UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The use of artificial neural networks to enhance numerical weather prediction model forecasts of temperature and rainfall

by

Hester Gerbrecht Marx

A dissertation submitted in part fulfilment of the requirements for the degree

MASTER OF SCIENCE

in the

Department of Geography, Geoinformatics and Meteorology
Faculty of Natural and Agricultural Science

UNIVERSITY OF PRETORIA

July 2008

DECLARATION

I hereby declare that the dissertation that I hereby submit for the degree MSc(Meteorology) at the University of Pretoria is my own work, and that it has not been previously submitted by me for degree purposes at any other university. I also declare that all the sources I have quoted have been indicated and acknowledge by complete references.

_____ _____

Signature Date

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The use of artificial neural networks to enhance numerical weather prediction model forecasts of temperature and rainfall

Hester Gerbrecht Marx

Supervisor: Prof. C.J.deW. Rautenbach
Co-Supervisor: Dr. W.J. Tennant
Department: Geography, Geoinformatics and Meteorology
University: University of Pretoria
Degree: Master of Science (Meteorology)

SUMMARY

Statistical post-processing techniques are used to remove systematic biases in modeled data. Models have shortcomings in the physical parameterization of weather events and have the inability to handle sub-grid phenomena successfully. The accuracy of forecasts interpolated to station points is limited by the horizontal resolution of the model. The magnitude of the bias at a station point depends upon geographical location and season. A neural network (NN) is a statistical downscaling method that seeks to model the linear or non-linear relationship between a set of different predictors and the predictand. NN's have a training rule whereby the weights of connections between predictors and the predictand, are adjusted on the basis of the data. NN systems have been developed by using as input, different model variables from the NCEP Ensemble Prediction System (EPS) and Eta model to forecast minimum/maximum temperature and rainfall (Quantitative Precipitation Forecast (QPF) and Probability of Precipitation (PoP)), respectively. Results show some potential for improved NN forecasts over the forecast generated by the Numerical Weather Prediction (NWP) models. The implementation of a NN system can serve as a guidance tool in operational forecasting but with one difficulty that the NWP model has to be frozen, meaning no upgrades or changes on the model.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to the following persons and organisation for their assistance and contribution to make this dissertation possible:

- My supervisors, Prof Hannes Rautenbach and Dr Warren Tennant, for their help and guidance.
- The South African Weather Service (SAWS) for the data and use of their computer resources.
- The SAWS personnel for their support.
- Karin and Anastasia for finding the research articles.
- Special thanks to my husband, Frans, and the rest of my family.

This dissertation is dedicated with love to:

My father and mother – Kobus and Magda Pienaar

My husband and daughter – Frans and Miané Marx

CONTENTS

LIST OF ABBREVIATIONS

AE	Absolute Error
AMS	American Meteorological Society
ARPS	Advanced Regional Prediction System
BMJ	Betts-Miller-Janjic
Brel	Reliability skill score
Bres	Resolution skill score
Bs	Brier score
Bss	Brier skill score
CAPE	Convective Available Potential Energy
CIN	Convective Inhibition
CGM	Conjugate Gradient Method
CSI	Critical Success Index
ECMWF	European Center for Medium-Range Weather Forecasts
ENS	Ensemble forecast
EPS	Ensemble Prediction System
FAR	False Alarm Ratio
FCSTER	Forecaster
GL	Generalization Loss
HSS	Heidke Skill Score
KSS	Kuiper Skill Score
LRC	Likelihood Ratio Chi-Square
MOS	Model Output Statistics
MSE	Mean Square Error
NCEP	National Centers of Environmental Prediction
NN	Neural Network
NWP	Numerical Weather Prediction
PCA	Principle Component Analysis
POD	Probability of Detection
POFD	Probability of False Detection
PoP	Probability of Precipitation
PP	Perfect Prognosis
PQPF	Probabilistic Quantitative Precipitation Forecast
QPF	Quantitative Precipitation Forecast
RMSE	Root Mean Square Error
ROC	Relative Operating Characteristic
RSM	Regional Spectral Model
RV	Relative Value
SAST	South African Standard Time
SAWS	South African Weather Service
SDT	Signal Detection Theory
SCG	Scaled Conjugate Gradient
SNNS	Stuttgart Neural Network Simulator
Tmin	Minimum Temperature
Tmax	Maximum Temperature
UK	United Kingdom
USWRP	United States Weather Research Program

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

BACKGROUND

1.1 Introduction

Minimum and maximum temperature (Tmin and Tmax) forecasts and precipitation forecasts are vital products to be provided by any operational weather forecasting office. Forecasts generated by Numerical Weather Prediction (NWP) models are imperfect due to errors in the initial and boundary conditions that are fed into the model. Therefore there is a need to apply statistical post-processing techniques on modeled forecast fields to improve forecast quality and value. These enhanced forecasts can serve as objective guidance supplements to their daily forecasting procedure.

1.2 The importance of temperature and precipitation forecasts

Tmin's and Tmax's that exceed certain thresholds delay many open-air activities and public services. Temperatures can reach extreme low or high values that lead to discomfort and even death, especially among small children and old people. Very low Tmin's can result in damage on agriculture sectors by the formation of frost and can lead to poor visibility due to temperature inversions at low levels in the atmosphere. Temperature values in South Africa are fairly moderate in nature and seldom produce extreme cold or hot conditions relative to the rest of the world. Figures 1.1 and 1.2 show the monthly average and extreme temperature values for the Pretoria and Cape Town station respectively which were acquired from climate data archived at the SAWS. Temperatures were averaged over a 30-year period from 1961 to 1990. The choice of an inland station, Pretoria and a coastal station, Cape Town are to highlight the different temperature characteristics between such stations.

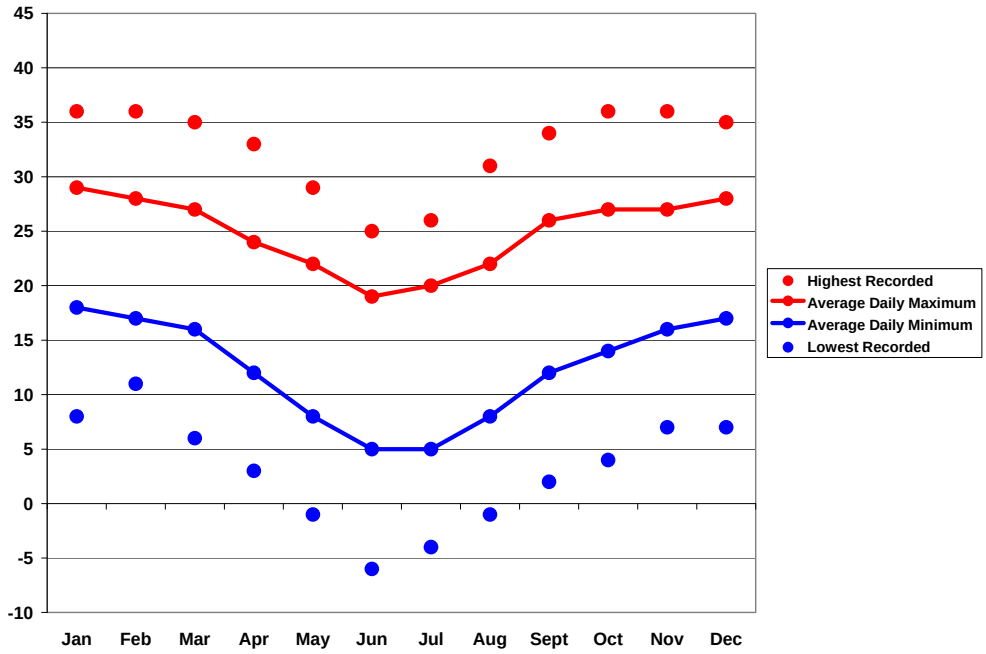

Figure 1.1 **The monthly average and extreme temperature values for the Pretoria station acquired from climate data archived at the SAWS.**

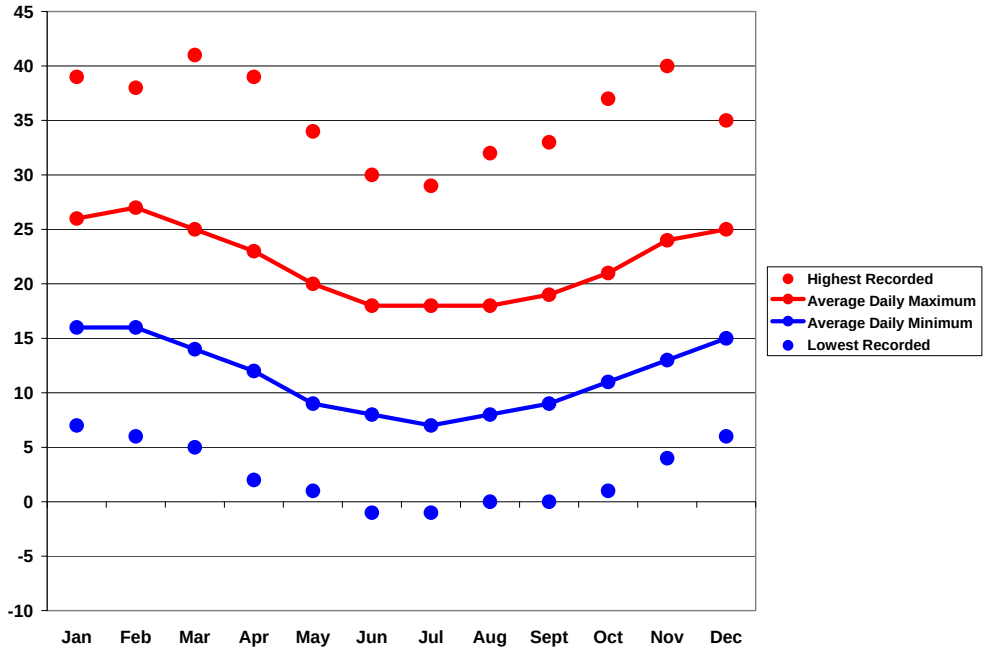

Figure 1.2 **The monthly average and extreme temperature values for the Cape Town station acquired from climate data archived at the SAWS.**

Tmax's at Cape Town are mostly lower than those at Pretoria throughout the year. During the months after mid-winter (August to October) the Tmax's at Cape Town are slower to increase towards summer than at Pretoria. Extreme Tmax's at Cape

Town are on average 14 °C higher than the monthly averages which are much higher than the average 8 °C difference found at Pretoria. Tmax's at Cape Town during the warm season range from 21 °C to 27 °C and Tmax's at Pretoria range from 27 to 29 °C. During the cold season Cape Town Tmax's range from 18 to 23 °C and from 19 to 26 °C at Pretoria.

The Tmin's at Cape Town are mostly cooler than the Tmin's at Pretoria but during the mid-winter months the Tmin's at Pretoria are cooler than the Tmin's found at Cape Town. The average difference between the monthly Tmin means and extreme Tmin's are 9 °C for both Cape Town and Pretoria. Tmin's during the warm season range between 11 and 16 °C at Cape Town and between 14 and 18 °C at Pretoria. During the cold season Tmin's at Cape Town ranges between 7 and 12 °C and between 5 and 12 °C at Pretoria.

The diurnal cycle is fairly constant throughout the year at Cape Town and shows a difference of 10 to 11 °C between daily Tmin's and Tmax's. The diurnal cycle at Pretoria changes with season. During the warm season the difference between Tmin's and Tmax's are between 11 and 13 °C while in the cold season the difference increases to between 14 and 15 °C.

The different temperature characteristics between Pretoria and Cape Town can be attributed to the difference in distance that these stations are located from the ocean. Cape Town is on the coast and is therefore more subject to the rate of daily temperature change of the ocean which is slower than the rate of daily temperature change found inland.

Precipitation is classified as the most important forecast variable, but is the least predictable entity in meteorology. It affects all aspects of life including the general public, economy, agriculture, fisheries, transportation, and many others. Precipitation occurrence is the main contributor to the occurrence of floods and droughts, which in turn gives it a high importance in the field of hydrology. Precipitation is very complex in nature. The occurrence of precipitation is characterised by a large variability in space and time and is largely dependent on small-scale processes, local topography and ocean temperatures around the continent which are characterized by the warm

Agulhas current along the east coast and the cold Benguela Ocean current along the west coast. The South African continent consists of a high plateau that is surrounded by an escarpment which is steep on the eastern side. The Vaal Dam catchment which is the study area of the precipitation forecasting section of this research is located on the eastern side of the plateau and forms part of the major watershed of South Africa. The area which embraces the Vaal Dam catchment area experience annual rainfall between 600mm and 800mm based on the monthly averages over a period of 30 years (1961-1990). According to the change of percentage mean monthly rainfall at monthly intervals which were determined by Taljaard (1986) the area which encloses the Vaal dam catchment receives most of its rainfall during the months October to January. Rainfall events decrease during the month of February towards the cold season until there is a slight increase in September.

1.3 NWP Model Predictions

NWP is the integration of governing equations of hydrodynamics by numerical methods subject to specified initial conditions. Dynamical schemes that are used in a simulation are only approximations of what takes place in the real atmosphere (American Meteorological Society (AMS) Glossary). The physical processes involved in the dynamics of the atmosphere are part of a complex system and need to be simplified by using a simple model. Dynamical equations are simplified through geostrophic and hydrostatic approximations which filter out sound and gravity waves responsible for the initiation of noise into the data fields. Parameterization is the method used to approximate unresolved physical processes in terms of resolved variables. The parameterization of convection is important because it affects vertical heat transfer in the atmosphere which maintains the tropospheric lapse rate and moisture distribution. Other parameterizations include the approximation of cloud-radiation interactions. NWP models are constantly being improved and horizontal and vertical resolutions increase as computer resources improve, but there will never be a perfect forecast. Model accuracy is also affected by a complex terrain of the area covering the model domain. Poor quality and scarcity in observations gives a model an uncertain analysis field to begin with which leads to the accumulation of error as forecast lead-time increases. Furthermore, the atmosphere is a chaotic system and any

error, no matter how small, present in the initial conditions or in the model grows exponentially, gradually contaminating all scales until a forecast is rendered useless.

The bias present in modeled temperature forecasts depends on the geographical location thereof and the relevant season for which the forecast was made. Qian and Leung (2005) found that mesoscale temperature forecasts have less error than global temperature forecasts due to the better resolved topography in mesoscale models. Zhang and Frederiksen (2003) found that localized interactions between land surface and the overlaying atmosphere are the primary contributor to the impact that soil moisture conditions has on the temperature forecasts.

Simulation of precipitation in NWP occurs at a subgrid-scale due to the mesoscale nature of precipitation. Subgrid-scale precipitation is simulated through convective parameterization schemes. In parameterization schemes model physics are represented by oversimplified parameters rather than by effects that are directly consequences of the model dynamics themselves (AMS Glossary). Some of the existing convective parameterization schemes are the modified Kuo scheme (Krishnamurti et. al., 1983), the Betts-Miller-Janic (BMJ) scheme (Betts (1986), Betts and Miller (1993), Janjic (1994)), the Kain-Fritsch scheme (Kain and Fritsch (1990,1993)), the Tiedtke scheme (Tiedtke, 1989), the Arakawa-Schubert scheme (Arakawa and Schubert, 1974) and the Grell scheme (Grell, 1993, Grell et. al., 1994).

The National Centers for Environmental Prediction (NCEP) Eta model forecasts are used as input to the precipitation Neural Network (NN) in this research and therefore a brief description is given on some of its characteristics. This model uses the BMJ scheme. The BMJ scheme nudges both the temperature and moisture profiles toward a reference profile simultaneously. If a cloud layer is too dry no convection will be initiated. The shallow convection component of the scheme imitates condensation and evaporation at the base of the cloud and the cloud top, respectively, and triggers convection when the cloud depth is more than 10 hPa and less than 200 hPa. The deep convection part of the BMJ scheme identifies instabilities in the lower 130 hPa of each grid point then calculates the depth of the cloud. If the cloud depth is more than 200 hPa the scheme modifies the temperature and moisture profile. Convective

intensity simulated by the BMJ scheme is greatly dependent on the amount of moisture present in the vertical layers.

Parameterization greatly degraded the possible accuracy of precipitation forecasts until recent times. But computer resources are developing at a fast rate and explicit representation of precipitation are already possible. Replacing convective parameterization with explicit convection will improve model precipitation forecasts (Fritsch and Carbone, 2004), but still not solve the problem of errors in precipitation forecasts.

Precipitation forecasts may be expressed in two ways, namely, Quantitative Precipitation Forecasts (QPF) and Probability of Precipitation (PoP) forecasts. A QPF is a prediction of the amount of precipitation that will fall at a given location in a given time interval (AMS Glossary). QPF's are of great value to hydrology as it operates as an important input to flood forecasting. Modelled QPF's lack accuracy due to the variability and intermittency of precipitation amount in reality. A PoP forecast is a prediction of the occurrence of precipitation at a location. Mohanty et. al. (2001) generated both QPF and PoP forecasts over Delhi, India. They found both types of forecasts to be skilful, but based on the percentage of correct forecasts; PoP forecasts outperformed QPF's. Giorgi and Marinucci (1996) found that precipitation amounts vary with model resolution. Local effects such as surface forcing are missing in coarse resolution models. Golding (2000) describes different QPF aspects as it is in the United Kingdom (UK). Various studies were done to improve the accuracy of QPF's. The United States Weather Research Program (USWRP) has come up with a strategy to improve QPF (Fritsch and Carbone, 2004). This strategy emphasizes the need to provide user guidance in a probabilistic form. They claim that certain issues are significantly important in improving QPF. Table 1.1 contains a summary of these issues.

Initial activities	Continuing activities
Determine user needs in the forecast product	Stimulate partnership between a diverse group of scientist to broaden the spectrum of end users
Increase expertise in data assimilation	Integrate cloud-resolving model advancements with new data assimilation systems
Practice the development of data assimilation techniques	Determine the benefit of construction a retrospective archive of control model output and ensemble model output
Assess the benefit of using very-high horizontal and vertical observations	Design an optimal ensemble forecasting system
Improve physics in cloud-resolving models	Determine factors that causes a deviation in the modal characteristics of moist convection
Collect a database of high resolution precipitation properties	Assess the benefit of incorporating nonlinear methods in the Model Output Statistics (MOS) approach
Use suitable verification techniques to assess forecast performance	
Blend ensemble forecasting and statistical post-processing techniques to develop accurate forecasts	Build up a baseline of forecast quality and predictive skill
Develop nowcasting techniques to render 0-6 hour forecasts	Develop seamless forecast ranging from nowcasts to medium range to climate
Develop techniques to assimilate QPE and probabilistic QPF's into hydrologic forecasts	

Table 1.1 **A summary of important issues to improve QPF as listed by the United States Weather Research Program (USWRP).**

1.4 Statistical Post-Processing

To improve the performance of NWP model forecasts some statistical post-processing techniques (Wilks, 1995) can be used. Common used techniques are the Perfect Prognosis technique (PP) (Klein et. al., 1959) and Model Output Statistics (MOS) (Glahn and Lowry, 1972). Both these techniques search for suitable statistical relationships between the predictors and the predictand. The relationships are determined by statistical methods such as linear/non-linear regression and multiple discriminant analysis. The determined relationships are then applied to the output of a NWP model to yield forecasts of the predictand. The main difference between the PP and the MOS technique is while both these techniques use predictors derived from observed weather elements, the MOS technique also use predictors derived from model forecasts. PP was the first statistical approach to be applied to NWP model output and makes no bias corrections on the assumed "perfect" model data. The more preferred MOS technique accounts for model biases by including model data in the development of the statistical relationships. These techniques have some pros and cons to consider (Table 1.2 and Table 1.3)

Perfect Prognosis	
Advantage	**Disadvantage**
Requires only observations	Systematic model errors not removed
Improves as raw model forecasts improve	Important derived model parameters can not be used
Use of multiple predictors	Deterioration of model forecast with lead time can't be accounted for

Table 1.2 Advantages and disadvantages of Perfect Prognosis.

MOS	
Advantages	**Disadvantages**
Removes systematic model errors and lead-time accuracy deterioration	Requires an extensive data set of historical model data
Select model variables with useful information	Equation are model dependent
Use of multiple predictors	Modify when model change
Better skill for longer range forecasts	Precipitation requires regional equations

Table 1.3 Advantages and disadvantages of Model Output Statistics.

MOS performs better than Perfect Prognosis when the underlying functions are presumed to be linear because it renders smaller bias, variance, and mean squared error (Marzban et al., 2005). In MOS, statistical relationships between model variables and weather observations are determined, to be used to improve model forecasts or render forecasts not done explicitly by the model (AMS Glossary). There are also non-linear MOS techniques which include NN. Various studies were done using NN's to render enhanced forecasts such as significant thunderstorm forecasts (McCann, 1992), severe-hail size forecasts (Marzban and Witt, 2001), short-range visibility forecasts (Pasini and Potestà (1995), Marzban (2005)), convective airmass classification (Pankiewicz, 1997), cloud identification (Pankiewicz, 1995), wind damage forecasts (Marzban and Stumpf, 1998), ozone forecasts (Narasimhan et. al., 2000), tornado diagnoses (Marzban and Stumpf (1996), Marzban (2000)) and most importantly temperature and rainfall forecasts.

Schoof and Pryor (2001) used circulation indices and cluster frequencies to downscale daily Tmax and Tmin and daily and monthly precipitation totals using regression techniques and NN's. The temperature models performed relatively well with the NN model superior to the regression models while both the precipitation models failed to represent the variability of precipitation. Coulibaly et. al (2004) compared generated daily temperature and precipitation forecast by NN's with those generated by multiple linear regression. They found that the NN is a more effective method than the regression model. Marzban (2003) used temperature forecasts from The Advanced

Regional Prediction System (ARPS) as input to a NN and found that the NN forecasts improved on those of the model in terms of a variety of performance measures at 31 different stations. The Mean Square Error (MSE) was reduced by 40%, the bias was reduced by 70% and the variance was reduced by 20%.

Silverman and Dracup (2000) showed that a NN can be used to make long-range precipitation predictions as well as in the identification of important parameters in long-range precipitation prediction. Hall et. al (1999) used the Eta model forecasts and upper air soundings as input to a NN which rendered good forecasts for both the probability and amount of precipitation. The sharpness of the PoP forecasts is remarkably high. Olsson et. al. (2004) tried to overcome the problem of forecasting the intermittency and variability of precipitation accurately. They used two different NN's in series of which the first network determines the rainfall occurrence and the second to determine the intensity of rainfall for rainy periods only. This method improved on the fraction of zero rainfall days which represents the intermittency aspect but still failed to capture extreme rainfall events. The second method they used was to set up a NN that categorizes rainfall into intensity categories which showed no significant improvement but can be useful for probability forecasting.

Applequist et. al. (2002) compared different methods to generate Probabilistic Quantitative Precipitation Forecasts (PQPF). Five methods were included, namely, linear regression, discriminant analysis, logistic regression, NN's, and a classifier system. Results showed that the classifier system has greater skill than linear regression at higher thresholds, and that discriminant analysis and NN's gave mixed results. They also concluded that logistic regression shows the greatest skill. Yuan et. al. (2005) used a 3-layer feed forward NN to calibrate PQPF's from the Regional Spectral Model (RSM) ensemble system. This method proved to correct conditional wet biases, and improved forecast skill. An advantage of using statistical post-processing techniques is that it is tailor-made for the particular region it was developed for.

Due to the complex nature of precipitation regarding its high variability in space and time it was decided to employ a non-linear statistical technique such as the NN. Both linear and non-linear NN's will be trained to forecast temperature of which the best

performing network will be selected. A linear NN is obtained when the network contains no hidden layer.

