks: a case study. *Hydrological Processes*, 22(2):229–241, 2008.

[45] Ozgur Kisi and Mesut Cimen. Precipitation forecasting by using wavelet-support vector machine conjunction model. *Engineering Applications of Artificial Intelligence*, 25(4):783–792, 2012.

[46] Kin-Nam Lau, Pui-lam Leung, and Ka-kit Tse. A mathematical programming approach to clusterwise regression model and its extensions. *European Journal of Operational Research*, 116(3):640–652, 1999.

[47] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003.

[48] Gwo-Fong Lin, Guo-Rong Chen, Ming-Chang Wu, and Yang-Ching Chou. Effective forecasting of hourly typhoon rainfall using support vector machines. *Water resources research*, 45(8), 2009.

[49] Gwo-Fong Lin and Bing-Chen Jhong. A real-time forecasting model for the spatial distribution of typhoon rainfall. *Journal of Hydrology*, 521:302–313, 2015.

[50] Gwo-Fong Lin, Bing-Chen Jhong, and Chia-Chuan Chang. Development of an effective data-driven model for hourly typhoon rainfall forecasting. *Journal of Hydrology*, 495:52–63, 2013.

[51] T Mandal and V Jothiprakash. Short-term rainfall prediction using ANN and MT techniques. *ISH Journal of Hydraulic Engineering*, 18(1):20–26, 2012.

[52] F Mekanik, MA Imteaz, S Gato-Trinidad, and A Elmahdi. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology*, 503:11–21, 2013.

[53] Andrew Mercer, Jamie Dyer, and Song Zhang. Warm-Season Thermodynamically-Driven Rainfall Prediction with Support Vector Machines. *Procedia Computer Science*, 20:128–133, 2013.

[54] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2015. R package version 1.6-7.

[55] et al. Mller, K.-R. Predicting time series with support vector machines. In *Artificial Neural Networks-ICANN'97*, pages 999–1004. Springer, 1997.

[56] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2015.

[57] Priya Narayanan, Ashoke Basistha, Sumana Sarkar, and Sachdeva Kamna. Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. *Comptes rendus geoscience*, 345(1):22–27, 2013.

[58] J Eamonn Nash and Jonh V Sutcliffe. River flow forecasting through conceptual models part 1A discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.

[59] Munir Ahmad Nayak and Subimal Ghosh. Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theoretical and applied climatology*, 114(3-4):583–603, 2013.

[60] Simon Newcomb. A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, pages 343–366, 1886.

[61] Hien Nguyen. Finite mixture models for regression problems. 2015.

[62] Australian Bureau of Meteorology. The key climate groups. http://www.bom.gov.au/iwk/climate_zones/map_1.shtml.

[63] Australian Bureau of Meteorology. State of the Climate. http://www.bom.gov.au/state-of-the-climate/, 2014.

[64] Australian Bureau of Meteorology. Annual Climate Statement. http://www.bom.gov.au/climate/current/annual/aus/, 2015.

[65] Australian Bureau of Statistics. Australia's Environment: Issues and Trends, 2007. http://www.abs.gov.au/AUSSTATS/abs@.nsf/9e1dd9680bdd9821ca257090002029cc/330bc8fdfd50bee4ca2573c6001049f9!OpenDocument, 2007.

[66] Australian Bureau of Statistics. Year Book Australia, 2012. http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1301.0~2012~Main%20Features~Water~279, 2012.

[67] Australian Bureau of Statistics. Environmental Issues: Water use and Conservation. http://www.abs.gov.au/ausstats/abs@.nsf/mf/4602.0.55.003?OpenDocument, 2013.

[68] Australian Bureau of Statistics. Annual Climate Statement 2015. http://www.bom.gov.au/climate/current/annual/aus/, 2015.

[69] Australian Bureau of Statistics. Water Account, Australia, 2013-2014. http://www.abs.gov.au/ausstats/abs@.nsf/mf/4610.0, 2015.

[70] Intergovernmental Panel on Climate Change. Climate Change 2014, Synthesis Report. https://www.ipcc.ch/report/ar5/syr/, 2014.

[71] Bambang Widjanarko Otok. Development of rainfall forecasting model in indonesia by using ASTAR, transfer function, and ARIMA methods. *European Journal of Scientific Research*, 38(3):386–395, 2009.

[72] Karl Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894.

[73] Cristian Preda and Gilbert Saporta. Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 49(1):99–108, 2005.

[74] Guoqi Qian, Yuehua Wu, and Qing Shao. A procedure for estimating the number of clusters in logistic regression clustering. *Journal of classification*, 26(2):183–199, 2009.

[75] Richard E Quandt. A new approach to estimating switching regressions. *Journal of the American statistical association*, 67(338):306–310, 1972.

[76] Maria Cleofe Valverde Ramirez, Haroldo Fraga de Campos Velho, and Nelson Jesus Ferreira. Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *Journal of hydrology*, 301(1):146–162, 2005.

[77] K. Schliep and K. Hechenbichler. *kknn: Weighted k-Nearest Neighbors*, 2016. R package version 1.3.1.

[78] Queensland Chief Scientist. Understanding floods. http://www.chiefscientist.qld.gov.au/images/documents/chiefscientist/pubs/floods/understanding-floods_full_colour.pdf, 2011.

[79] Asaad Y Shamseldin and Kieran M O'Connor. A nearest neighbour linear perturbation model for river flow forecasting. *Journal of Hydrology*, 179(1):353–375, 1996.

[80] Qing Shao and Yuehua Wu. A consistent procedure for determining the number of clusters in regression clustering. *Journal of Statistical Planning and Inference*, 135(2):461–476, 2005.

[81] Vigya Sharma, Shashi van de Graaff, Barton Loechel, and Daniel M Franks. *Extractive resource development in a changing climate: Learning the lessons from extreme weather events in Queensland, Australia*, volume 4. National Climate Change Adaptation Research Facility, 2013.

[82] Ravi P Shukla, Krishna C Tripathi, Avinash C Pandey, and IML Das. Prediction of Indian Summer monsoon rainfall using niño indices: a neural network approach. *Atmospheric Research*, 102(1):99–109, 2011.

[83] SILO. Scientific Information for Land Owners, 2015.

[84] VK Somvanshi, OP Pandey, PK Agrawal, NV Kalanker, M Ravi Prakash, and Ramesh Chand. Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, 10(2):141–151, 2006.

[85] H. Späth. Algorithm 48: A fast algorithm for clusterwise linear regression. *Computing*, 29:175–181, 1981.

[86] Helmuth Späth. Algorithm 39 Clusterwise linear regression. *Computing*, 22(4):367–373, 1979.

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

[88] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.

[89] Michel Wedel and Wayne S DeSarbo. A mixture likelihood approach for generalized linear models. *Journal of Classification*, 12(1):21–55, 1995.

[90] Svante Wold, Arnold Ruhe, Herman Wold, and WJ Dunn, III. The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3):735–743, 1984.

[91] Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, S Yu Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.

[92] S Yakowitz and M Karlsson. Nearest neighbor methods for time series, with application to rainfall/runoff prediction. In *Advances in the Statistical Sciences: Stochastic Hydrology*, pages 149–160. Springer, 1987.

[93] Bin Zhang. Regression clustering. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 451–458. IEEE, 2003.