1.5 Forecasting by the forecaster

The role of a forecaster in the forecasting process is indispensable (Snellman 1977; Doswell and Brooks 1998). The task of a forecaster can be characterized by making decisions in the presence of uncertainty. Sources of uncertainty originate from initial condition uncertainty, model uncertainty, how to make use of a flood of information in the appropriate manner, and the duality of error (false negatives can only be reduced by false positives in a Yes/No forecast) Development of objective forecasting techniques has become more important and are employed as a guidance to human forecasters. Forecaster must be careful not to focus their training on automated forecast products only but first of all on fundamental scientific understanding and physically based conceptual forecast models (Pliske et. al., 1997). Olssen et. al (1995) stated that the ability of the forecaster to forecast accurate QPF's is more dependent on numerical model guidance than on the number of years of experience.

At the South African Weather Service (SAWS) Tmin and Tmax forecasts are forecast twice daily, for the present day and the next day respectively. These forecasts are done at 96 stations across South Africa. 24-hour rainfall forecasts are created daily and are made available on maps for the next day. The rainfall forecasts are presented in the form of a region presenting a certain percentage of stations to receive rainfall. Verification of temperature and precipitation forecasts by forecasters at SAWS was done by Banitz (2001). Results show that Tmin's are more difficult to forecast during the winter and the Absolute Error (AE) increases with lead time, with the present day forecast having an AE of 1.75º and the next day forecast having an AE of 2.3º). Tmax's are more difficult to forecast in the summer. There is a slight increase in the AE as lead time increase (2. 25º to 2.3º). Forecasts of rainfall occurrence have a fairly low bias. The hit rate for both summer and winter is more than 60%, but less accuracy are evident in periods of heavy rainfall. Rainfall tends to be over-forecast and is slightly better than chance.

1.6 Aim of Study

Variations in dynamical and thermodynamical processes produce irregularities in the atmospheric system which is non-linear in nature (Shukla, 1985and Tsonis, 2001). Dynamical systems are chaotic, but an atmospheric quantity which is apparently varying randomly in time may contain information about the underlying chaotic process (Tsonis and Elsner, 1989) such as the occurrence of precipitation for example.

Given the importance of rainfall and temperature forecasts and the need for some form of model post-processing to help improve the skill of these forecasts, the objectives of the research are to:

a. study previous work on objective forecasting techniques, especially the structure and use of the artificial NN
b. collect a data set of model parameters that stretch over a significant period of time and determine the best predictors from the set of model parameters to use in the training of a neural net to forecast rainfall occurrence.
    - *This was done by determining the length of archived data available for each accessible model data set. Model data of the NCEP Ensemble Prediction System (EPS) was used as input to the temperature neural net systems and NCEP Eta data were used as input to the precipitation neural net systems. The number of available model parameters of the Eta model was reduced by selecting the best predictors for precipitation as predictand by using a technique known as bivariate analysis.*
c. develop NN system to forecast rainfall occurrence and intensity and a system to forecast Tmin and Tmax
    - *This was done by a thorough study of the NN structure to determine an appropriate system for each type of forecast. The Stuttgart Neural Network Simulator (SNNS) software was used to set up each network.*
d. verify neural net forecasts for both temperature and rainfall

- *This was accomplished by selecting relevant verification measures that can assess the performance of probability as well as quantitative forecasts.*

1.7 Hypotheses

The hypotheses to be tested are:

1. The application of NN's on model forecasts improves 14-day Tmin and Tmax forecasts in comparison to temperature forecasts rendered by the model and by the forecaster.
2. The application of NN's on model forecasts provides 24-hour PoP forecasts which show improved forecast resolution relative to the human forecaster PoP forecast.
3. The application of NN on model forecasts improves on the 24-hour QPF by the model in terms of bias and variance.

The dissertation consists of five chapters. **Chapter one** describes the need for accurate temperature and precipitation forecasts. A short introduction is given on statistical post-processing techniques and studies done on the techniques. The forecasters' needs and role in the forecasting process are explained. The objectives and hypotheses are introduced in this chapter. **Chapter two** introduces the structure and capabilities of the artificial NN. **Chapter three** discusses the NWP model data and observations used in the research. The different study domains of the temperature and rainfall forecasts are discussed. The methodology used to accomplish the research aim is explained in detail. In **Chapter four** the results of the temperature and the precipitation NN's will be discussed. **Chapter five** introduces the best predictors found to forecast precipitation forecasts and discuss the results explained in chapter four. It also explains the use of PoP's in the decision-making process. **Chapter six** summarizes and concludes the findings of this research by referring back to the hypotheses made.

CHAPTER 2

NEURAL NETWORKS

2.1 Introduction

The structure and different types of NN's are discussed in this chapter. A successful NN needs to be trained on a comprehensive set of data consisting of different forecast variables and the relevant observations of the desired parameter to be forecast by the NN. A learning algorithm is used to train the underlying function of the target parameter in terms of the given input variables or predictors. Numerous references describing NN's can be found on a website by Sarle (1997). Other important references that give an introduction to NN's are Bishop (1995) and Ripley (1996).

2.2 Structure of a neural network

There is no perfect definition for a NN as existing definitions do not suit the diverse group of users simultaneously. To name a few, Haykin (1994), defined that:

"A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. *Knowledge is acquired by the network through a learning process.*
2. *Interneuron connection strengths known as synaptic weights are used to store the knowledge."*

According to Nigrin (1993):
"A neural network is a circuit composed of a very large number of simple processing elements that are neurally based. Each element operates only on local information. Furthermore each element operates asynchronously; thus there is no overall system clock."
Zurada (1992), stated that:

"Artificial neural systems, or NN's, are physical cellular systems which can acquire, store, and utilize experiential knowledge."

The NN consists of three different layers, the input layer, one or more hidden layers, and the output layer (Figure 2.1). The input layer nodes receives the input parameters which consist of model variables and/or observations, indicated by $x_1,x_2,x_3,....,xi$. It should be noted that too much input parameters can lead to overtraining or complicate the learning of the relationship between the input and the desired output.

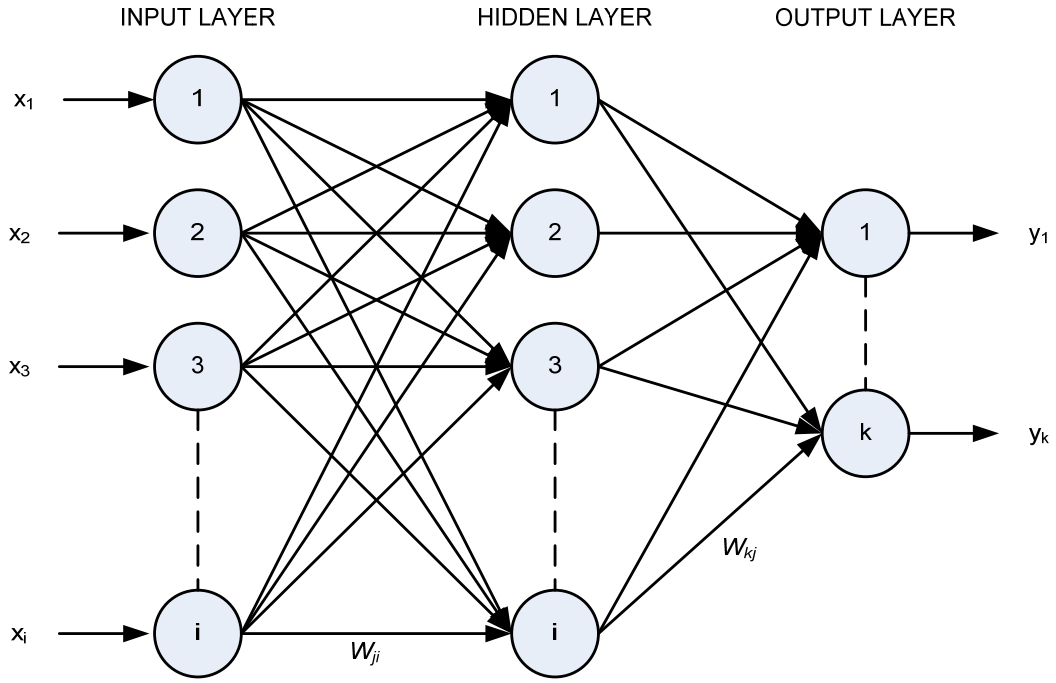

Figure 2.1 The structure of a 3-layer feedforward neural network.

The input parameters are also known as the predictors of the NN and the output variable is defined as the predictand. The input nodes are completely connected to the nodes $h_1,h_2,...,h_j$ in the hidden layer. Each connection has an independent weight, w attached to it. The number of hidden nodes and layers are dependent upon the function to be solved and are determined by the trial and error while training the network. The input parameters, x_i, are fed through the input nodes to the next level where each of the input values are multiplied by the relevant weight w on each connection to create a weighted input value. The weighted inputs wx_i at each hidden node or also called the artificial neuron, are summed. Artificial neuron refers to the similar computations that

take place in the human brain. An activation function *a,* is applied to each of the summed values from which it is determined if the neuron is activated or stays inactive (Equation 2.1). Activated neurons contribute to the prediction of the desired output or targets, $y_1,\ldots,y_k$. A similar process takes place between the hidden nodes and the output nodes (Equation 2.2).

$$h_j = a_j \sum w_{ji} x_i \tag{2.1}$$

$$y_k = a_k\left(\sum w_{kj} h_j\right) = a_k\left(\sum w_{kj} a_j\left(\sum w_{ji} x_i\right)\right) \tag{2.2}$$

2.3 Types of neural networks

There are two kinds of learning algorithms that are used in NN's, namely supervised and unsupervised learning (Sarle, 1997). During supervised learning target values are known during the training process. Opposite is true for unsupervised learning where the NN is trained without target values. The former learning algorithm is usually used in regression analysis, while the latter are used in data clustering.

Feedforward and feedback networks are two different kinds of network topologies Sarle, 1997). In feedforward networks there are no cycles present in connections between units and information travels only one way from the input layer towards the output layer. There is no feedback from the output layer to any of the other layers. The feedback network is therefore simple to train and is used to associate inputs with outputs. These networks are usually used in pattern recognition applications. Feedback or recurrent networks have cycles and therefore information travels from the input layer to the output layer and back again. This type of network is powerful but can develop into a very complex structure and difficult to train effectively. The feedback network is dynamic and changes state continuously until it reaches an equilibrium point.

Both categorical and quantitative data sets are accepted by a NN. Categorical variables can be in the form of symbols but should be encoded to numbers before it can be feeded into a network. For example a Yes/No forecast can be encoded to binary numbers 0 and 1. Training a network with categorical target values is called classification. Quantitative variables are numerical measurements of some elements, like for example temperature in degrees Celsius. Training a network using supervised learning with quantitative target values is called regression.

2.4 Training a neural network

During the learning process a NN needs a training data set as well as a validation data set to train successfully. The training data set is used to learn the network parameters or in other words to set the weights of each input parameter. These parameters need to be tuned to set the correct network architecture and this is where the validation set comes in. Usually a third data set namely the test set are also needed to assess the performance of the network (Ripley, 1996).

There are various functions involved in a NN (Sarle, 1997). Combination functions combine the values from units that are fed via synaptic connections with the values of the following units. Functions that transform the net input to generate unit activation values are called activation functions. If the activation function is a linear function then the network will also be a linear model. A common choice of activation function is sigmoid functions like the logistic, hyperbolic/arc tan, and Gaussian function. The hyperbolic and arc tan functions contain both positive and negative values which provide faster training than functions with only positive values. Sigmoid functions like for example the logistic function (Figure 2.2) is preferable over threshold functions or step function (Figure 2.3) when applied at hidden units because it is easier to train from.

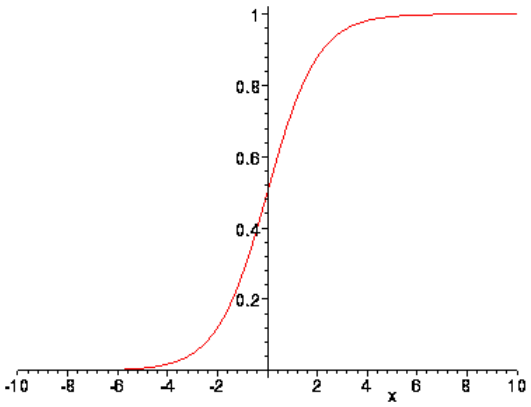

Figure 2.2 **The logistic function is an example of a sigmoid function.**

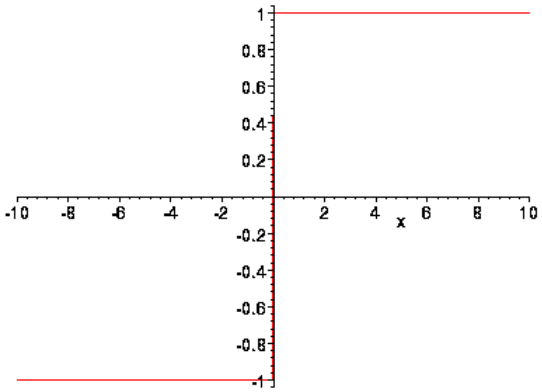

Figure 2.3 **The stepwise or threshold function that is bounded by the -1 and 1 thresholds on the y-axis.**

Table 2.1 shows some sigmoid and step functions and the relevant ranges which are available to use as activation functions.

Function	Definition	Range
Identity	x	(-inf,+inf)
Logistic	$\frac{1}{1-e^{-1}}$	(0,+1)
Hyperbolic	$\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	(-1,+1)
Exponential	e^{-x}	(0, +inf)
Softmax	$\frac{e^{x}}{\sum_{i} e^{x_i}}$	(0,+1)
Unit sum	$\frac{x}{\sum_{i} x_i}$	(0,+1)
Square root	$\sqrt{x}$	(0, +inf)
Sine	$\sin(x)$	[0,+1]
Ramp	$\begin{cases} -1 & x \le -1 \\ x & -1 < x < +1 \\ +1 & x \ge +1 \end{cases}$	[-1,+1]
Step	$\begin{cases} 0 & x < 0 \\ +1 & x \ge 0 \end{cases}$	[0,+1]

Table 2.1 A list of available activation functions to generate unit activation values.

Activation functions at output units should be choose according to the distribution of the target values. Thus data that is:

- binary, should use logistic functions.
- categorical, should use the softmax function.
- continuous-valued,
    - and bounded, should use logistic or hyperbolic tan functions
    - positive and has no upper bound, should use exponential functions
    - and have no bounds, should use linear functions.

Error functions measure the discrepancy between network output and target values. Then finally the actual linear or non-linear function to be learned by the network is called the objective function.

There are two methods to learn a network using a data set of cases. In batch learning all cases are processed during training after which the weights are updated. During incremental learning weights are updated after each individual case are processed.

There are various kinds of NN's available. The most widely used network kind is the standard backpropagation network (Rumelhart and McClelland, 1986). In this network the gradient of the case-wise error function are computed with respect to the weights for a feedforward network. It is also known as the delta rule.

The Scaled Conjugate Gradient (SCG) network (Moller, 1993) is a supervised learning algorithm and a feedforward NN. The SCG forms part of the Conjugate Gradient Methods (CGM). CGM methods utilize second-order techniques where standard back propagation techniques use first-order techniques. This means that CGM makes use of second-order derivatives of the objective function to be learned and standard back propagation uses first-order derivatives. SCG therefore finds a better way to the local minimum error but at high computational cost. Standard back propagation proceeds down the gradient of the error function, while the SCG proceeds in the direction which is conjugate to the direction of the previous step. In this way the minimization of the previous step are not partially undone by the next such as it occurs when using standard back propagation and other conjugate methods. SCG is considerably faster than standard back propagation and will be utilized in this research.

With each SCG iteration k a new weight w_{k+1} are calculated by, $w_{k+1} = w_k + \alpha_k \cdot p_k$, where p_k is the new conjugate direction and α_k is the size of the step in this direction. The α_k is also a function of $E''(w_k)$ which is the second derivative matrix of the error function E and which is also known as the Hessian matrix. Other CGM's use a slow line search procedure to approximate α_k but SCG approximates s_k (Equation 2.3) which is an important component in the computation of α_k.

$$s_k = E''(w_k)\cdot p_k \approx \frac{E'(w_k + \sigma_k \cdot p_k) - E'(w_k)}{\sigma_k}\ ,\ 0 < \sigma_k << 1 \tag{2.3}$$

SCG also uses a scalar λ_k to regulate the indefiniteness of the Hessian matrix because it is not always positive definite which leads to poor performance. This is done by adding $\lambda_k \cdot p_k$ to each iteration and also adjusting λ_k after each iteration.

Some guidelines (Sarle, 1997) can be followed to ensure that the network will be able to learn the underlying function of the relationship between the input and output. The guidelines are:

a. Check for outliers in the desired output set and remove if the data are not supposed to be rare-case type of data such as in the case of tornado observations, or any other severe weather variables.
b. Standardize input data to ensure that some parameters do not get larger weights due to a larger range of values.
c. Encode categorical inputs
d. The amount of training cases should be larger that the amount of input parameters – approximately 10 times more training cases as input parameters.
e. Standardize quantitative output values to simplify the learning process.
f. The arctan activation function is recommended for the hidden layer.
g. Use the bias term at a hidden node to prevent constraining the hyperplane at this hidden node to pass only through the origin of the space created by the input values.
h. Set initial weights to random numbers.
i. Use software that consists of good conventional numerical optimization techniques.
j. Use batch training rather than incremental training.

2.5 Preprocessing data

It is recommended to standardize input and output data (Sarle,1997). Data can be standardized to have a mean value of zero and a standard deviation of one (Equation 2.4) or to have a midrange of 0 and a range 2, thus a range of [-1,1] (Equation 2.5).

$$\bar{x} = \frac{\sum x}{N}$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

$$S = \frac{x - \bar{x}}{\sigma} \tag{2.4}$$

$$midrange = \frac{\max(x) + \min(x)}{2}$$

$$range = \max(x) - \min(x)$$

$$S = \frac{x - midrange}{range/2} \tag{2.5}$$

The x is the value to be standardized, the $\bar{x}$ is the mean value of x, the σ is the standard deviation of x, and S is the standardized value. The data will then have a Gaussian distribution. Standardizing input values decrease complexity in data and training will take place at a faster rate. Standardizing output values reassure good initial weights. Apart from standardizing data there is also the option to rescale or to normalize data. Rescaling consists of adding or subtracting a constant and then multiplying or dividing a constant from a parameter. Data are normalized by scaling the values to lie between 0 and 1.

2.6 Generalization

Generalization are reached when the output values rendered by the trained network are significantly close to the desired output after input from a set of data were introduced to the network that was not utilized in the training process (Sarle, 1997).

To reach good generalization the following conditions should be present:

a. Input parameters used to train the network must be related to the predictand in such way that there exists a mathematical equation to describe the relationship.
b. The underlying function to be learned must approach a smooth line. Small variations in input values assure small variations in output values.
c. The training data set must consist of a reasonably large amount of cases. There should be at least ten times the amount of training cases to the amount of input units.

Generalization is limited by the noise present in data. Statistical noise are variations in target values that are unpredictable from the input values, while physical noise are variations in target values that are naturally unpredictable regardless of what inputs are used. Noise also increases the possibility of overfitting. Overfitting causes excessive variance in the network output data while underfitting causes excessive bias (Sarle,1995). Overfitting can also occur when there are less than approximately 10 times as many cases as input parameters. Multi-colinearity can cause weights to become extremely large. There are some approaches to prevent under- and overfitting.

a. Use a large amount of training cases of approximately 30 times more training cases than weights.
b. Apply jittering on training data. Adding noise to data increase the amount of training cases. One should take note that too much noise will not have the desired effect.
c. Apply early stopping and/or weight decay during the training process.
d. Use Bayesian learning in which probability is used to represent uncertainty about the relationship being learned. (Neal, 1996)
e. Combine networks.

Early stopping described by Prechelt in Orr and Mueller (1998) were used in this research to prevent overtraining of the network. In this process both a training and validation data set are utilised. During training errors are calculated periodically for both data sets. As can be expected the training error and the validation error will decrease as the training iterations proceeds. At some point the validation error will stop decreasing and will actually start to increase again. It is at this point were overfitting of data are appearing and at this point where the training process are stopped. It should be noted that the validation error is not a true reflection on the performance of the trained network. To determine the true performance or in other words the generalization error, a third data set is needed which contains data that are totally independent of the training process, and is called the test set. The generalization error is then determined by feeding the trained network with cases from the test set. This method has the advantage of being very fast in detecting overfitting to such an extent that it can be applied to networks of which the amount of weights are much larger than the amount of training cases. It also has the advantage that it only needs one decision to made by yourself which in turn will have an influence on the effectiveness of the stopping process, and this is to decide on the most effective proportion of cases present in the training and validation data set. The splitting of the data set into a training data set and a validation set are known as split-sampling.

2.7 Resampling

Split-sampling is one of different resampling techniques. Other resampling techniques are cross-validation, bootstrapping and jackknifing (Efron and Tibshirani, 1993; Efron, 1982; Hjorth, 1994).

In split-sampling the full data set are split into a training set, a validation set, and an additional test set. Running the trained network with the test set provides an unbiased estimate of the performance of the network. The disadvantage of these techniques is that it limits the amount of data available to be used in the training process. Cross-validation solves this problem to a degree.

In *k*-fold cross-validation the full data set are divided into *k* subsets of approximately the same size. The network is then trained *k* times, each time leaving out one subset to be used to compute the validation error as it was done by the validation set in the split-sampling technique. The *k* validation errors are then averaged to find the generalization error. This method is superior to split-sampling because it utilizes the entire data set and all cases are used in the training process. This technique is ideal if only a small data set is available for training (Goutte, 1997). The 10-fold cross validation technique is the most popular choice under researchers. The disadvantage of this technique is the vast consumption of computer resources as one has to retrain the network *k* times.

Bootstrapping is again an improvement on cross validation but is even more computational expensive. Instead of retraining on subsets of the data as done in cross validation, retraining is done on subsamples of the data, which can amount to anything between 50 to 2000 subsamples. The subsample consists of a random sample with replacement from the full data set. Although the results of the network trained with the bootstrapping technique have low variance, it can also be extremely biased.

CHAPTER 3

DATA AND METHODOLOGY

3.1 Introduction

Predictors need to be identified from the available NWP forecast variables from which the best predictors are chosen to be able to learn a function that will be able to explain the association of the predictors with the selected observational output. The proper NN architecture needs to be selected for the training process and proper verification measures need to be selected to assess the performance of the trained network.

3.2 NWP model predictors

Input parameters used to train the precipitation forecasting network were acquired from the NCEP Eta model. This model is an eta-coordinate regional forecast model (Mesinger, et. al., 1988). The Eta model with a matching data assimilation system became operational at the SAWS in November 1993. The model underwent several upgrades until recently, but further upgrades are static at the moment due to the transformation to a new modeling system that will take place in 2006. The model is characterized by hydrostatic primitive equations with a step-mountain formulation for the vertical coordinate. It has an Arakawa E-grid on a transformed lat-lon grid and time integration is done by a split explicit adjustment scheme, the Euler backward advection scheme. Precipitation modeled by the Eta model is a product of the Betts-Miller convective- and large-scale precipitation parameterization schemes. The Eta model covers a 133 X 79 model point domain which consists of South Africa and its surrounding oceans. The longitudinal boundaries are from 13° W to 53° E and the latitudinal boundaries are from 48° S to 9° S. The Eta model had a horizontal resolution of 48 km and vertical resolution of 38 layers from August 1998, but was upgraded to a 32 km, 45 layer model, in Jul 2003. The Eta data used in this research covers the period, January 2000 to May 2005 and therefore results may be affected by the horizontal resolution change in this period.

Input parameters used to train the temperature forecasting network were acquired from the NCEP EPS. Small error in the model initial conditions grows with lead time until there is a total loss in predictability (Lorenz, 1969). Ensemble forecasts are made out of an ensemble of initial conditions. Lots of information is captured through this set of ensemble members. This ensemble of forecasts describes the chaotic character of the atmosphere. The average value of the ensemble members is a more reliable forecast then any of the members on its own. Computer generated products that are derived from the EPS renders 14-day forecasts which form part of the value-added public forecasts that are issued at SAWS on a daily basis. EPS data are downloaded daily from NCEP from which daily maps are created. These maps are used by the forecasters to assemble text descriptions of the 8 to 14-day forecasts. The EPS forecasts have a horizontal resolution of 2.5 km and cover a domain (7 X 6 model points) over South Africa from 17.5° E to 32.5° E and from 35° S to 22.5° S. The EPS data used in this research covers the period, November 2000 to Jul 2004.

Eta model variables that were available for this research are listed in Table 3.1 and EPS model variables that were available are listed in Table 3.2.

06Z, 12Z, 18Z, 24Z and 30Z Eta forecasts of:
2-D model variables: 1. Mean sea level pressure 2. Soil water 3. Lifted index 4. 4-Layer lifted index 5. Precipitable water 6. Low cloud cover 7. Mid-cloud cover 8. High cloud cover 9. Total cloud cover 10. Surface height 11. Surface temperature 12. 2-meter temperature 13. 2-meter dew point temperature 14. 2-meter relative humidity 15. 10-meter horizontal wind component 16. 10-meter vertical wind component 17. Surface Convective Available Potential Energy (CAPE) 18. Surface Convective Inhibition (CIN) 19. 180-0 mb CAPE 20. 180-0 mb CIN **3-D model variables at 1000, 850, 700, 500, and 250 hPa levels:** 21. Geopotential height 22. Temperature 23. Potential temperature 24. Relative humidity 25. Dew point temperature 26. Specific humidity 27. Vertical velocity 28. Horizontal wind component 29. Vertical wind component
24-hour accumulated Eta variables:
30. Total precipitation 31. Convective precipitation

Table 3.1 Available variables of the NCEP Eta model.

00Z and 12Z EPS forecasts for:
1. Mean sea level pressure 2. Relative humidity at the 700 hPa level 3. 10-meter horizontal wind component 4. 10-meter vertical wind component 5. Geopotential height at the 500 hPa level
24-hour EPS forecasts for:
6. Precipitation 7. Tmin 8. Tmax

Table 3.2 Available variables of the NCEP Ensemble Prediction System (EPS).

Statistical relationships between predictors and the predictand can be stationary or might fluctuate in time as in the case of climate change. If there is no change in the mean, variance and auto-correlation of the relationship, it implies stationarity. The assumption is made that all the predictor-predictand relationships in this study is stationary to simplify the development and training of the NN systems.

3.3 Observations

Eight rainfall stations that are approximately located in the Vaal Dam catchment area were chosen as locations at which precipitation networks were trained. The reason for including a catchment in the research is to investigate the possible hydrological advantages that can be initiated from the results. The eight stations include Ermelo, Frankfort, Bethal, Standerton, Vrede, Bethlehem, Van Reenen, and the Royal Natal National Park. Although the stations at Van Reenen and the Royal Natal National Park are actually not part of the Vaal Dam catchment, they were still included to test the ability of NN to forecast precipitation at locations with complex topography. The location of each of the rainfall stations is indicated on Figure 3.1. Terrain elevation measured in meters is indicated by the color bar. Precipitation is measured with a rain gauge for a 24-hour period between 08:00 SAST and 08:00 SAST the next day. Rain gauges used by the SAWS consist of cylinder with a sharp rim that has a diameter of 127 mm and is mounted 1.2 m above the ground.

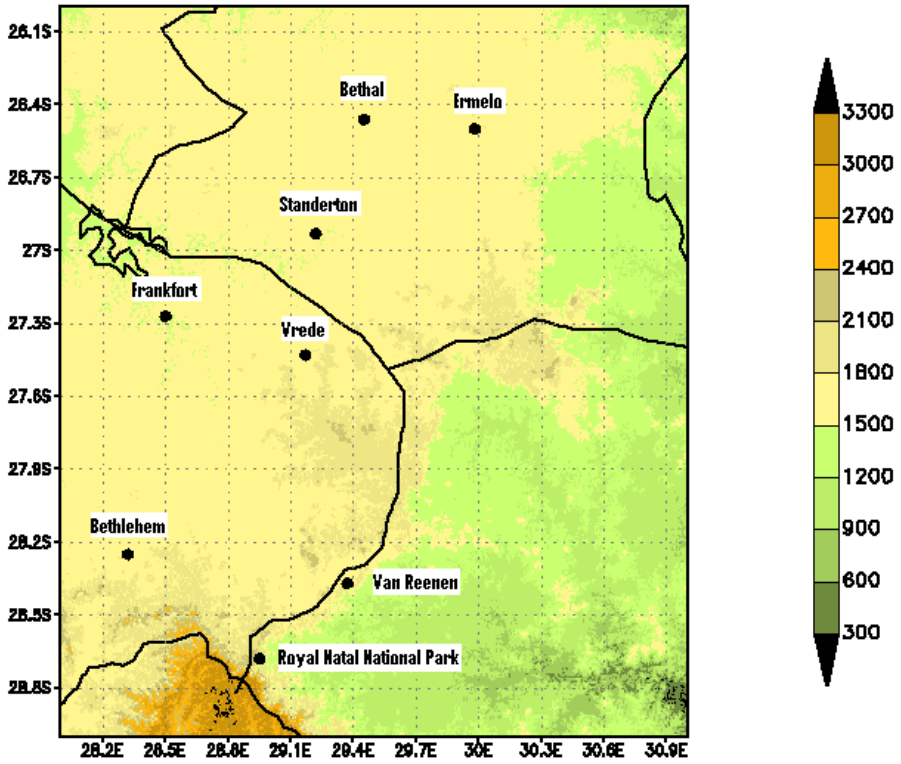

Figure 3.1 The locations of eight rainfall stations in the Vaaldam Catchment area.

Temperature observations were collected for one coastal station and one inland station, namely Cape Town and Pretoria. This is to test the model's ability to reproduce maritime influence on coastal stations and its ability at higher altitudes. If the model fails in any of these situations then it can be determined if a NN has the ability to enhance the temperature forecasts.

3.4 Best predictors

When the number of predictors in a data set exceeds the number of cases available in the set the statistical model will rapidly overfit the function. One way to overcome this problem is to reduce the amount of predictors (Marzban et al., 1999). There are two 2 methods. One method is the Principle Component Analysis (PCA) method which keeps only the linear or non-linear combinations of predictions that explain most of the variance in the data set. This method is inappropriate in selecting the best predictors because no reference is made to the dependent or target value. Another method is to take the linear or non-linear combinations that make up a set of best predictors of the event.

Through a bivariate approach one predictor gets analyzed at a time. This approach guarantees no contamination by problems arising in data with multiple variables, such as multicollinearity. Three bivariate approaches are discussed. The first two approaches are linear while the third approach is non-linear. Predictors that got penalized by the linear approaches might show significant predictive strength through the non-linear approach due to a possible non-linear relationship with the predictand.

a. Pearson's linear correlation coefficient r_{xy}

This approach measures the level of association between two variables *x* and *y,* assuming that *x* and *y* are continuous (Equation 3.1).

$$r_{xy} = \frac{\sigma_{xy}^2}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

$$\sigma_{xy}^2 = \overline{xy} - \left(\bar{x}\right)\left(\bar{y}\right)$$

$$\sigma_x^2 = \overline{x^2} - \left(\bar{x}\right)^2$$

$$\sigma_y^2 = \overline{y^2} - \left(\bar{y}\right)^2 \tag{3.1}$$

The σ_{xy}^2, σ_x^2, and σ_y^2 are the covariance and variance terms, respectively. The correlation values range between -1 and 1. The correlation between each predictor is calculated to determine the collinearity between the predictors. If predictors are collinear it is advised to remove one of the predictors to exclude multiple relationships that are identical to each other. The predictive strength of each predictor is determined by the amount of correlation between the relevant predictor and the target value or predictand. A high positive or negative correlation coefficient points out that the predictor has noteworthy predictive strength in regard to the predictand. Small coefficients do not necessary show on low predictive strength but can be indicating a

nonlinear association as Pearson's correlation coefficient is after a linear measure of association.

b. Maximizing measures of performance

Marzban (1998) found that some measures of performance are biased in that they induce under- or overforecasting in rare-event situations, like the occurrence of tornadoes. One measure of skill (Equation 3.2) that is relatively unbiased is the Heidke skill score (HSS). The Critical Success Index (CSI) and the Likelihood Ratio Chi-square (LRC) are moderately biased measures (Equations 3.3 and 3.4, respectively). CSI measures accuracy and LRC measures skill. LRC also has the ability to differentiate between the predictive strengths between predictors and the predictand.

$$HSS = \frac{2(C_1C_4 - C_2C_3)}{N_0(C_2 + C_4) + N_1(C_1 + C_3)} \tag{3.2}$$

$$CSI = \frac{C_4}{C_2 + C_3 + C_4} \tag{3.3}$$

$$LRC = -\sum_{i=1}^{4} C_i \log \frac{E_i}{C_i}$$

$$E = \frac{1}{N}\begin{Bmatrix} (C_1 + C_2)(C_1 + C_3) & (C_1 + C_2)(C_2 + C_4) \\ (C_3 + C_4)(C_1 + C_3) & (C_3 + C_4)(C_2 + C_4) \end{Bmatrix} \tag{3.4}$$

The C_1, C_2, C_3, and C_4 are the four elements of a contingency table, which indicates the number of correctly classified non-events, the number of misses, the number of false alarms, and the number of hits, respectively. Each case of a predictor is dichotomized to a 2X2 contingency table by a certain decision threshold. The matrix E contains the expected values of the events which are calculated from the elements of the contingency table. N_1 and N_0 indicate the number of events and non-events, respectively and N is the total number. It is also useful to look at other measures of

forecast quality, such as the Probability of Detection (POD), the False Alarm Ratio (FAR), and the bias (Equations 3.5-3.7).

$$POD = \frac{C_4}{N_1} \tag{3.5}$$

$$FAR = \frac{C_2}{C_2 + C_4} \tag{3.6}$$

$$bias = \frac{C_2 + C_4}{N_1} \tag{3.7}$$

c. Probabilistic Approach

Another approach to determine the importance of a predictor is to examine the posterior probability $P_1(x)$ which is the conditional probability of an event at a given value of a predictor *x* (Equation 3.8). The posterior probability is a function of the conditional frequency distributions of the events and non-events.

$$P_1(x) = \frac{N_1(x)}{N_1(x) + N_0(x)} \tag{3.8}$$

Equation 3.8 is also known as the Bayes' theorem. If $P_1(x)$ changes significantly with *x,* it is an indication that this predictor has significant importance. The correlation ratio η (Croxton and Crowden, 1955) combines the multiple information that can be found in a plot of $P_1(x)$ into a single scalar value to simplify the process of finding the best predictors. Correlation ratio can be described as a non-linear generalization of the linear correlation coefficient. The square of the correlation ratio is given in Equation 3.9.

$$\eta^2 = \frac{N_0 + N_1}{N_0 N_1} \sum_x \left[N_0(x) + N_1(x)\right]\left[P_1(x) - p_1\right]^2 \tag{3.9}$$

The p_1 component in Equation 3.9 is the priori or climatological probability of an event which is given by $p_1 = N_1/(N_0 + N_1)$. Some of the information in the posterior probability is lost when it is refined to the scalar value of the correlation ratio and it is recommended to examine both measures in the analysis.

3.5 Neural Network architecture

Precipitation has more complicated characteristics than temperature and therefore difficult to represent with a model due to its intermittency and variability. There are three types of precipitation networks trained for this research. The first precipitation network (NN_{event}) will be trained to predict the occurrence of precipitation in a binary format of 0 or 1 which indicates a non-event or an event respectively. The second precipitation network (NN_{amount}) will be trained to predict precipitation amount in millimeters. NN_{amount} will only be trained for cases on which precipitation was observed to eliminate the effect of intermittency (Olsson et al, 2004) and will only be utilized to forecast precipitation amount if NN_{event} forecasts a precipitation event. After finding the predictors with the highest predictive strength towards the precipitation predictand, these predictors were scaled in such manner that some of the variables do not obtain more weight due to the range in which the values are located. For example the values of geopotential height are in a much larger scale than the values of the wind components. The predictors of NN_{event}, and NN_{amount} were standardized by Equation 4 to have a mean value of 0 and a standard deviation of 1. The observations of the predictand of NN_{event} were converted to binary values of 0 for precipitation less than 0.1 mm and values of 1 for precipitation larger and equal to 0.1 mm. The observations of the predictand of NN_{amount} were scaled to range [-1; 1] by using Equation 5. The precipitation observations will also altered by logarithmic transformation before rescaling to test the possibility to eliminate some of the variability in this way. Logarithmic transformation stabilizes the variance of a sample (Bland, 2000) and positively skewed frequency distributions are normalized. After all statistical techniques works best with frequency distributions that are single-peaked and symmetric. The activation function used at all the nodes of NN_{event} is the logistic function (Equation 3.10) to ensure network output between 0 and 1.

$$a_i = \frac{1}{1+e^{-x_i}} \tag{3.10}$$

The activation function of NN_{amount} at all the nodes is the arc tan function $a_i = \tanh(x_i)$ to ensure network output with values between -1 and 1. The kind of network used is the SCG network due to its ability to train rapidly. The kind of network used is the SCG network due to its ability to train rapidly. The amount of hidden nodes was varied [0, 2, 4, 8, 16, 32] with each training procedure to determine the optimum amount. The full data set were divided in 10 subsamples of approximately 60 cases each to apply 10-fold cross validation.

NN_{event} is trained through 5 trials from which the best generalization error are chosen. Each of the 10 subsamples is used in a training procedure which consists of 2000 epochs. The training MSE and validation MSE are computed after each epoch. Both these errors decrease as the network is learning the underlying relationship. As soon as the validation MSE increases in size, the network is on its way to overfit the function. The minimum validation MSE that was reached before it increased again is called a local minimum. The validation error curve has more than one local minimum and the smallest of the minima is called the global minimum. Therefore some stopping criterion is necessary which will yield the best possible generalization error. The stopping criterion used for the precipitation networks is called the generalization loss *GL(t)* (Equation 3.11).

$$GL(t) = 100 \cdot \left(\frac{E_{va}(t)}{E_{opt}(t)} - 1 \right) \tag{3.11}$$

$E_{va}(t)$ is the validation MSE and $E_{opt}(t)$ is the lowest $E_{va}(t)$ reached up until the last epoch trained. As soon as *GL(t)* exceeds a certain threshold α training is stopped. For the precipitation network the α is set to 2. Training is also aborted if the ratio, *(validation MSE)/(training MSE)* is larger than 2.5 because this shows that the function that was learned from the training subsample does not explain the trend in

the validation subsample sufficiently. The generalization error for each of the five trials are determined by summing the validation errors calculated after training each of the 10 subsamples, and dividing the sum with the amount of subsamples. The trial with the lowest generalization error is taken to be the optimal trained network.

The EPS forecasts have only eight different predictors available for the temperature network and therefore no effort were made to reduce the amount by selecting the best predictors. The predictors and the observations of the predictand were scaled to be enclosed in the range [-1; 1] by the following equation,

$$x_{scaled} = \frac{x - \bar{x}}{\max_{i=1}^{n} \left| x_i - \bar{x} \right|} \tag{3.12}$$

where n denotes the amount of x values. The activation functions of the input, the hidden and the output nodes are the arc tan function which ensures that the network output also range between -1 and 1. The kind of network used is the again the SCG network. The network was trained with different amounts of hidden nodes [0, 2, 4, 8, 16] to determine the amount of hidden nodes for optimal training. Split-sampling was used to enhance network generalization. The full data set was divided into three subsamples. The training set contains 75% (± 1000 cases), the validation set contains 15% (± 200 cases), and the test set contains 10% (± 100 cases) of the full data set. Two temperature networks were trained for each station to render separate Tmin and Tmax forecasts. The predictors of the Tmin forecasts consists of all the 00:00z forecasts and the predictors of the Tmax forecasts consist of all the 12:00Z forecasts.

The temperature networks are trained for an amount of trials with each trial having a maximum of 1000 training epochs. An epoch is one cycle through the entire set of cases. The training MSE and the validation MSE are calculated after each set of 500 epochs, using the training and validation subsamples respectively. If the validation MSE starts to increase with trials then training is aborted. Training is also aborted if the ratio, *(validation MSE)/(training MSE)* is larger than 2.5 because this shows that the function that was learned from the training subsample does not explain the trend

in the validation subsample sufficiently. Training for a trial is stopped automatically after a maximum of 1000 epochs. The maximum number of epochs allowed in the training procedure depends on the rate at which the network is trained to ample performance. The test MSE is calculated after each trial to determine the performance of the trained network on the test subsample which contains cases that are independent of the training procedure. The test MSE indicates the networks ability to forecast temperature for future cases. Five trials were done for this research from which the best performing network per station are selected according to test MSE values.

3.6 Performance assessment

Performance assessment is done to determine the extent of enhancement that can be reached by applying NN's on model output. The assessment is done by comparing model and NN forecasts against observations.

The forecasts and the observations of the temperature networks and NN_{amount} is continuous data. Verification on continuous forecasts against continuous observations can be done by plotting probability distribution (p(x) and p(f)) graphs which measure forecast sharpness. Conditional probability (p(x|f) and p(f|x)) graphs can be plotted which are used to create a reliability curve of E(x|f) and E(f|x). Frequency distributions of the residuals determine the distribution of errors. The MSE's of both data sets can be calculated from where it can be decomposed which determines the bias and variance contributions to the total error.

NN_{event} output consists of probabilistic forecasts while the observations are in binary format. The discrimination capability of the forecasts can be determined by plotting the histograms of p(f|x=0) and p(f|x=1). The sharpness or degree of agreement between the forecast and observational frequency of events can be measured by plotting the reliability curves of p(x=1|f) and p(x=0|f). Another way to determine the sharpness of the forecasts is by plotting the refinement plot p(f|x). To determine if the forecasts are skillful and reliable simultaneously, the reliability plots can be extended to an attributes diagram by adding some boundary lines. Plotting Relative Operating

Characteristic (ROC) curves also give more information on the performance of the forecasts.

CHAPTER 4

NEURAL NETWORK RESULTS

4.1 Performance Assessment

Murphy (1993) declared that a good forecast depends on three aspects namely on consistency which is determined by the forecaster's judgment and knowledge, and on quality or accuracy, and on value which determine the degree of benefit the forecast have on decision-making. For verification purposes, forecast quality is determined by nine different attributes which is accuracy, skill, sharpness, uncertainty, bias, association, discrimination, reliability, and resolution. Accuracy depends on the extent of the error between the forecast and the observation, as a lower error gives higher accuracy. Skill is determined by calculating the relative accuracy of a forecast over some reference forecast such as climatology. Sharpness describes the ability of the forecast to render extreme values. Uncertainty is determined by the variability of the observations. The association between the mean forecast and the mean observation is described by the bias attribute. The strength of the linear relationship between the forecast and the observation are described by the association attribute. The discrimination attribute shows the ability of the forecast to distinguish among observations. The last two attributes mentioned here, reliability and resolution are the main attributes that determine forecast quality.

Reliability describes the average agreement or statistical consistency between the forecast and the observation and is equivalent to the bias value. Forecasts having a lack of reliability can be improved through the application of bias correction on the forecasts.

Resolution describes the ability of the forecast to resolve the set of sample events into subsets with different frequency distributions. Resolution can be related to the standard deviation or variance of a data set. A lack of resolution is the prime concern for operational purposes because this deficiency of quality in forecasts can not be

corrected. Therefore the resolution attribute can be defined as a measure of potential forecast usefulness.

The reliability and resolution of forecasts are independent of each other and can be determined by a categorical verification score namely the Brier Score (Bs) which measures the MSE of the probability forecast. The Bs can be decomposed in three terms of reliability (rel), resolution (res) and uncertainty (unc) as shown in Equation 4.1.

$$Bs = rel - res + unc \tag{4.1}$$

The Bs is negative orientated meaning smaller values indicate better performance. The third term in the decomposed Bs equation (Murphy, 1973) is the uncertainty term which describes the variance of the observations in the verification sample. The uncertainty term specifies the inherent difficulty to forecast an event during the particular period. The relative skill of probabilistic forecasts over that of the climatology can be determined using the Bs and are known as the Brier Skill Score (Bss).

$$Bss = 1 - \frac{Bs}{Bs_{reference}} \tag{4.2}$$

The Bs of the reference forecast is the determined by the climatology of the data sample. The reliability and the resolution term have zero values for a climatology forecast therefore the Bs of the reference forecast is equal to the uncertainty term according to Equation 4.1. The Bss in Equation 4.2 can therefore be transformed to

$$Bss = 1 - \frac{Bs}{unc} \tag{4.3}$$

The skill scores for reliability and resolution are determined using the following

$$Brel = 1 - \frac{rel}{unc}$$
$$Bres = \frac{res}{unc}$$

(4.4)

The resolution or the ability of the probabilistic forecasts to discriminate between events and non-events are determined by the ROC curve. The ROC curve is created by plotting the POD values against the False Alarm Rate or Probability of False Detection (POFD) values for each probability threshold. The area formed below the ROC curve describes the skill of the forecast. If the ROC area is less or equal to 0.5 then the forecasts have no skill.

The continuous verification score namely the MSE is also a measure of the two main forecast quality attributes, because the MSE can be decomposed into the bias and the variance of the data sample (Equation 4.5), which represents the reliability and resolution attribute, respectively.

$$MSE = bias^2 + variance$$

(4.5)

Stanski (1989) stated that the reliability and resolution attributes interact in a certain manner. A forecast has no skill when the reliability as well as the resolution attribute has low values. If the forecasts have high reliability but low resolution then the forecasts are equivalent to climatology forecasts. Forecasts having low reliability and high resolution are still good forecasts but need to be improved by calibration such as bias-correction. The ideal forecast therefore has high reliability as well as high resolution.

4.2 NCEP Temperature Forecasts

An EPS is intended to make an ensemble of N forecasts available which are independent of each other out of a predicted probability distribution. Toth and Kalnay (1997) found that the NCEP EPS has a lack of spread. A consequence due to this lack of spread were illustrated by Atger (1999) who compared the performance of the

European Centre for Medium-Range Weather Forecasts (ECMWF) EPS and the NCEP EPS against each other, regarding a binary event which consist of the 500 hPa anomaly exceeding the 50 m threshold. He found that the NCEP EPS has lower skill than the ECMWF EPS especially in terms of reliability. Hamill et. al (2004) found through their experiments with the MOS approach that by applying this approach to the ensemble forecast the effective forecast lead time can be increased by several days. They also illustrate the fact that unprocessed numerical guidance have no skill and are highly unreliable such as the probability forecasts that were derived from the NCEP ensemble forecast's relative frequency.

Tennant (2006) did verification on the NCEP forecasts used at SAWS including the Tmax and Tmin forecasts. The ability to forecast temperature change of ±2 °C was verified using the Bss. Annual statistics for the year 2004 to 2005 were analysed for a coastal station located in Cape Town and an inland station located in Pretoria. This is to test the model performance at two geographical different stations as a coastal station is influenced by the ocean conditions and the inland station are influenced by different weather conditions at its higher level above the sea.

Verification results shows that the Tmin and Tmax forecasts have no skill for forecasting day 1 to 14 at the coastal station. All the Bss have positive values indicating that the NCEP forecasts are better than the climatological forecasts. The reason for the poor performance of the NCEP model at the coastal station is due to the location of the station close to the boundary between land and ocean. The NCEP model has a coarse horizontal resolution of 2.5 degrees which is approximately 250 km. The boundary between the land and ocean is accordingly not well defined. As a result a station point close to the ocean can be represented by a model grid point which is located in the ocean.

The Tmin forecasts at the inland station Pretoria showed low skill for the first seven forecasting days. The Tmax forecasts showed moderate to low skill for the first seven forecasting days.

4.3 Temperature Forecasts

Although the NN and the NCEP forecasts of temperature are rendered for 14 days ahead, only the first seven forecasting days were evaluated because the Figures 4.1 to 4.4 show the Root Mean Square Error (RMSE) of the NN forecast, the NCEP Ensemble forecast (ENS), and the Forecaster forecast (FCSTER) in relation to observations at the Pretoria and Cape Town synoptic weather station. For forecasting day 1 to 2 the FCSTER forecast for Tmin at Pretoria has the lowest RMSE value (Figure 4.1). For day 3 to 7 the ENS forecast and the NN forecast show less error than those of the forecaster. There is no significant improvement of the NN forecast over the ENS forecast and both hold a moderately constant error throughout all seven forecasting days. The RMSE of the FCSTER forecasts increased by approximately 1 degree towards the seven day forecast. The Tmax forecast at Pretoria (Figure 4.2) by the forecaster showed an increase in RMSE of approximately 2.5 degrees towards the seventh day forecast. The NN forecast and the FCSTER forecast had equally lower RMSE's than the ENS forecast for forecasting day 1. For days 2 to 7 both the NN forecast and the ENS forecast show lower RMSE than that of the forecaster with a smaller slope of increase toward day 7 of about 1.5 degrees. The NN forecast again proves to be better than the ENS forecast by 0.5 degrees RMSE.

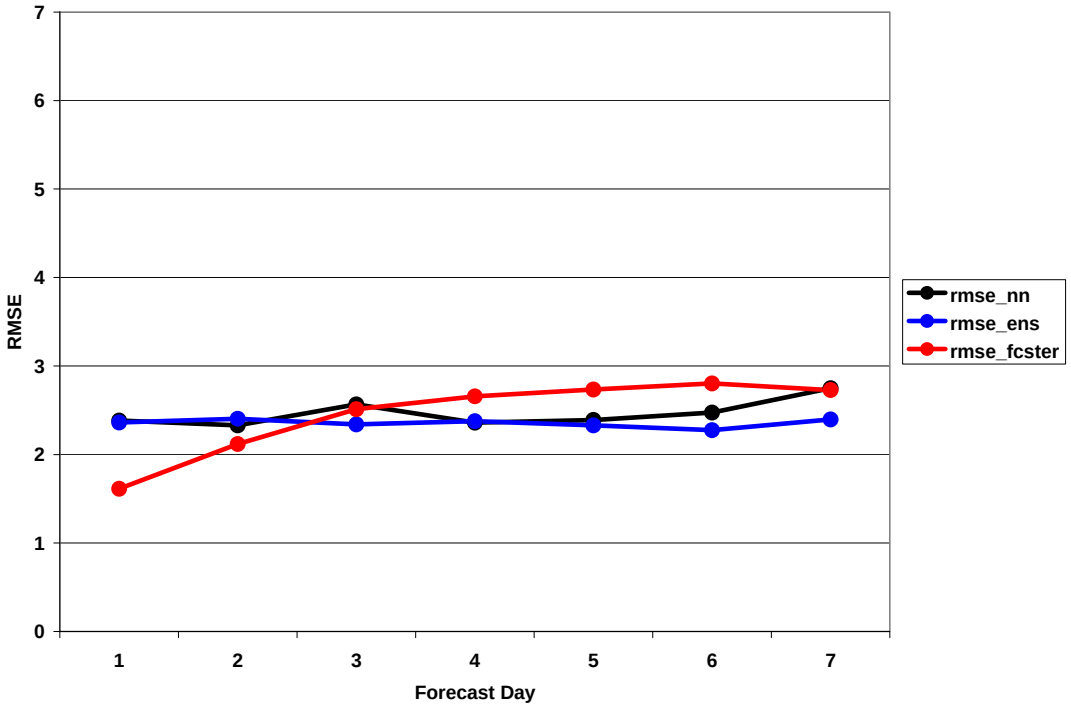

Figure 4.1 **RMSE values of the neural network forecast (NN), the NCEP ensemble forecast (ENS), and the forecaster forecast (FCSTER) for minimum temperature forecasts at the Pretoria synoptic weather station.**

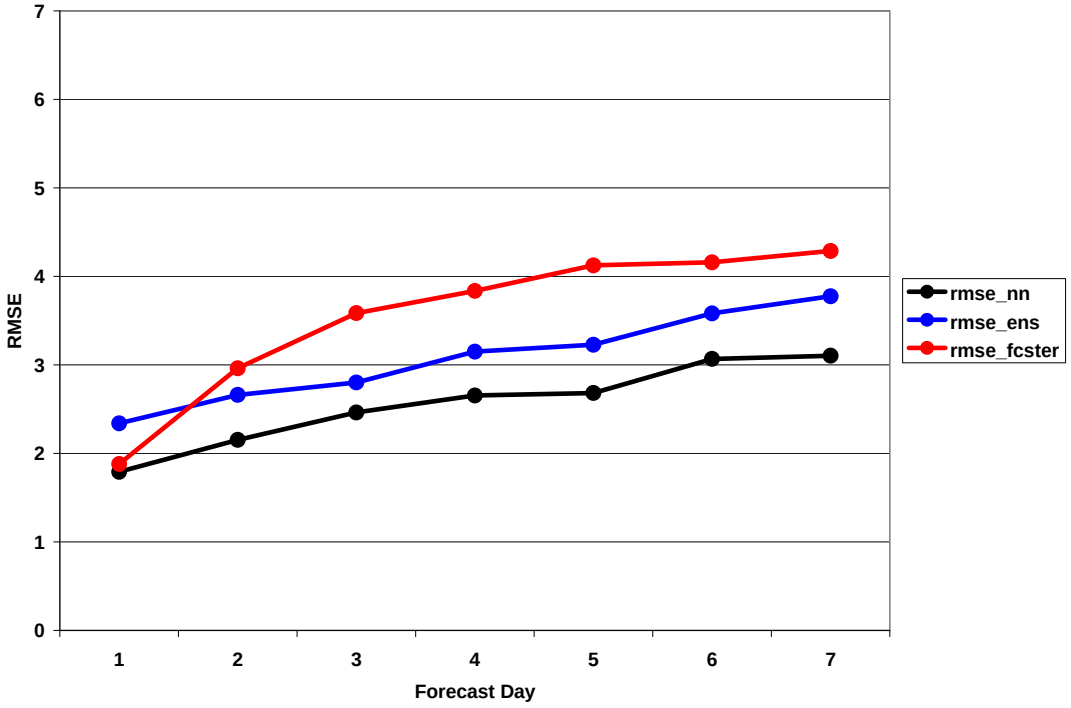

Figure 4.2 **RMSE values of the neural network forecast (NN), the NCEP ensemble forecast (ENS), and the forecaster forecast (FCSTER) for maximum temperature forecasts at the Pretoria synoptic weather station.**

RMSE's for Tmin at the Cape Town station in Figure 4.3 show (Tennant, 2006) that the ENS forecasts have no skill for forecasting days 1 to 14. The ENS forecasts show a constant RMSE of 6 degrees over the first seven forecasting days. Similarly the Tmax forecasts by the NCEP ensemble model show a constant RMSE of 5.5 degrees as shown in Figure 4.4. The Tmin temperature forecasts by the forecaster are the most skillful at day 1. At day 2 the NN forecast and the FCSTER forecast have approximately the same RMSE. For day 3 and 4 the NN forecasts are only slightly better than the FCSTER forecasts. For the rest of the period toward day 7 the NN forecasts and FCSTER forecasts have the same amount of error. For the Tmax forecasts in Figure 4.4 the FCSTER forecasts are again the best forecasts for day 1. At day 2 the NN forecasts are slightly better and then from day 3 to day 7 the NN forecasts show the best performance by an improvement of approximately 1 degree over the RMSE of the FCSTER forecasts.

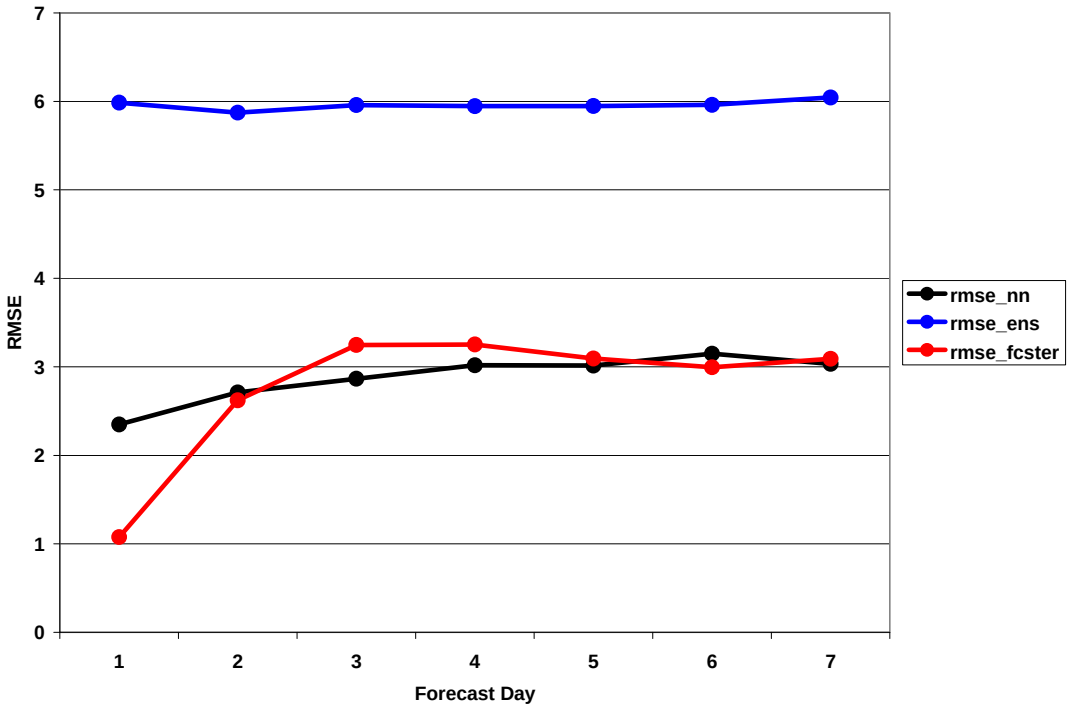

Figure 4.3 **RMSE values of the neural network forecast (NN), the NCEP ensemble forecast (ENS), and the forecaster forecast (FCSTER) for minimum temperature forecasts at the Cape Town synoptic weather station.**

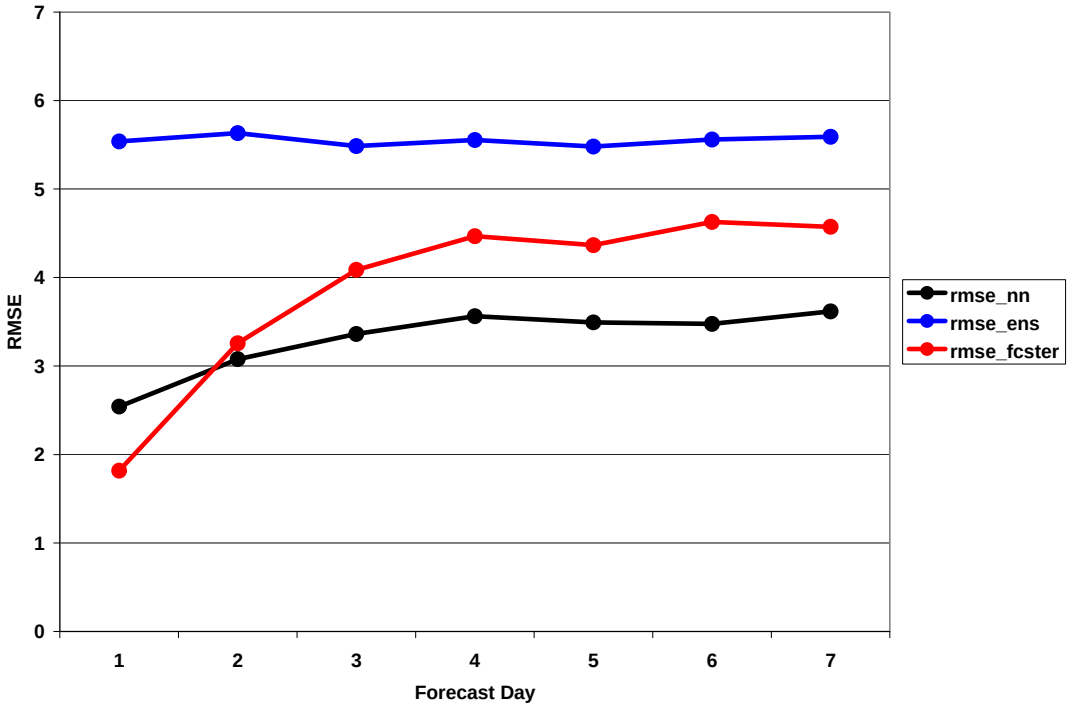

Figure 4.4 **RMSE values of the neural network forecast (NN), the NCEP ensemble forecast (ENS), and the forecaster forecast (FCSTER) for maximum temperature forecasts at the Cape Town synoptic weather station.**

Tmax forecasts have higher RMSE values than those of Tmin temperature for both the Pretoria and Cape Town station.

The correlation coefficient for Tmin temperatures at Pretoria (Figure 4.5) shows that the FCSTER forecasts were the best for forecasting day 1. For day 2 both NN forecasts and FCSTER forecasts showed the best performance. The NN forecasts are slightly better than the FCSTER forecasts for days 3 to 5. Day 6 shows again similar performance between NN forecasts and FCSTER forecasts, while day 7 has the NN forecasts as the best performer. All the different forecasts have very high correlation coefficients ranging from 0.89 to 0.90. The Tmax forecasts for Pretoria (Figure 4.6) have both the NN forecasts and the FCSTER forecasts as the best performers for day 1. For day 2 both the NN forecasts and the ENS forecasts have the highest correlation coefficients. Day 3 to 7 shows that the ENS forecasts are slightly better than the NN forecasts. The correlation coefficient of the FCSTER forecast decreases significantly by 50% towards day 7, while those of the NN forecasts and the ENS forecasts decreases only by 22% and 16 % respectively.

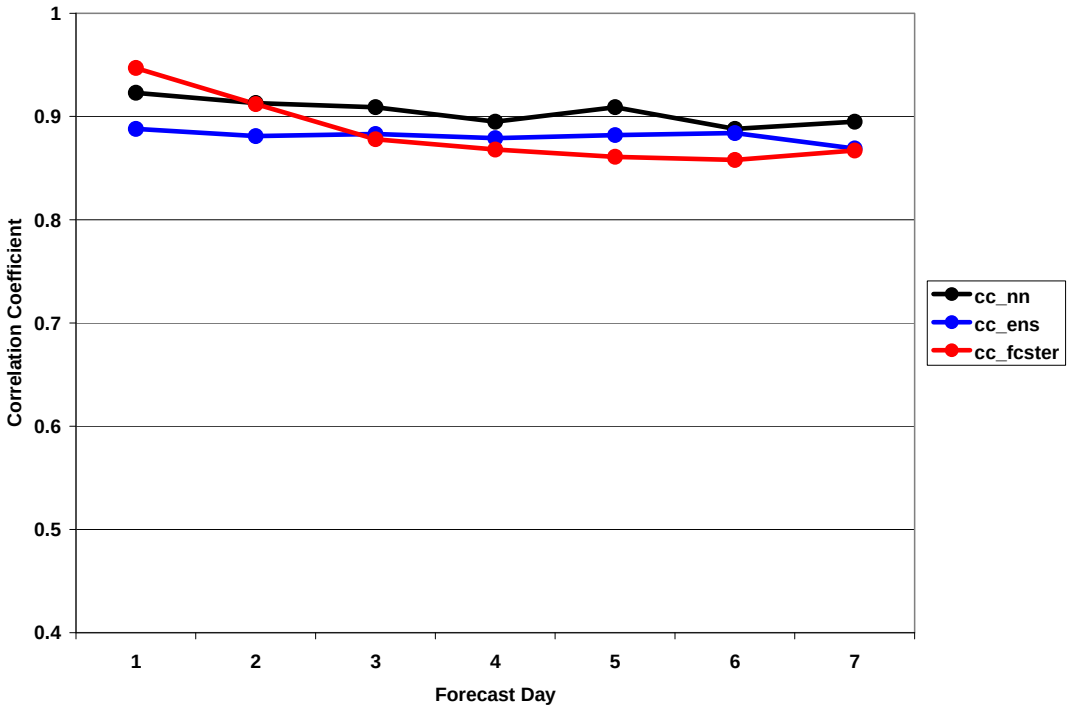

Figure 4.5 **The correlation coefficient for minimum temperatures at Pretoria.**

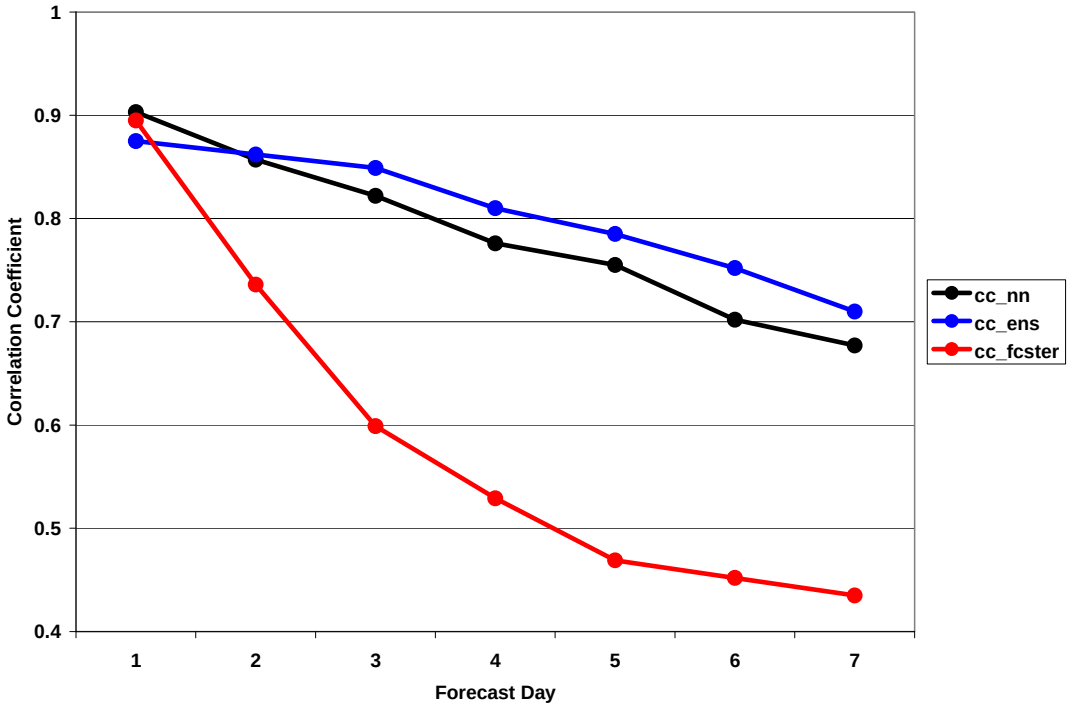

Figure 4.6 **The correlation coefficient for maximum temperatures at Pretoria**

The Tmin temperatures at Cape Town (Figure 4.7) show the highest correlation coefficient for the FCSTER forecasts for day 1 of approximately 0.97. The NN

forecasts and the FCSTER forecasts are the best performers in correlation for day 2. Day 3 has the NN forecasts as the best performer. By day 4 to day 7 all the model correlation coefficients are positioned close to each other with average correlation coefficient of 0.7. The Tmax forecasts at Cape Town (Figure 4.8) again have the FCSTER forecasts as the best performer with a correlation coefficient of 0.94. The correlation coefficient of the FCSTER however decreases rapidly towards day 7 by 41%. The NN forecasts and the ENS forecasts show a lesser declination in correlation values with a decrease of 17% and 6% respectively. For day 2 the NN forecasts and the FCSTER forecasts are very close in performance and by forecasting day 3 the NN forecasts outperform the FCSTER forecasts. For days 4 to 7 The NN forecasts are still having the highest correlation but are closely followed by the ENS forecasts.

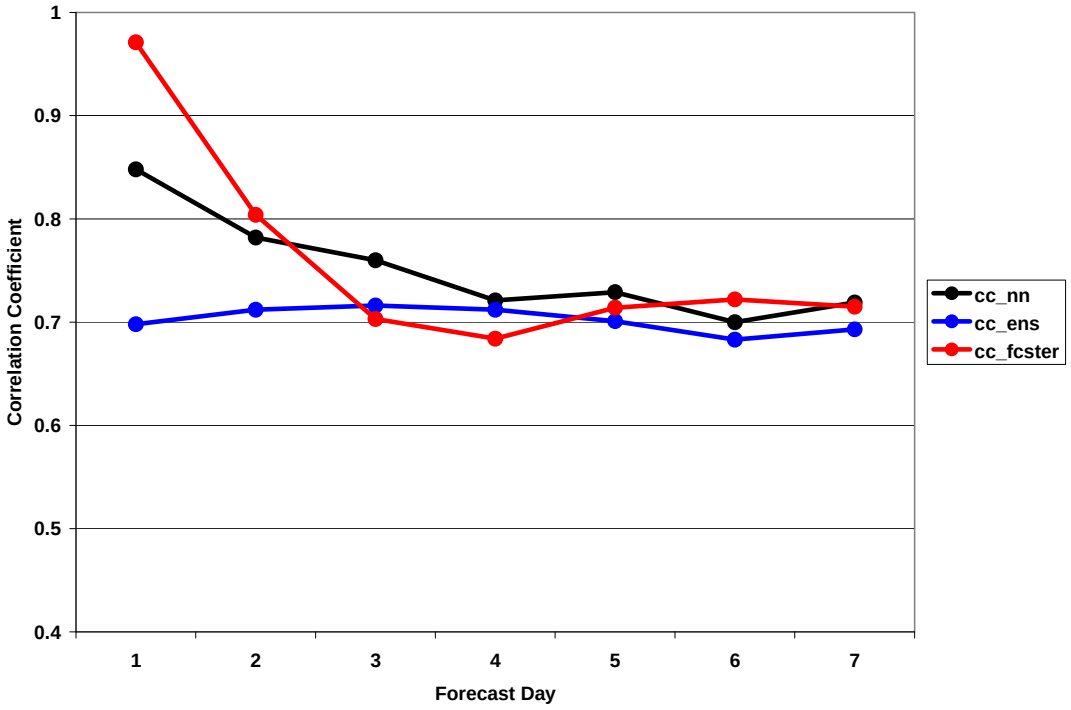

Figure 4.7 **The correlation coefficient for minimum temperatures at Cape Town.**

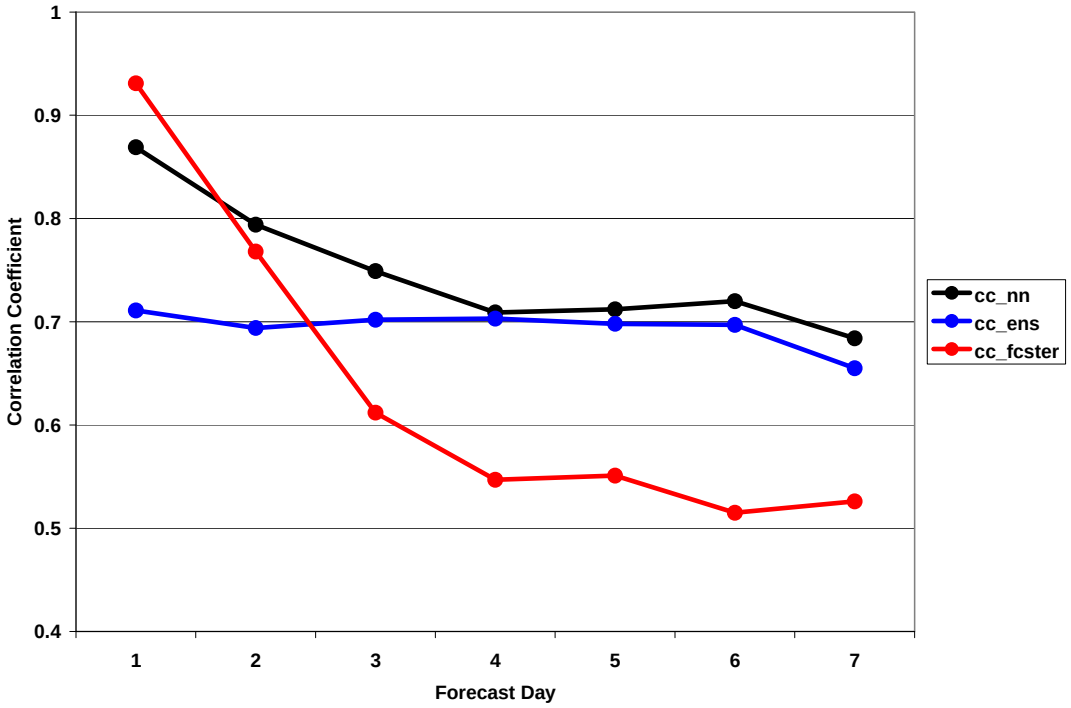

Figure 4.8 **The correlation coefficient for maximum temperatures at Cape Town.**

4.4 Eta Precipitation Forecasts

Several studies were done to verify the performance of the Eta model precipitation forecasts. Colle, Jones, and Tonque (2005) stated that the type of convective parameterization scheme used has a large affect on Eta cloud cover and precipitation forecasts. In another study done by Grams et al. (2005), it was found that for almost all types of convective precipitation, the 12-km Eta model produced lower average rain rates, peak rainfall amounts, and total rainfall than the observed values. They concluded that the explanation for the lower forecast values is the use of the BMJ convective parameterization. In terms of the capability of the Eta model to forecast precipitation intensity, Gallus and Segal (2004) found that a 10-km Eta model is more skilled in detecting regions where the occurrence of convective precipitation are most likely than in predicting the precipitation intensity.

For this research Eta precipitation forecasts were evaluated at 8 synoptic stations as was mentioned in chapter 3 for the winter and summer months of the year 2005. The Eta model forecasts consist of a matrix or field of values for each model variable. Model grid points have one forecast value assigned to it for each of the variable fields.

The model grid point value does not present the forecast condition at that specific geographical point but rather the average condition in the area surrounding the grid point. Still model forecasts are used to determine conditions at specific station locations as the general public are interested in weather conditions at their current destination. For example a precipitation forecast rendered by the Eta model for the city of Pretoria is in actual fact a forecast for the 32x32 km^2 area surrounding a model grid point which is located closest to Pretoria. This model attribute is problematic in the case of real weather parameters especially precipitation which is highly variable in space. Model precipitation forecasts at station points comprise of a large degree of uncertainty. Bernardet (2002) found bias scores to be higher when computed at the model grid points instead of at the station locations.

Figure 4.9 and 4.10 shows the accuracy rate for summer and winter months respectively. Different colored lines depict different synoptic stations. The Eta model forecast accuracy seems to increase from 0.1mm and onwards until it reaches a peak round about the 5mm threshold for both summer and winter months. The accuracy rate becomes constant for thresholds at 50mm and higher during summer months and at 20mm and higher during winter months. This trend emerges when there are hardly any or none forecasts appearing at these threshold ranges. Accuracy rates are higher for winter months than for summer months, because the frequency of 0mm forecasts is much higher during winter such that fewer penalties are placed upon model performance if precipitation intensity is forecast incorrectly. All 8 synoptic stations are located in a summer rainfall region. In general the accuracy rates of precipitation forecasts are moderately good except for the 0.1mm threshold during summer months which are slightly higher than 0.5 for some of the stations meaning that the forecasts were almost as many times erroneous as it was accurate.

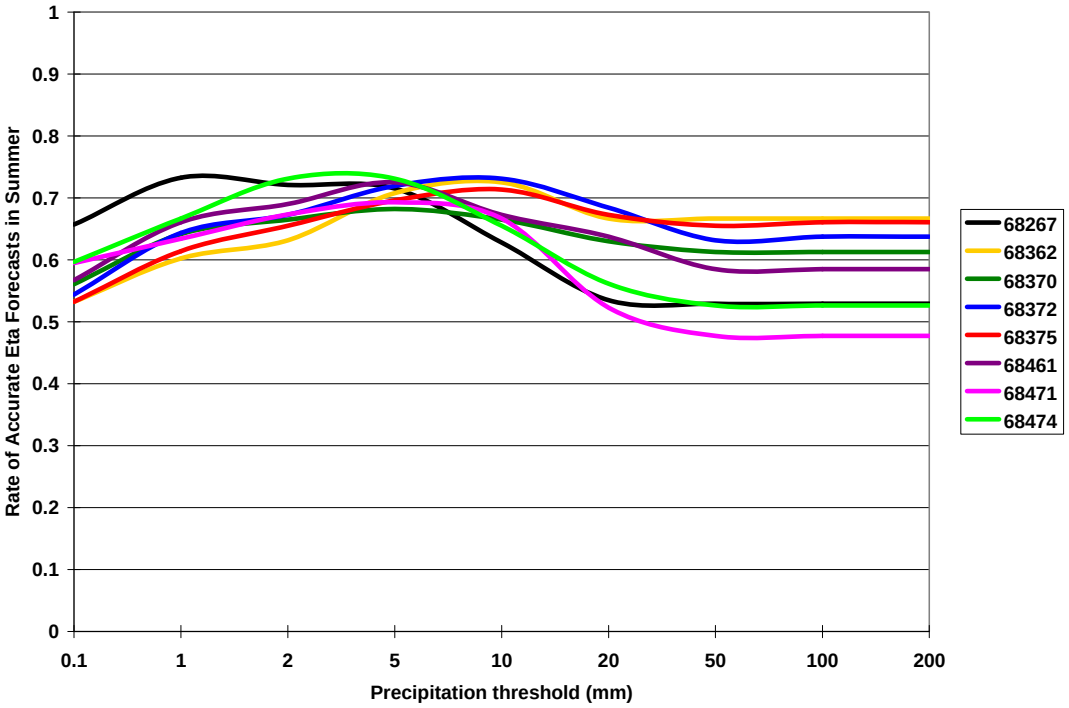

Figure 4.9 **The accuracy rate of Eta precipitation forecasts for the summer months.**

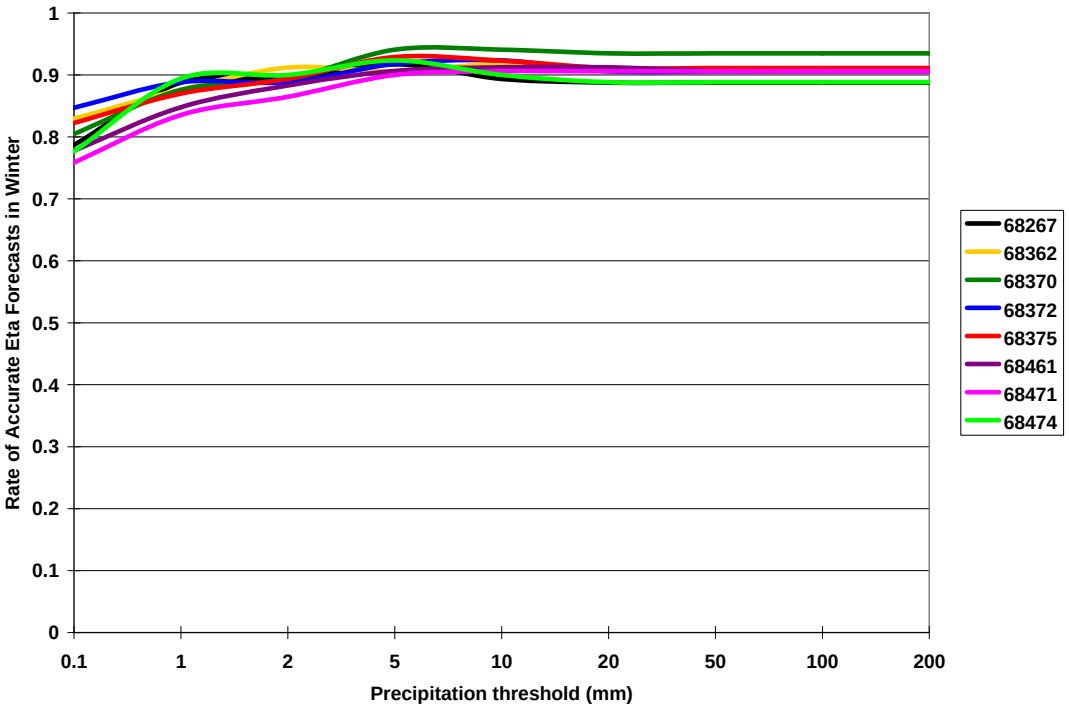

Figure 4.10 **The accuracy rate of Eta precipitation forecasts for the winter months.**

The POD decreases more rapidly with precipitation threshold during winter (Figure 4.12) than summer (Figure 4.11) which peaks between the 0.1mm and 2mm thresholds before descending. The descending POD lines illustrate that the model is

more capable in detecting the lower range of thresholds. Higher range precipitation thresholds are mostly associated with convective precipitation systems which are complicated to model with a NWP model. Convective systems are modeled by a convective parameterization scheme for the reason that precipitation is classified as a subgridscale phenomenon in relation to the horizontal resolution of the Eta model. The nearly absence of convective systems during winter months explains the rapid declining of the POD lines.

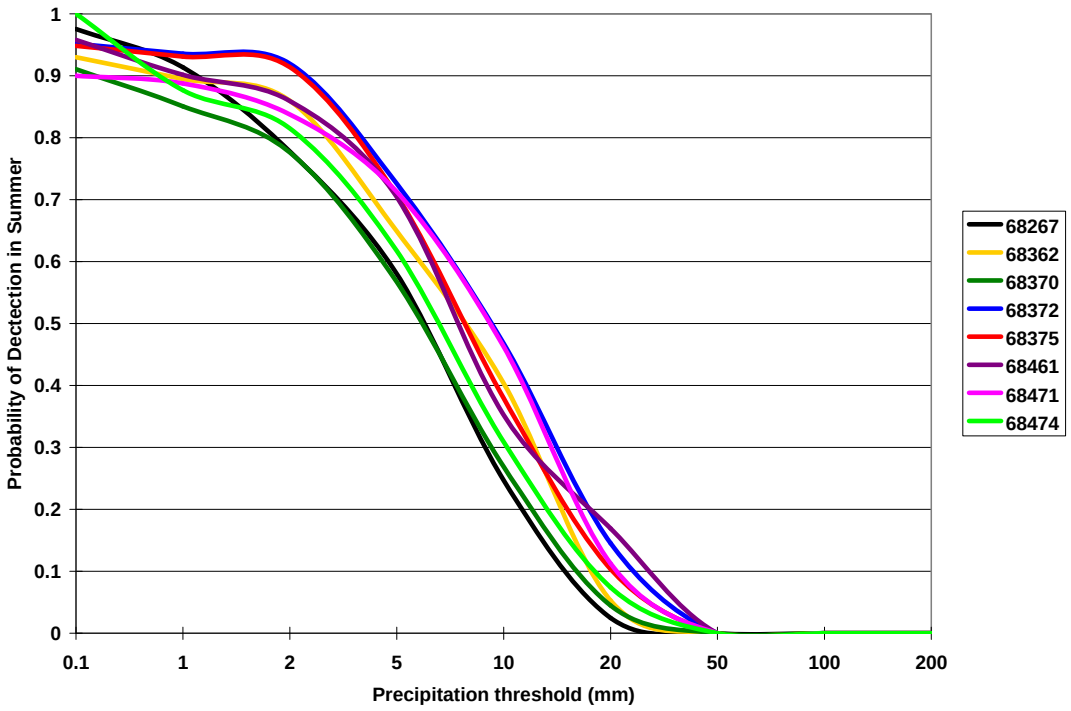

Figure 4.11 The probability of detection of Eta precipitation forecasts for the summer months.

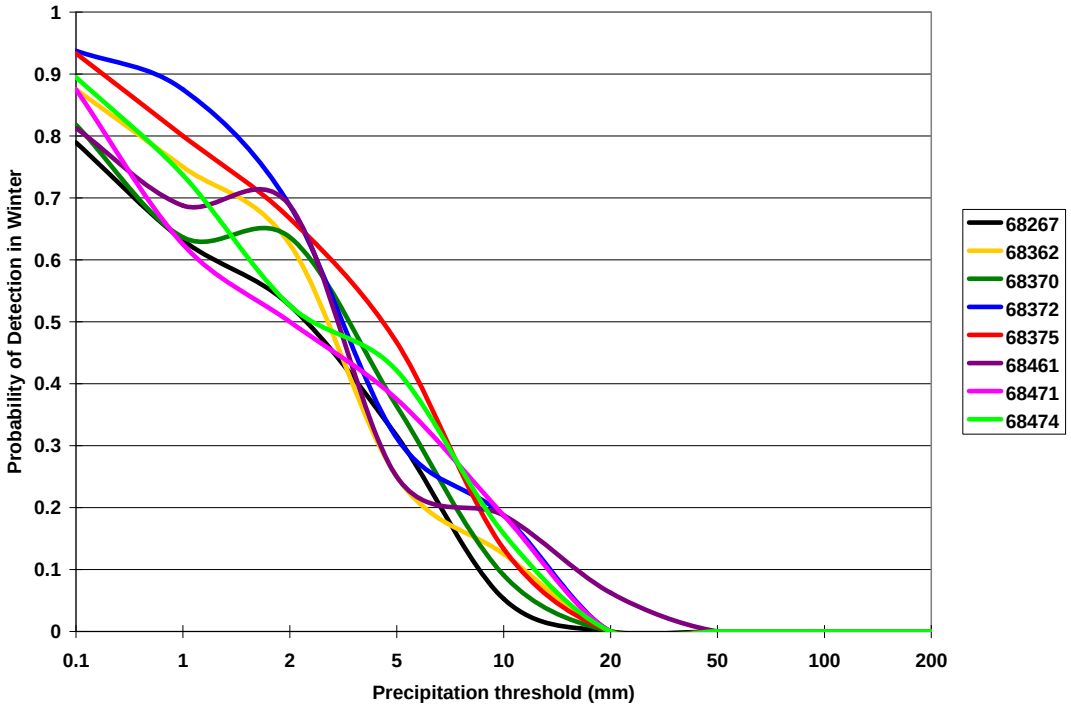

Figure 4.12 **The probability of detection of Eta precipitation forecasts for the winter months.**

Figures 4.13 and 4.14 illustrate the bias rate of the summer and winter months respectively. Bias rate values larger than 1 depicts over-forecasting, while values equal to 1 are associated with perfect forecasts and values smaller than 1 depicts under-forecasting. During the summer and winter months the precipitation intensities of 0.1mm towards 5mm were over-forecast. Perfect forecasts are found at more or less 5mm in summer and between 2mm and 5mm during winter months. Thresholds above 5mm were under-forecast for both seasons. The tendency of the model to forecast precipitation intensities mostly in the lower range, 0.1 to 5mm, shows a lack of ability in the adequate presentation of convective systems and an inappropriate convective parameterization scheme.

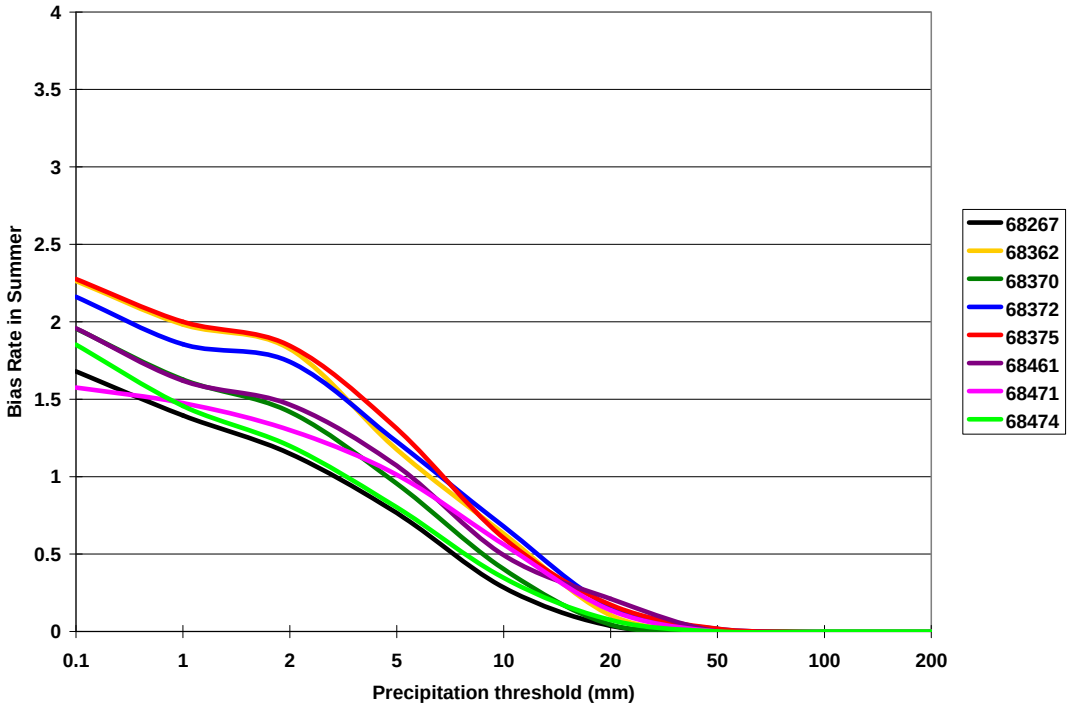

Figure 4.13 **The bias rate of Eta precipitation forecasts for the summer months.**

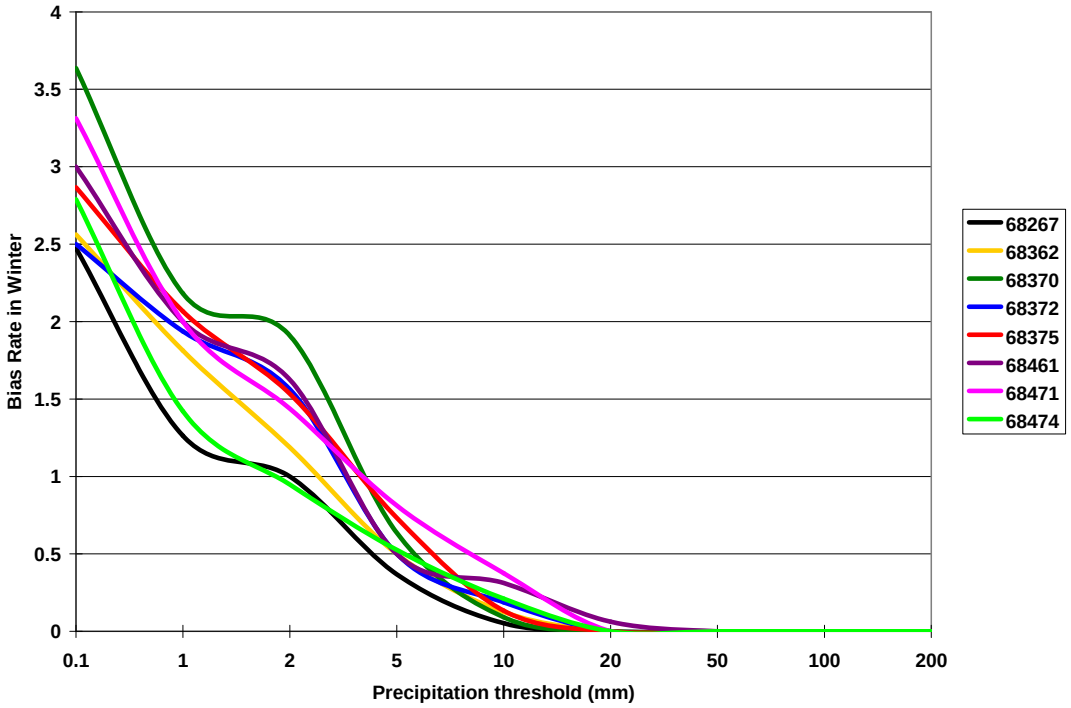

Figure 4.14 **The bias rate of Eta precipitation forecasts for the winter months.**

The equitable threat score determines the correspondence between forecast rain-events and observed rain-events. This score is equitable since it penalizes missed events and false alarms in a similar manner and is therefore suitable in the verification of precipitation forecasts which consist of wet and dry regimes. The equitable threat score for the summer months reached a peak between 1mm and 10mm (Figure 4.15) and between 1mm and 5mm during the winter months (Figure 4.16). The peaks reached ranged from 0.2 to 0.35 which are relatively acceptably in terms of NWP model performance. Rogers et al. (1996) confirmed that the ETS of NWP models rarely exceeds the value of 0.4.

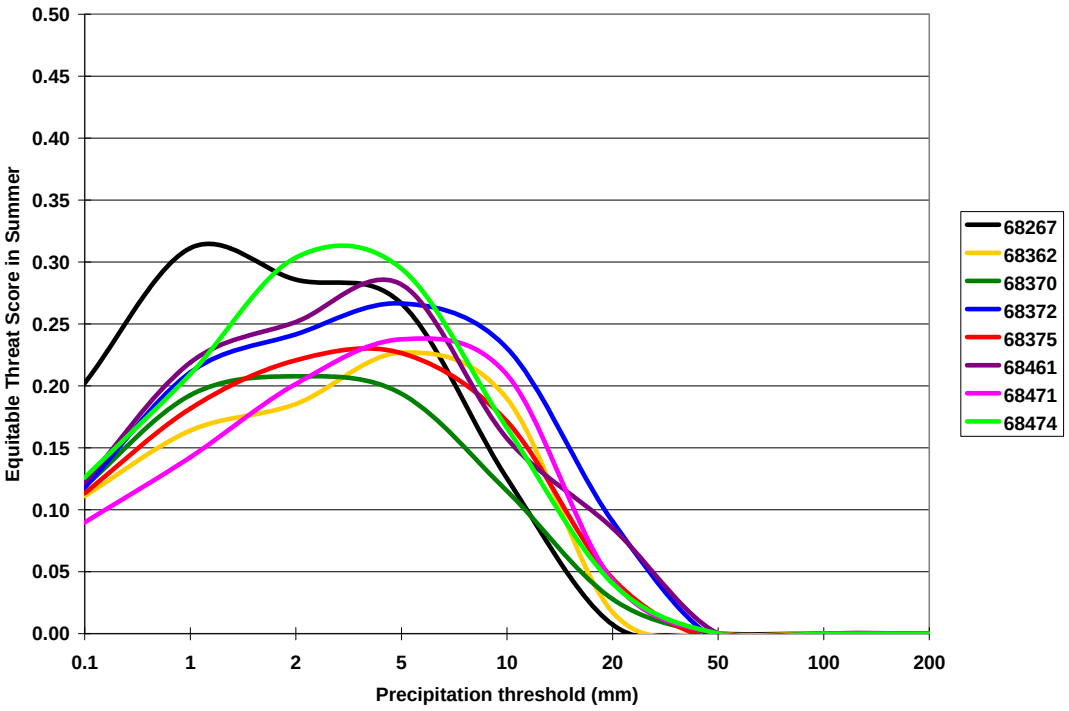

Figure 4.15 **The equitable threat score of Eta precipitation forecasts for the summer months.**

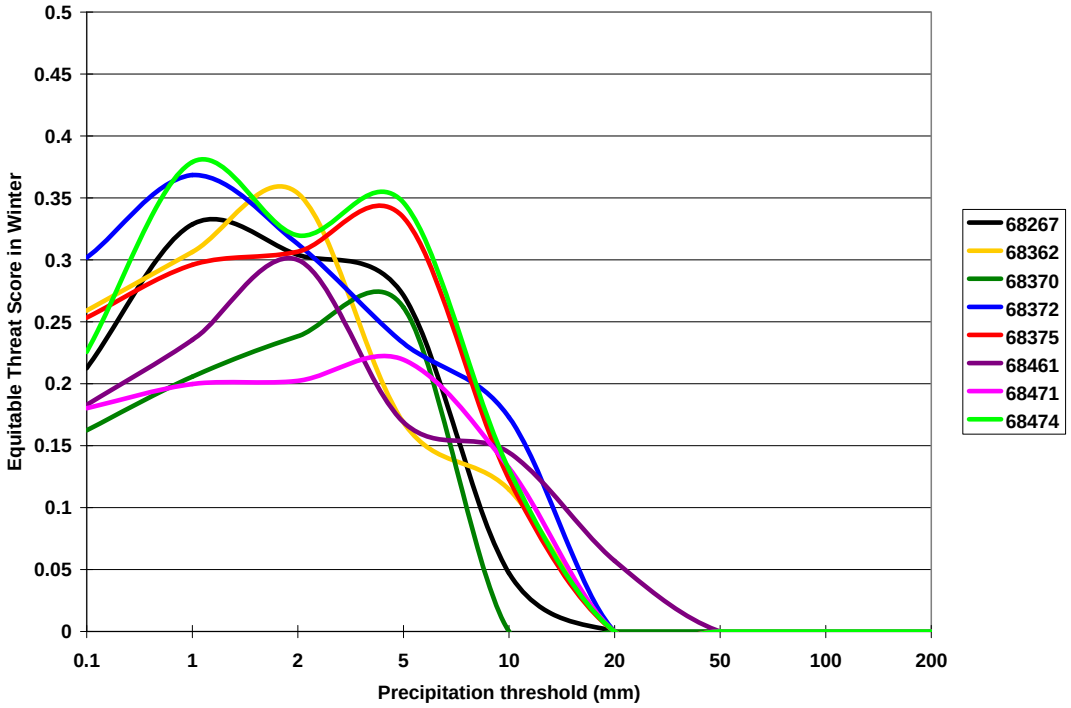

Figure 4.16 **The equitable threat score of Eta precipitation forecasts for the winter months.**

4.5 Probability of Precipitation Forecasts

A NN attempts to learn the association between the coarse area-averaged model parameters and the actual precipitation at a specific point. This is rather an immense task to perform when considering the complexity of precipitation. Therefore the availability of a vast amount of cases consisting of model forecasts and its associated precipitation observations has considerable influence on NN forecast performance. The question of how large such a data archive should be or what the limit of the capability of the NN is to get satisfactorily forecast performance can only be answered over a period of time as the volume of this archive grows.

The PoP forecasts consist of a continuous range of values between 0.0 and 1.0, which describes the probability of a no-precipitation and precipitation event, respectively. The probability forecasts of precipitation by the forecasters are similar in meaning but are expressed slightly different as the product between the forecaster confidence in the occurrence of an event and the areal coverage of the event (Poolman, 2004).

Verification on PoP forecasts was done for the period January 2005 to December 2005. Figure 4.17 and 4.18 illustrates the Bss, the Reliability skill score (Brel) and the Resolution skill score (Bres) for each of the 8 selected synoptic stations, during the summer months and winter months, respectively. These scores all range from a value of -∞ to a perfect score value of 1. All the scores have positive values which indicate that the NN performed better than the sample climatology. The Ermelo station has the highest Bss of 0.3 and 0.36, for summer and winter months respectively. The rest of the stations had Bss ranging between 0.15 and 0.2 during summer months and between 0.08 and 0.27 for winter months. The worst performing forecasts were made at the Frankfort station during summer and at the Van Reenen station during winter. Table 4.1 shows the availability of verification cases at each station as well as the amount of precipitation events that was observed during the verification period. The varying performances at the different stations seem to be linked with the amount of precipitation events in summer data sample. The winter forecast performances show the same relationship for the Bethal station which only had 11 precipitation events, but the low scores at the Bethlehem and Van Reenen might be due to complex topography. All the stations have approximately 10 missing cases during the summer months of 2005, except at the Van Reenen station which had 29 missing cases. For winter months each station had approximately 13 missing cases.

Synop Station	**Summer: Cases (Precip. events)**	**Winter: Cases (Precip. events)**
Ermelo	172 (81)	169 (19)
Frankfort	171 (57)	170 (15)
Bethal	173 (67)	169 (11)
Standerton	171 (62)	170 (16)
Vrede	171 (58)	169 (15)
Bethlehem	171 (71)	171 (16)
Van Reenen	153 (80)	170 (16)
Royal Natal National Park	171 (81)	170 (19)

Table 4.1 **The availability of verification cases at each station and the amount of precipitation events that was observed during the verification period in brackets.**

Figure 4.19 shows the Bss of the precipitation forecasts rendered by the forecasters over all South African stations. In the area of the Vaal dam catchment the Bss's have values ranging from 0 to 0.2, which is only slightly lower than the scores of the NN PoP forecasts.

The Bss can be decomposed into the different contributions by reliability, resolution and uncertainty. The Brel are very high for all stations. During summer months the Brel are higher than 0.9 and during winter months the Brel range from 0.81 to 0.92. The Bres are much lesser than the Brel, therefore the Brel are major contributor to the value of the Bss. The Bres range from 0.16 to 0.34 during summer months and from 0.26 to 0.47 during the winter months. The winter precipitation forecasts have more resolution than the summer precipitation forecasts but to the expense of reliability.

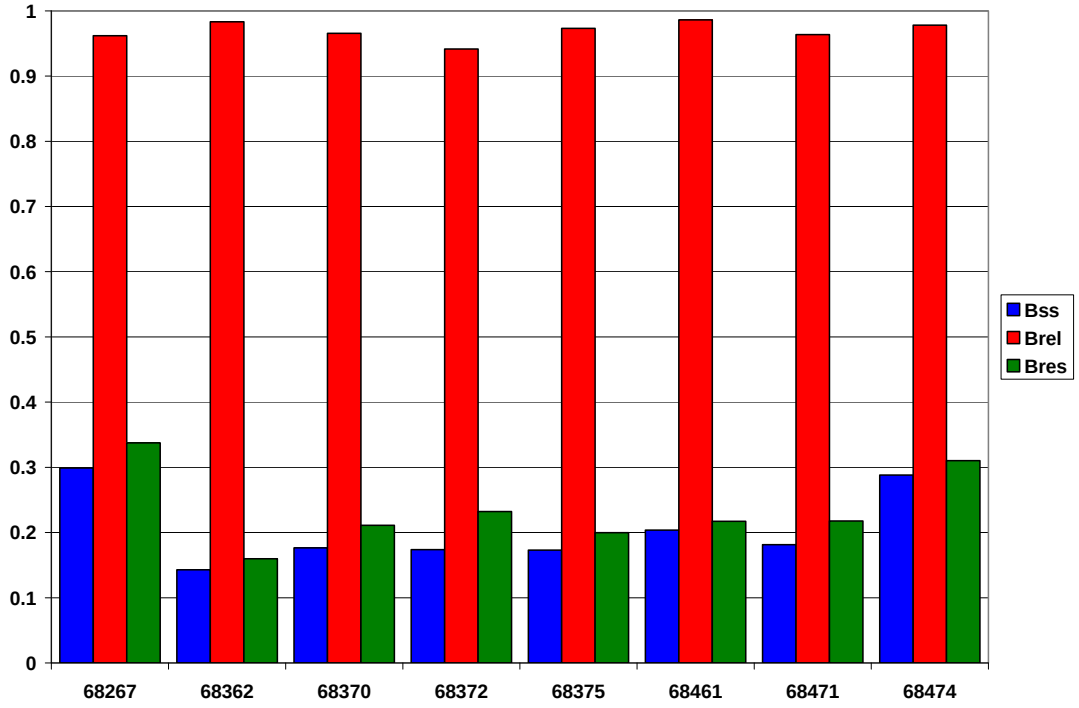

Figure 4.17 **The Brier skill score (Bss), the Reliability skill score (Brel) and the Resolution skill score (Bres) for each of the 8 selected synoptic stations, during the summer months.**

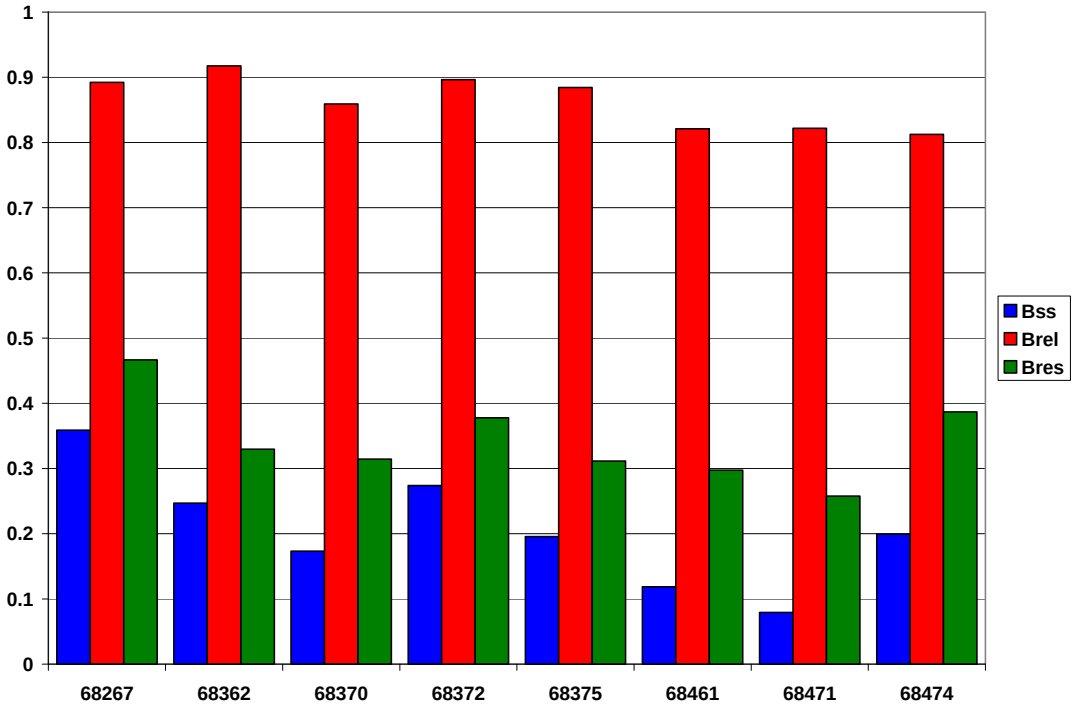

Figure 4.18 **The Brier skill score (Bss), the Reliability skill score (Brel) and the Resolution skill score (Bres) for each of the 8 selected synoptic stations, during the winter months.**

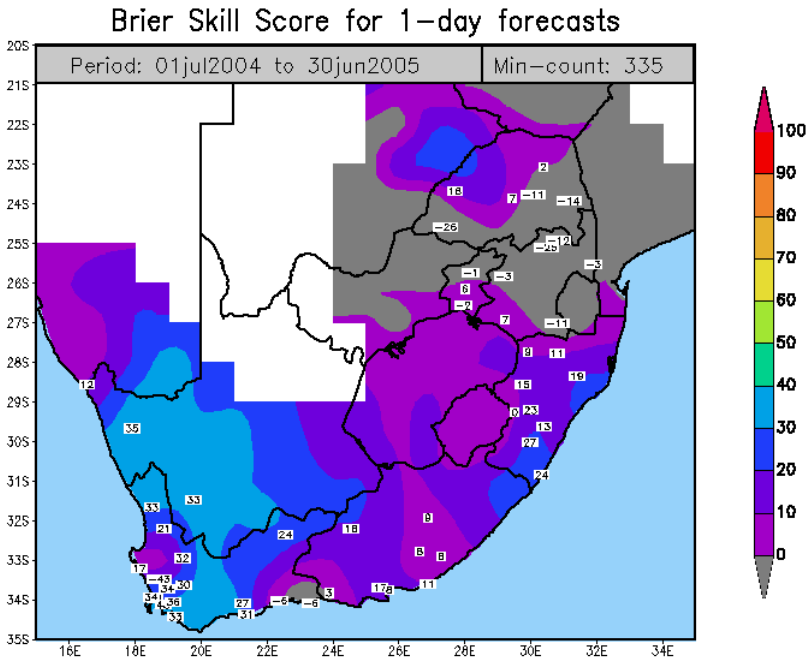

Figure 4.19 **The Brier skill score of the precipitation forecasts by the forecasters over all South African stations.**

The ROC curve was calculated at each station using the POD and POFD values. The area under the plotted ROC curve indicates the skill of the forecasts. If the ROC area is larger than 0.5 the forecasts are better than random forecasts and a value of 1 indicates a perfect forecast. In Figure 4.20 the area under each station ROC curve are illustrated for both the summer and winter months. The ROC areas for the summer months range from 0.74 to 0.84 and from 0.75 to 0.89 for winter months. Six of the stations have higher Bss in the winter months. Only the Ermelo and Bethlehem stations have slightly higher Bss during summer months. All ROC areas are above 0.5 and are relatively high.

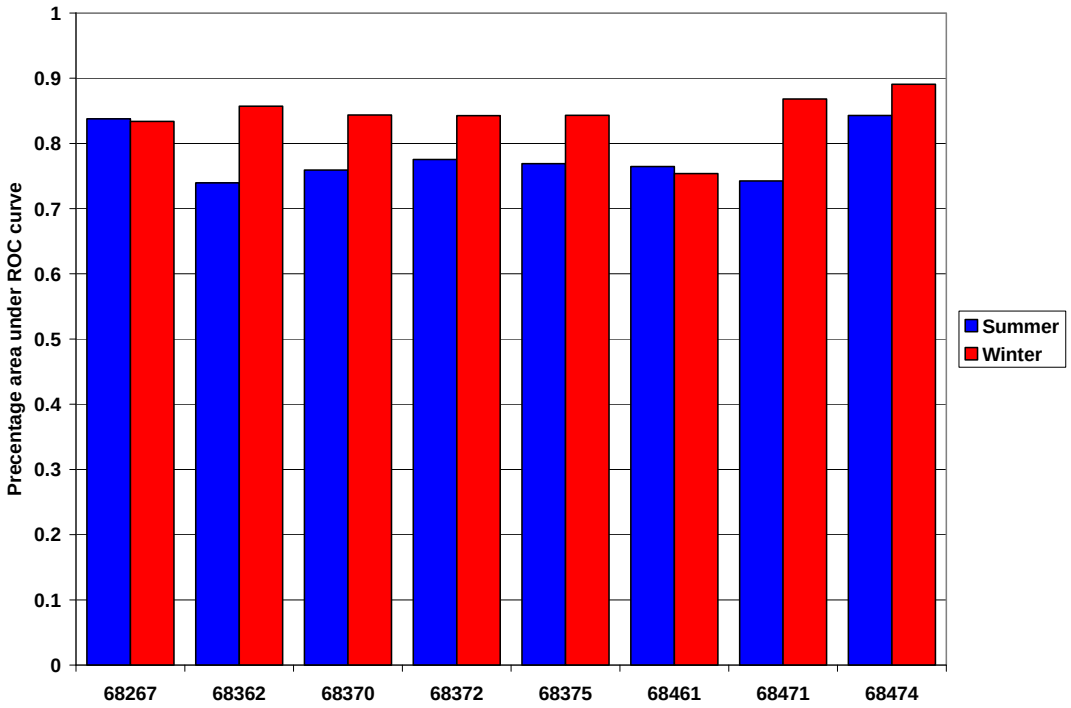

Figure 4.20 **The area under the ROC curve for each of the stations for both the summer and winter months.**

A frequency distribution plot shows the spread of the data over different forecast categories. Figure 4.21 to 4.32 shows the frequency distribution plots of 3 synoptic stations at which precipitation probability forecasts are made for 1 day to 7 days ahead. These stations are the Ermelo (Figure 4.21 to 4.24), the Standerton (Figure 4.25 to 4.28), and the Bethlehem station (Figure 4.29 to 4.32). The blue histogram bars shows the frequency distribution of forecaster forecasts in each forecast category of 0%, 30%, 60%, and 80%. The Eta precipitation forecasts were stratified into Yes/No event forecasts using the threshold of 0.1mm and the frequency distributions

of the Yes/No event forecasts are indicated by the red histogram bars. The green histogram bars illustrates the distribution of the NN PoP forecasts into a 0-10%, 20-40%, 50-70%, and an 80-100% category.

During summer months, for days when no precipitation was observed (Figure 4.21, 4.25, and 4.29) the NN forecasts showed no improvement over the Eta model forecasts regarding the amount of hits achieved by each. The performance of the Eta forecasts is quite similar to the performance of the forecaster's forecasts. It was noticed that approximately 40% of the forecaster's forecasts were made in the 30% chance of precipitation category. It was also found that a third to a half of the NN forecasts appeared in the 20-40% category. For summer days when precipitation was actually observed (Figure 4.22, 4.26, and 4.30) the NN forecasts proved to be a significant improvement over the Eta precipitation forecast as well as the forecaster's forecasts. The proportion of hits ranges from 94% to 98% of the total amount of forecasts. It was again noticed that round about 40% of the forecaster's forecasts were made in the 30% chance category. Due to poor performance of the NN forecasts during no-precipitation days and its significant performance during precipitation days, it was decided to change the stratification of the NN forecasts. The hits accomplished by NN forecasts for no-precipitation days increased from a range of 35% to 48% to a range of 54% to 66% after stratifying the no-precipitation category to 0-20% from the original 0-10% category. This change provides a NN model which performs better than both the forecaster and the Eta model. There was a slight decrease in the NN forecast performance in hits of about 10% after applying a similar stratification to the forecasts during which summer precipitation was observed, but were still able to perform slightly better than the forecaster except at the Standerton station.

The NN forecasts during winter months had hits for approximately 88% of the forecasts for no-precipitation days at the three synoptic stations (Figure 4.23, 4.27, and 4.31). These forecasts also performed better than the Eta model forecasts, but the forecaster's forecasts still performed the best with hits for approximately 91% of the days. By changing the no-precipitation category for NN forecasts to 0-20% rendered their performance to be similar to those of the forecaster. For winter precipitation days (Figure 4.24, 4.28, and 4.32) there were only 19 observations at Ermelo, 15 observations at Standerton, and 14 observations at Bethlehem. The Eta model and the

NN model seems to show better performance than forecasters for these kinds of days, but this can't be said with certainty due to small amount of cases available for winter precipitation at stations which are mainly located in a summer rainfall regime.

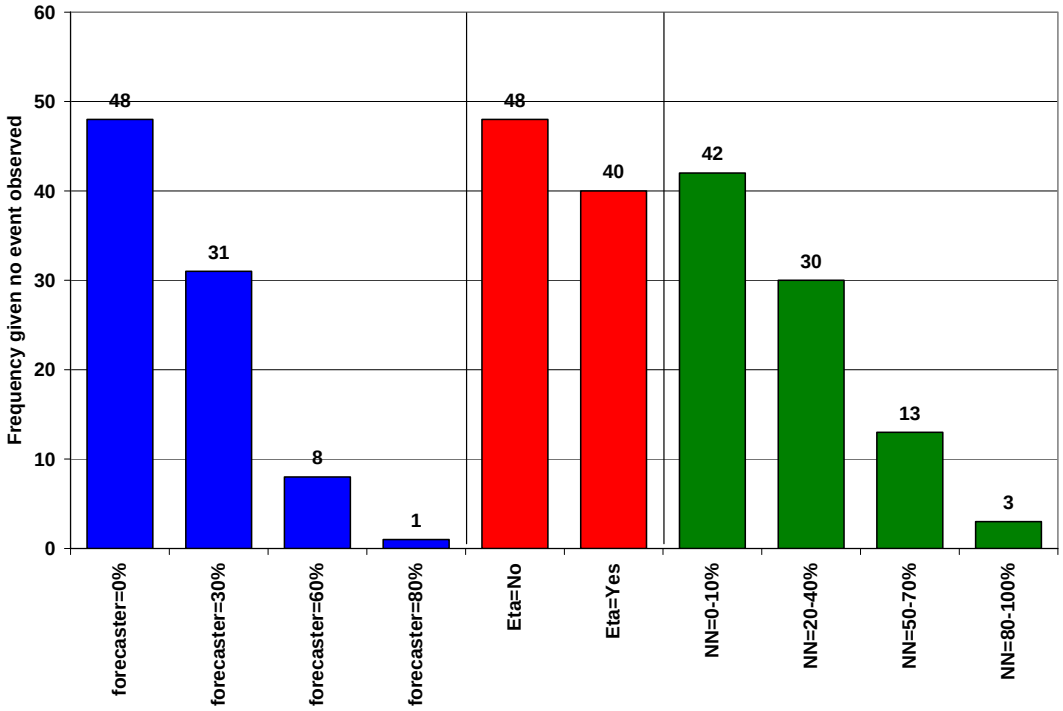

Figure 4.21 The frequency distribution plot of PoP forecasts for days during the summer months when no precipitation was observed at the Ermelo station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.

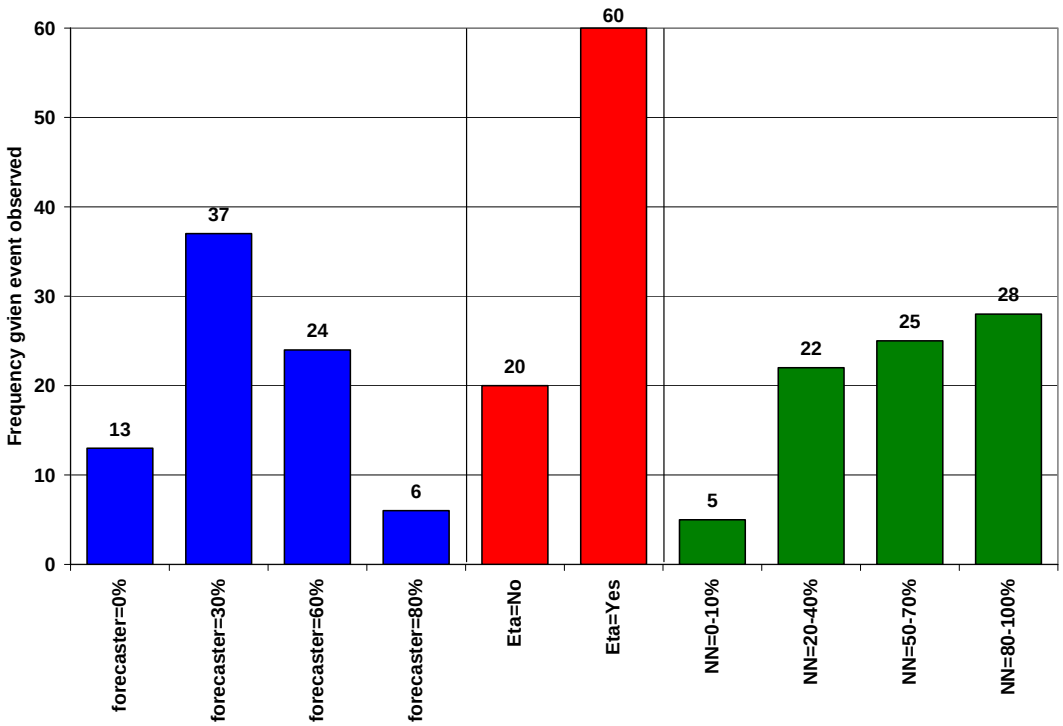

Figure 4.22 **The frequency distribution plot of PoP forecasts for days during the summer months when precipitation was observed at the Ermelo station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

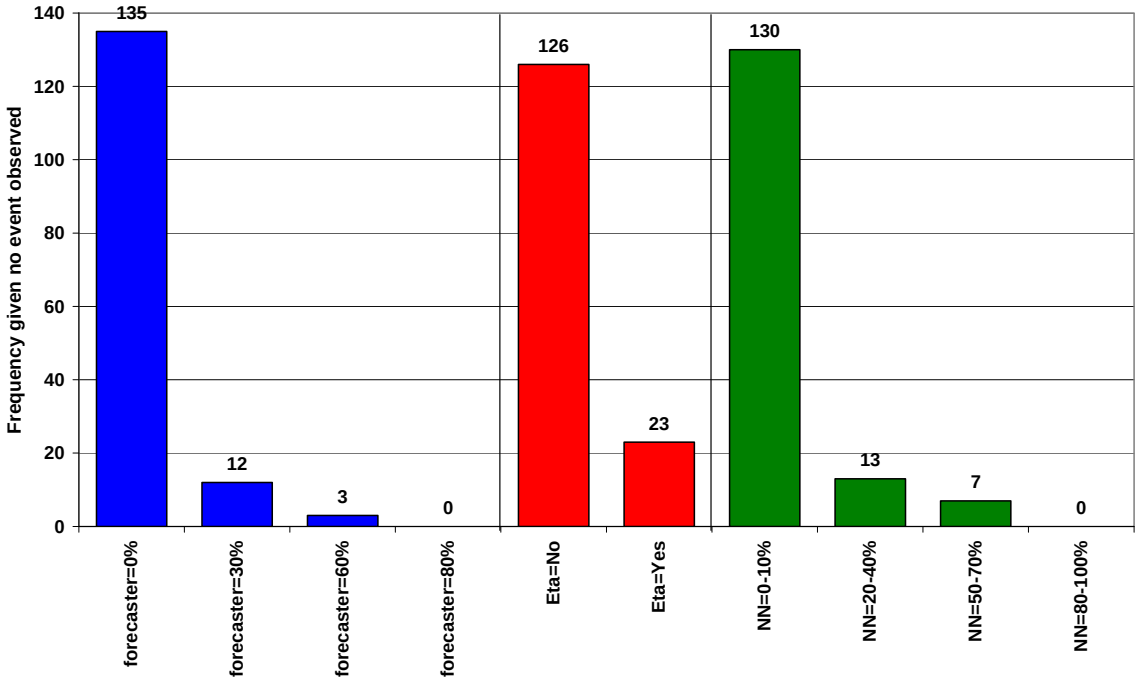

Figure 4.23 **The frequency distribution plot of PoP forecasts for days during the winter months when no precipitation was observed at the Ermelo station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

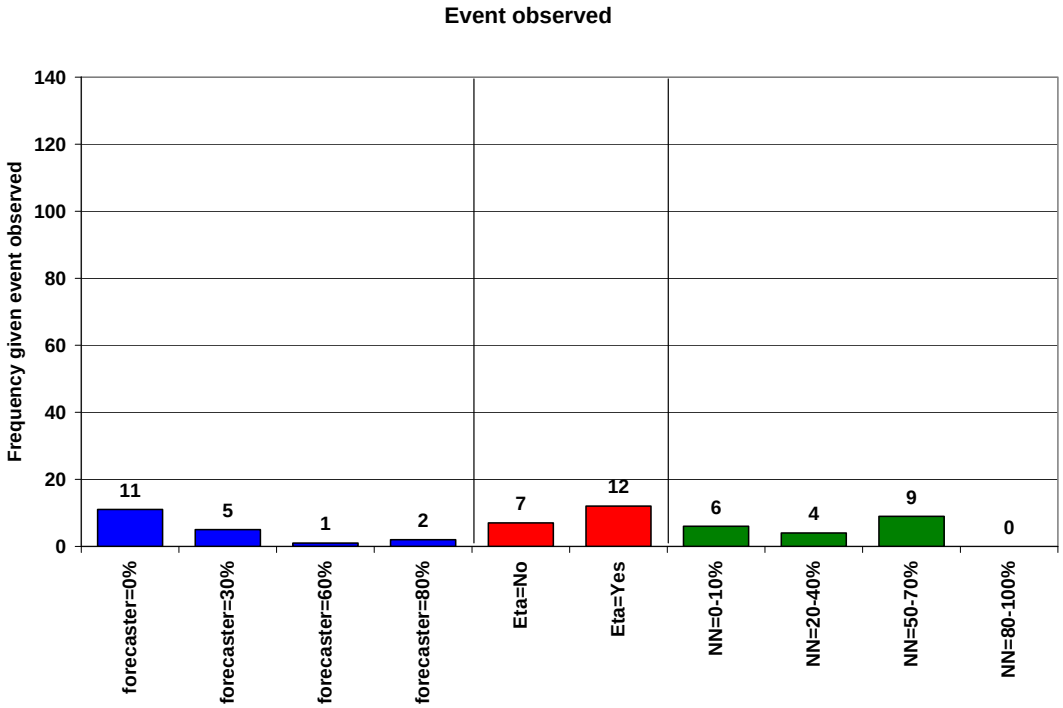

Figure 4.24 **The frequency distribution plot of PoP forecasts for days during the winter months when precipitation was observed at the Ermelo station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

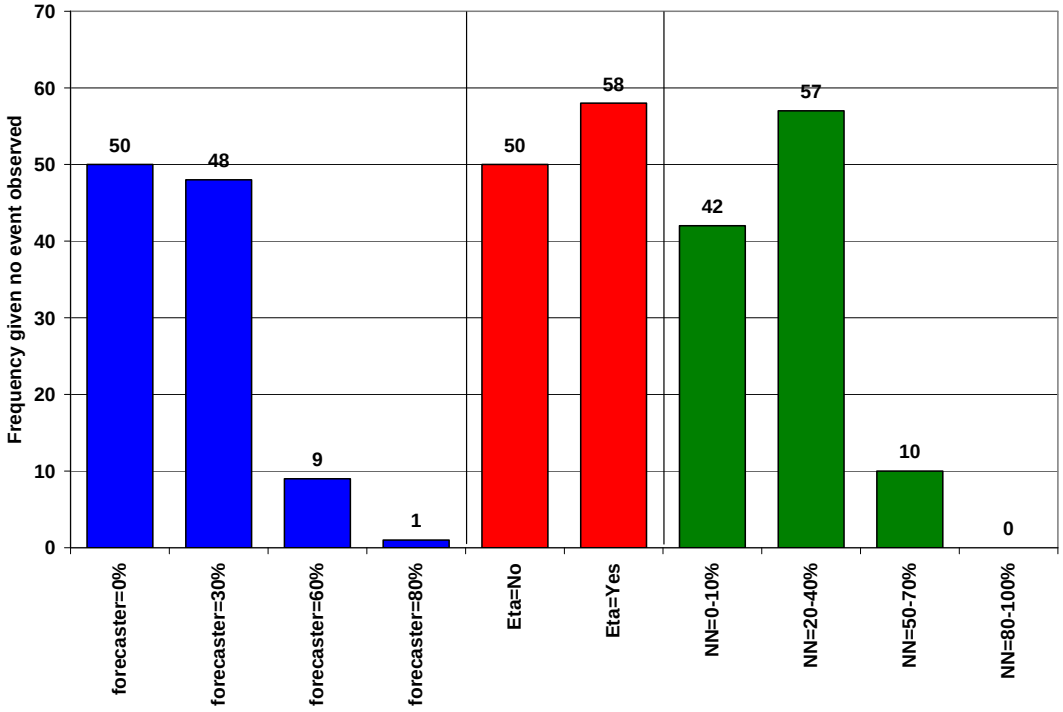

Figure 4.25 **The frequency distribution plot of PoP forecasts for days during the summer months when no precipitation was observed at the Standerton station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

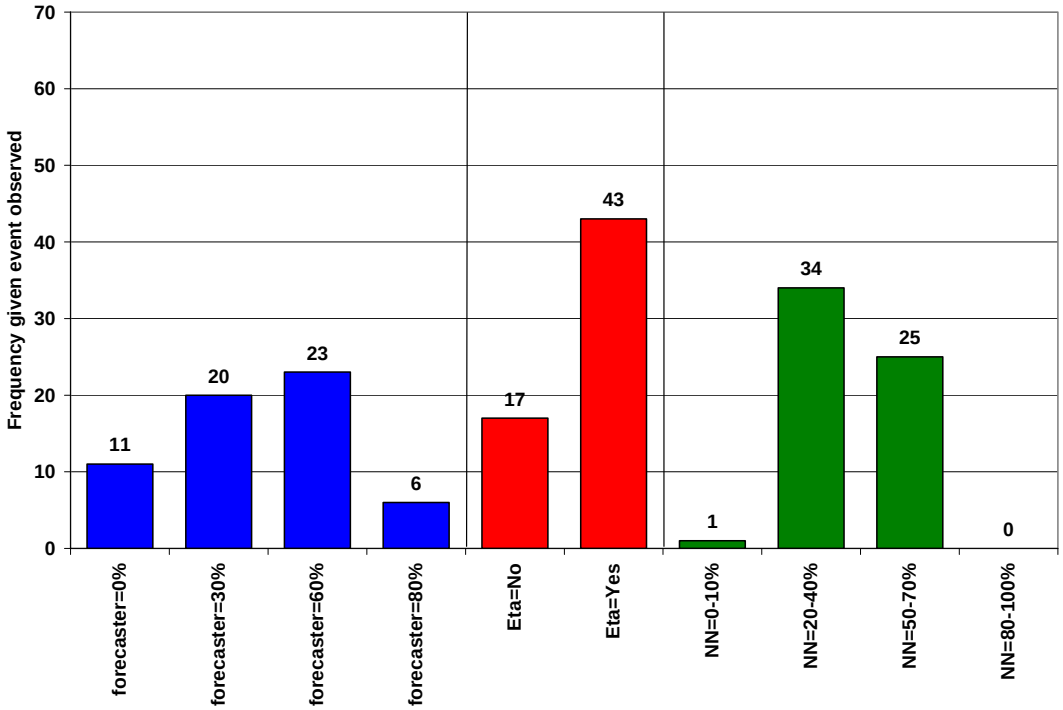

Figure 4.26 **The frequency distribution plot of PoP forecasts for days during the summer months when precipitation was observed at the Standerton station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

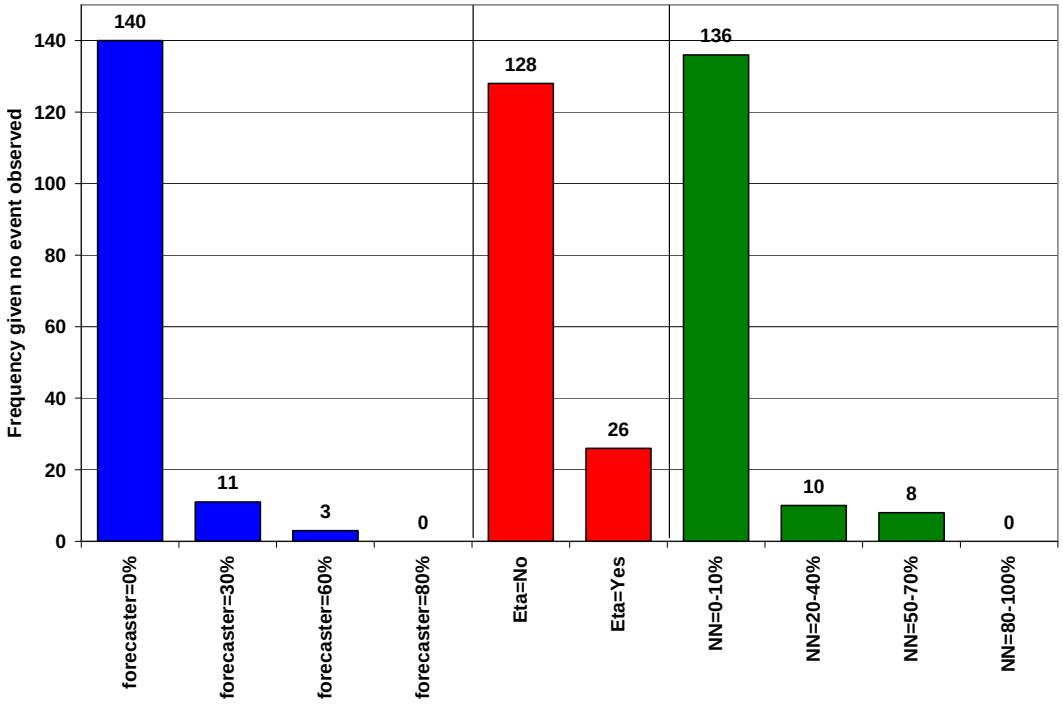

Figure 4.27 **The frequency distribution plot of PoP forecasts for days during the winter months when no precipitation was observed at the Standerton station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

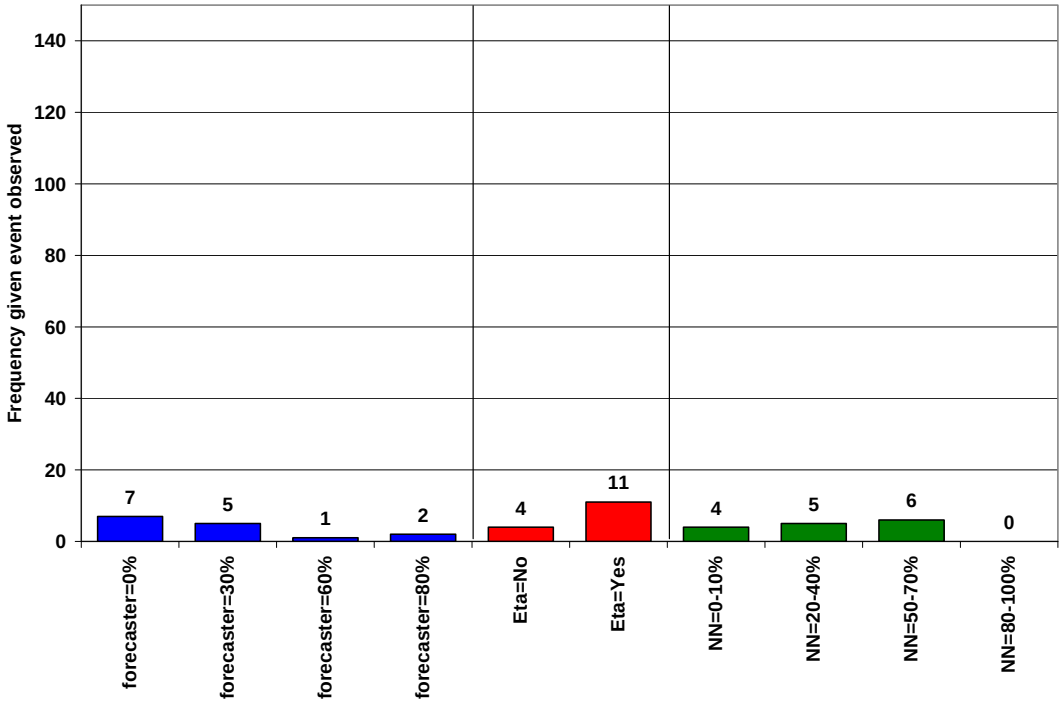

Figure 4.28 **The frequency distribution plot of PoP forecasts for days during the winter months when precipitation was observed at the Standerton station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

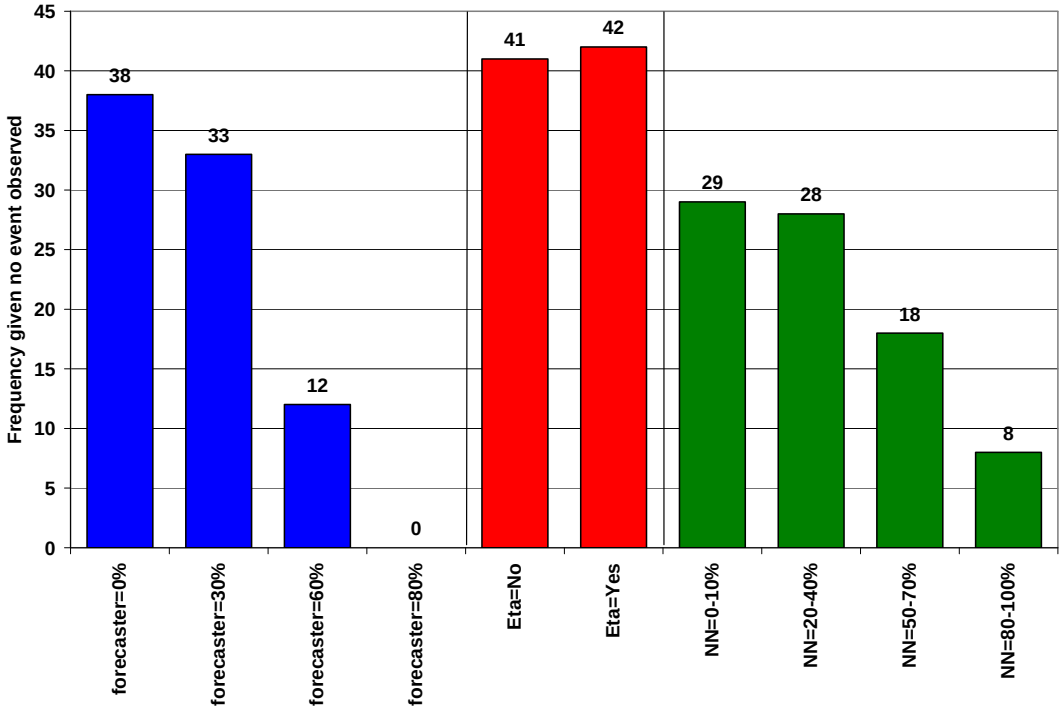

Figure 4.29 **The frequency distribution plot of PoP forecasts for days during the summer months when no precipitation was observed at the Bethlehem station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.**

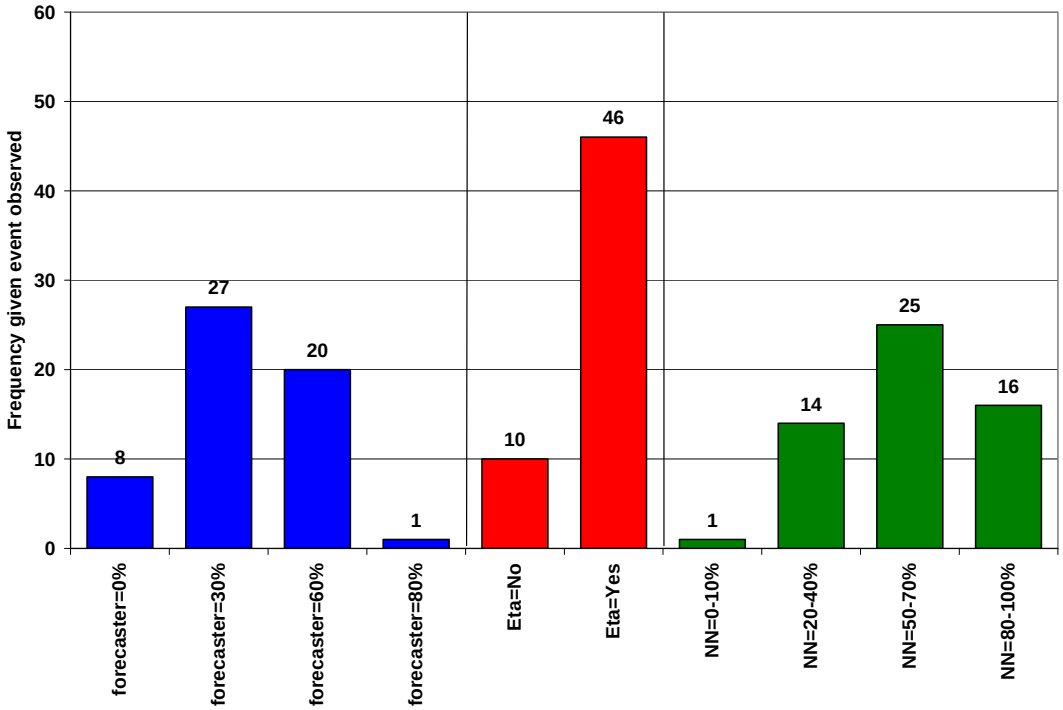

Figure 4.30 The frequency distribution plot of PoP forecasts for days during the summer months when precipitation was observed at the Bethlehem station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.

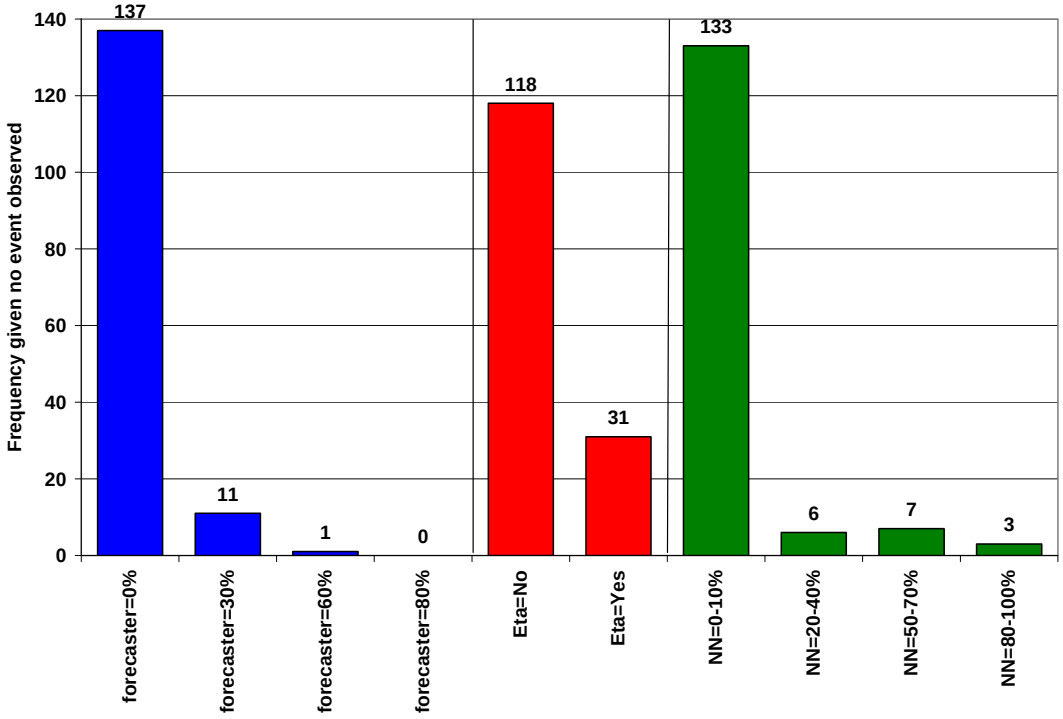

Figure 4.31 The frequency distribution plot of PoP forecasts for days during the winter months when no precipitation was observed at the Bethlehem station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.

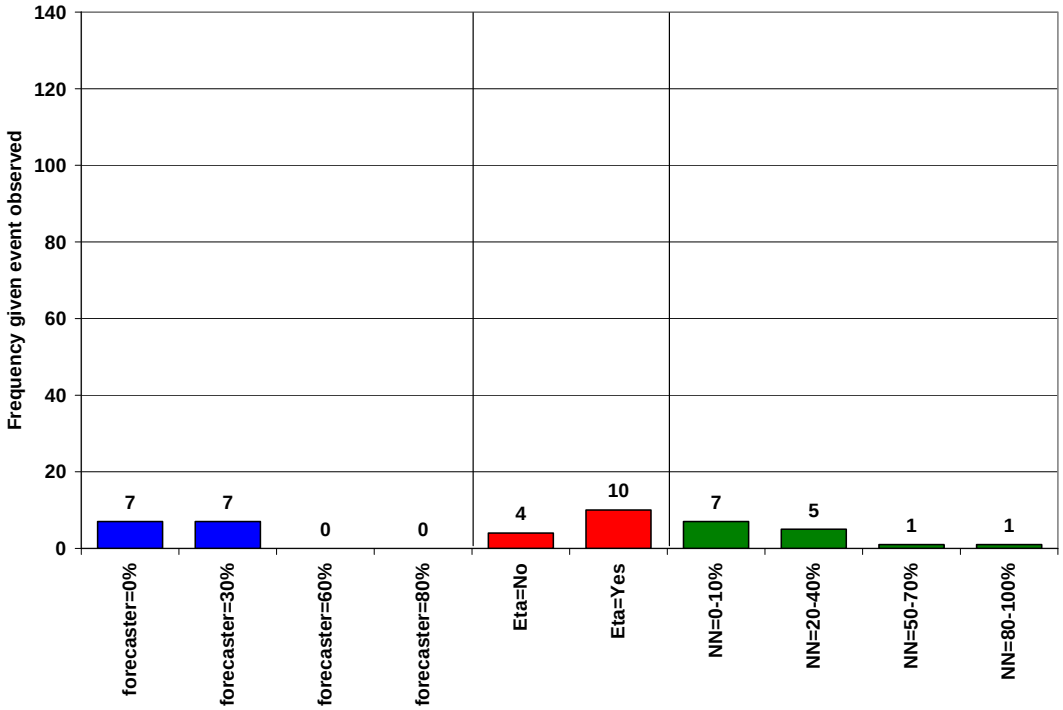

Figure 4.32 The frequency distribution plot of PoP forecasts for days during the winter months when precipitation was observed at the Bethlehem station. Blue bars indicate PoP forecasts by the forecaster, red bars by the Eta model, and green bars by the neural network.

4.6 Precipitation Intensity Forecasts

The NN forecasts and the Eta model forecasts were verified against the synoptic station observations for the summer and winter months of the year 2005. RMSE values for the summer months (Figure 4.33) show that the NN forecasts have less error than the Eta precipitation intensity forecasts for 7 of the 8 stations. But this slight decrease in error by approximately 1mm is not a significant improvement to the already high RMSE values of the Eta forecasts. It was also noted that the stations, Bethlehem, Van Reenen, and the Royal Natal Nat. Park, which are surrounded by more complex topographical features have the highest RMSE values for both the Eta model and the NN model. The RMSE value can be decomposed into the bias value and the standard deviation value, using the equation MSE = $Bias^2$ + Variance. Figure 4.34 and 4.35 shows the bias and the standard deviation for each of the stations for the summer months. According to these two images the standard deviation is the largest contributor to the RMSE. The large standard deviation values are due to the inability of both the Eta and NN model to represent the variability characteristic of

precipitation. NN forecasts at 5 of the 8 stations proved to have smaller bias values than the Eta precipitation forecasts.

The precipitation intensity forecast for the winter months show lower RMSE values than experienced during summer months at the stations except at the Frankfort station (Figure 4.36). This might be due to a lack of winter precipitation observations at this station during the training and verification process. Once more the NN forecasts at 5 of the 8 stations showed lower bias values than the Eta forecasts (Figure 4.37). The standard deviation (Figure 4.38) of the NN forecasts and the Eta forecast are again the major contributor to the error.

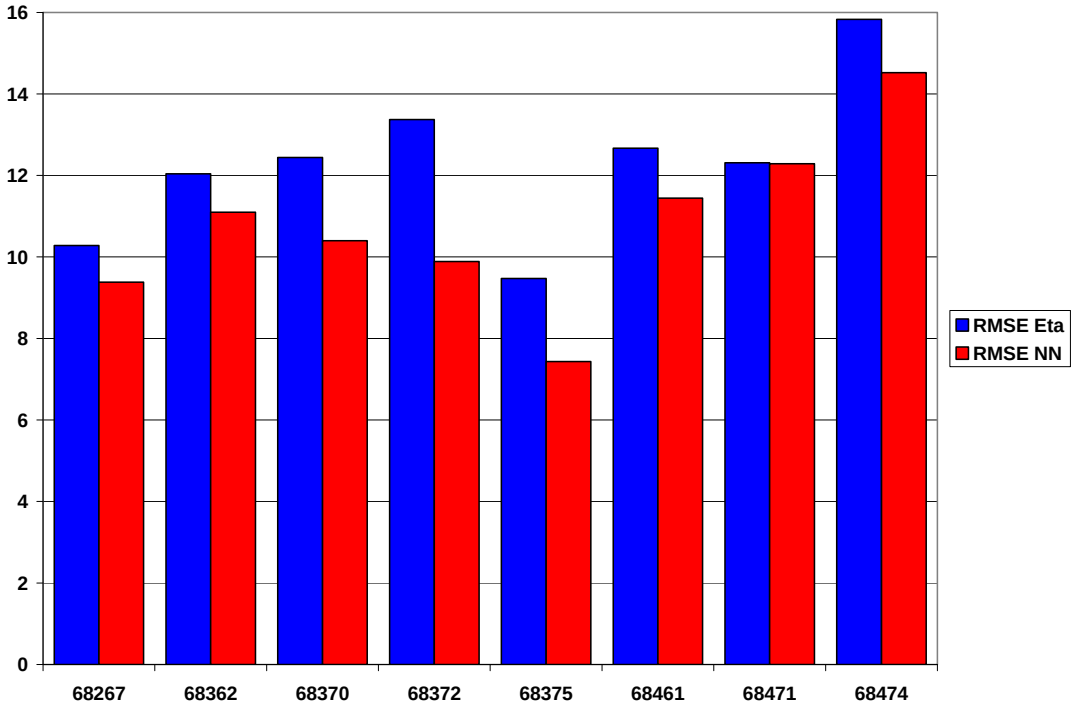

Figure 4.33 RMSE values for the summer months for each of the 8 selected stations.

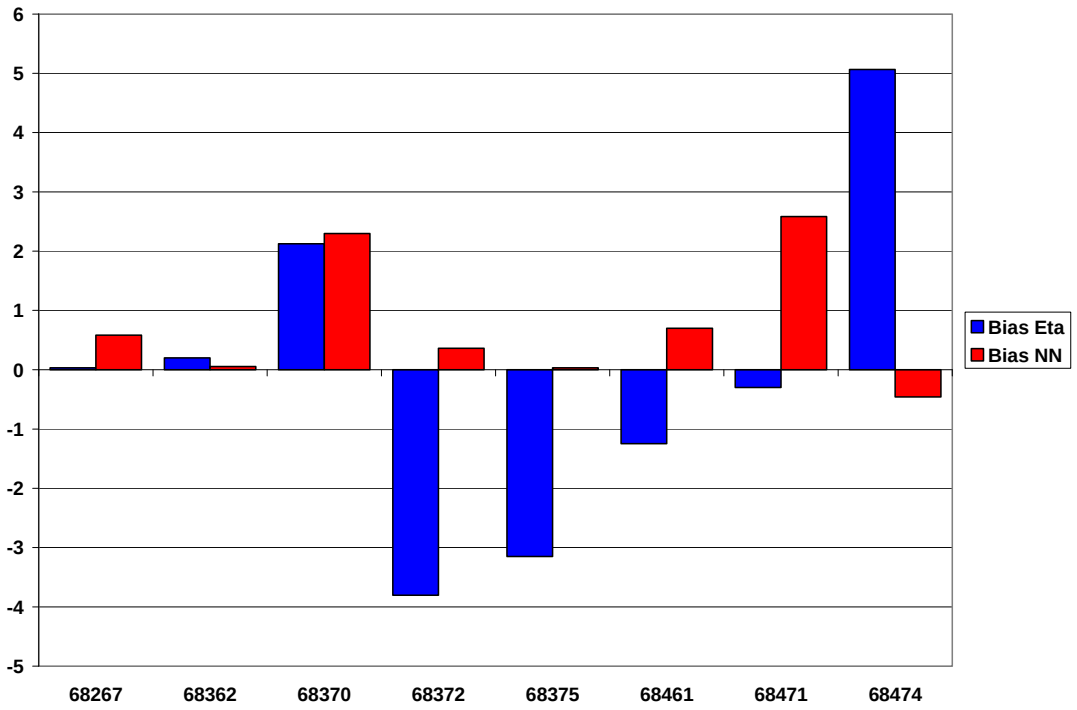

Figure 4.34 **Bias values for the summer months for each of the 8 selected stations.**

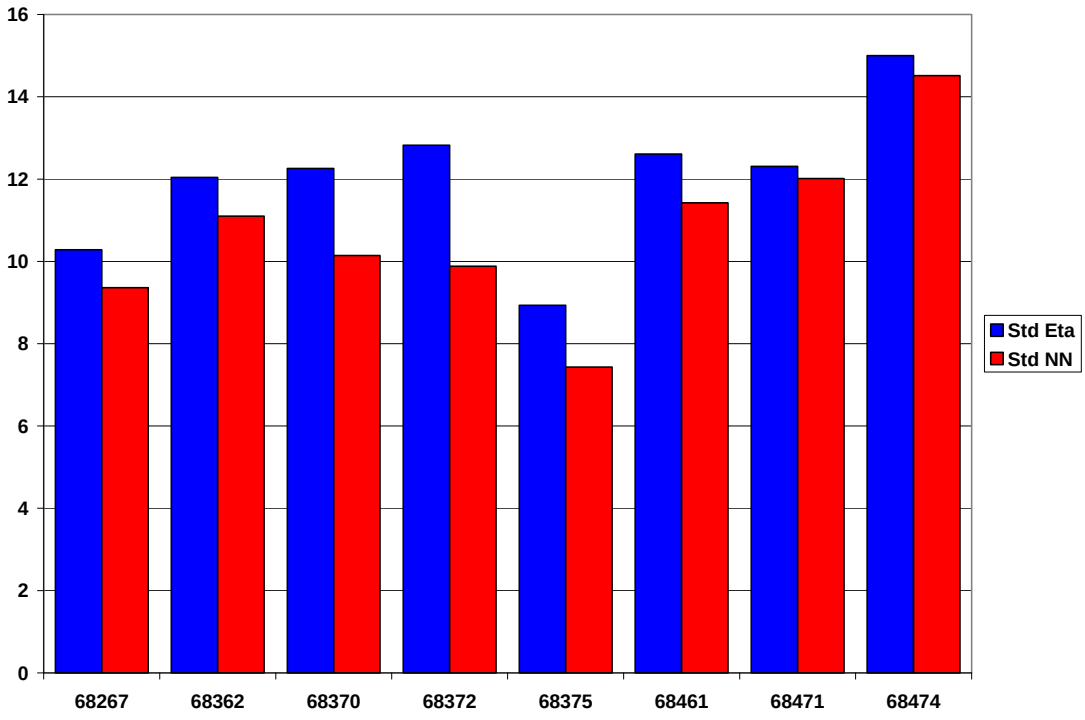

Figure 4.35 **Standard deviation values for the summer months for each of the 8 selected stations.**

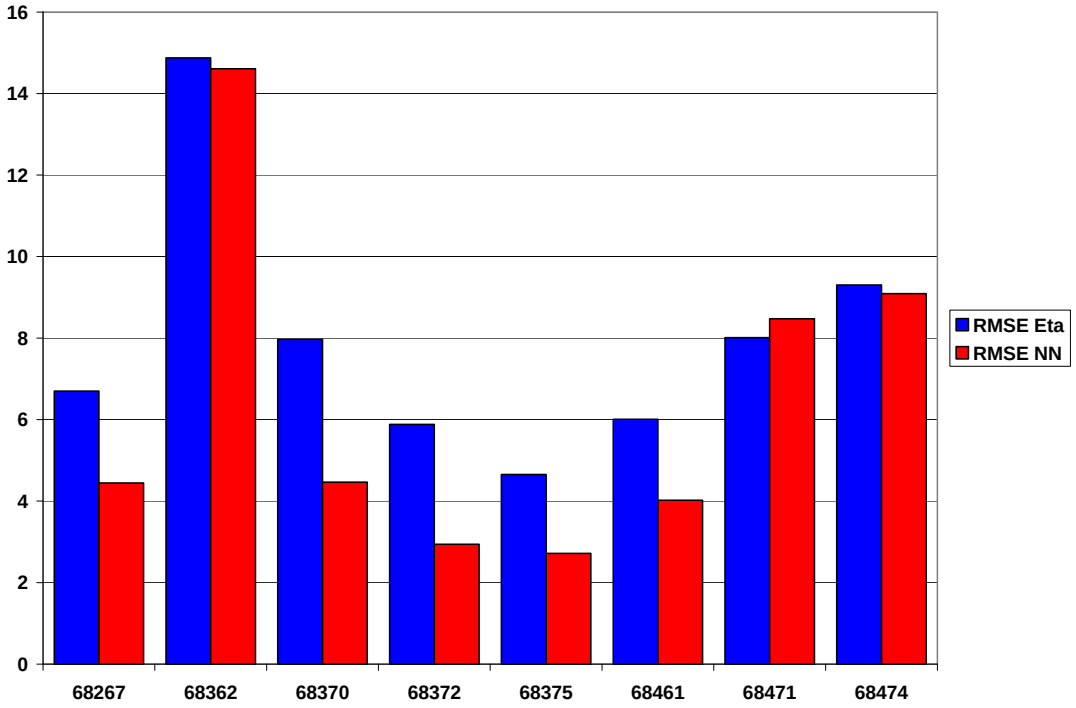

Figure 4.36 **RMSE values for the winter months for each of the 8 selected stations.**

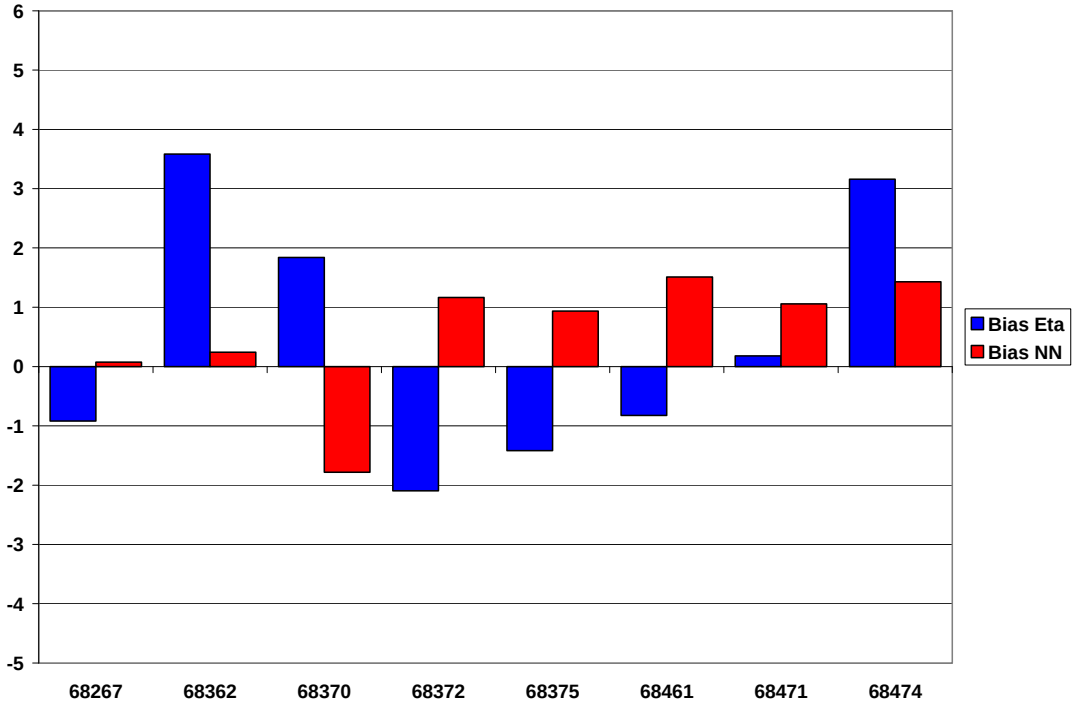

Figure 4.37 **Bias values for the winter months for each of the 8 selected stations.**

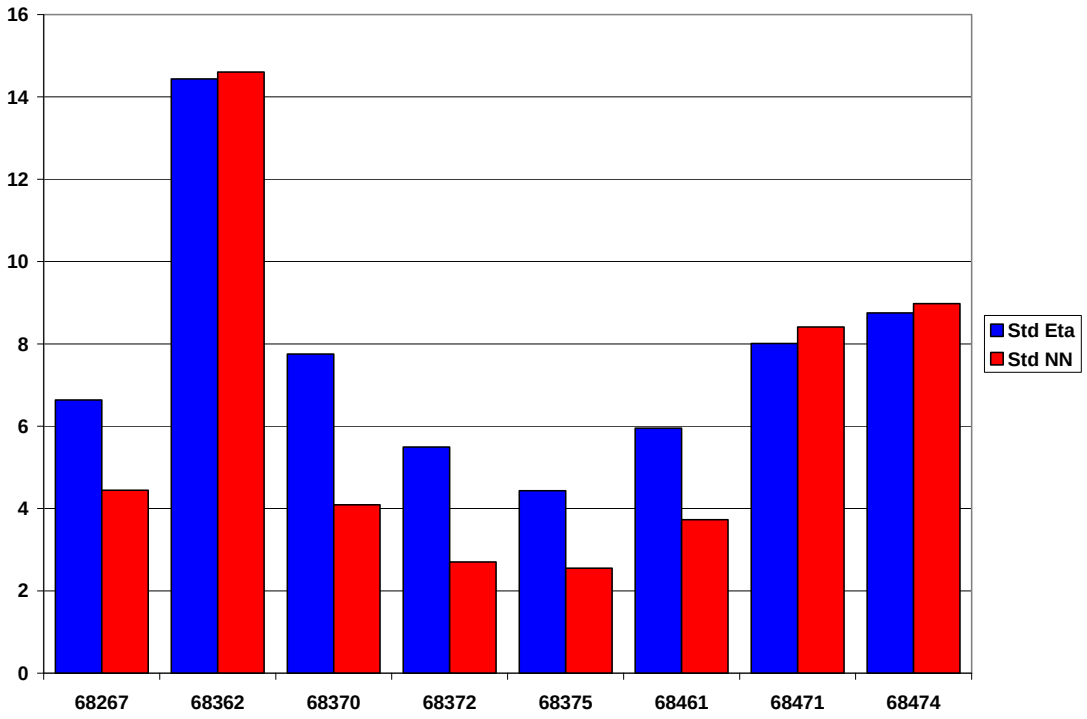

Figure 4.38 **Standard deviation values for the winter months for each of the 8 selected stations.**

The correlation coefficient describes the degree of association that exists between a forecast and observation. Figure 4.39 and 4.40 shows the correlation coefficient for summer and winter months, respectively. The correlation coefficient for the summer Eta precipitation forecasts are very low and trivial, ranging from -0.1 to 0.3. The NN precipitation forecasts could not improve on this. Only the correlation coefficient at the Bethlehem and Vrede station was slightly higher than that of the Eta model, but is still a trivial value. Forecasts for the winter months show higher association with observations. The correlation coefficient for these months ranged from -0.12 to 0.65 for the Eta model and from -0.15 to 0.85 for the NN model. There was a relatively good increase in correlation coefficient for 6 of the 8 synoptic stations when using the NN forecasts over the forecasts of the Eta model. The two stations which showed a decrease in correlation coefficient when using the NN forecasts were the Van Reenen and Royal Natal Nat. Park.

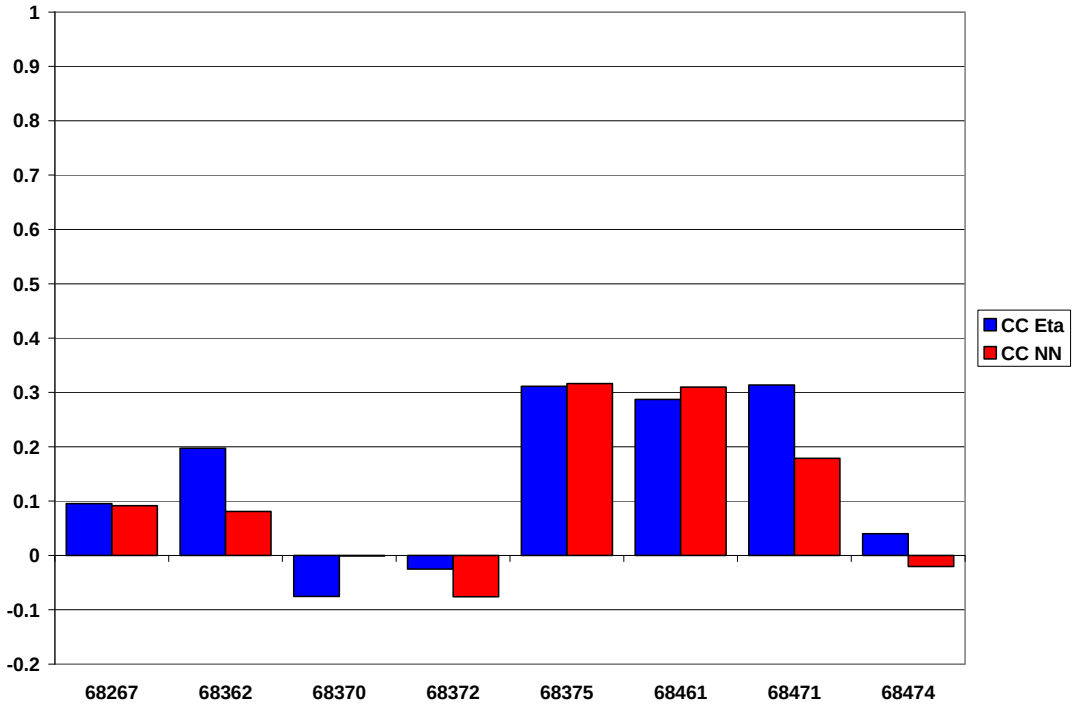

Figure 4.39 **Correlation coefficients for the summer months for each of the 8 selected stations.**

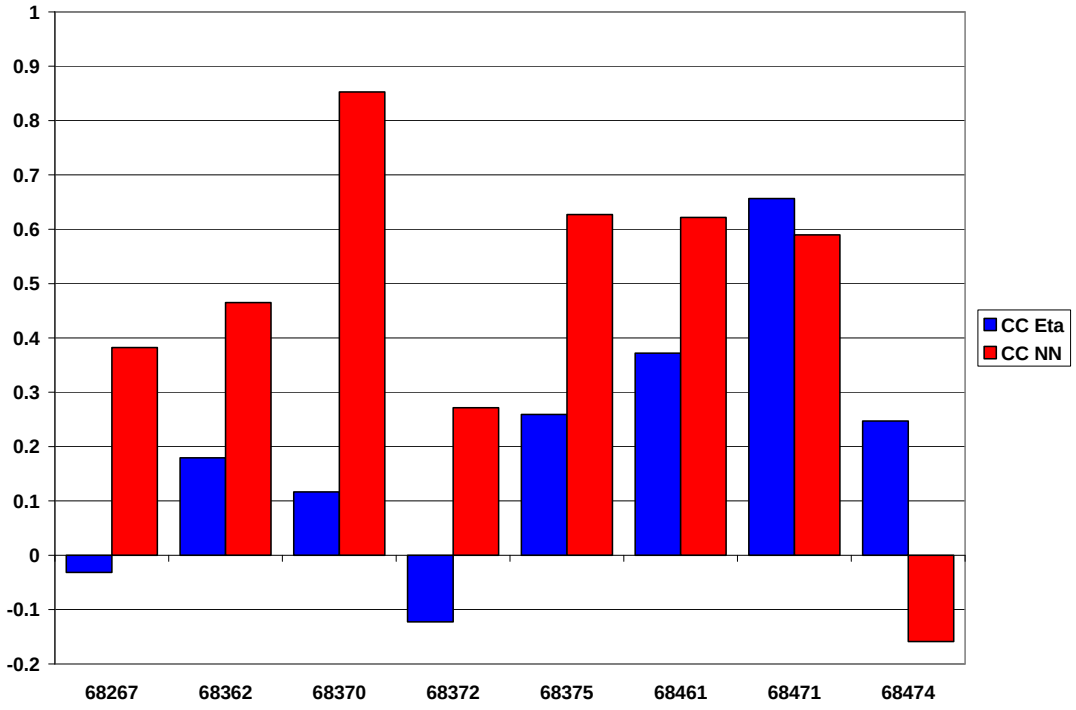

Figure 4.40 **Correlation coefficients for the winter months for each of the 8 selected stations.**

A summary and conclusion on the findings drawn from the statistical scores in this chapter will be given in chapter 5.

CHAPTER 5

DISCUSSION

5.1 Revisiting the aim of the study

A NN is a statistical post-processing technique or MOS technique that trains on a comprehensive data set of cases, containing model forecasts and the relevant observations, to learn the function which associates the predictors with the predictand. The trained function uses future model forecasts to forecast future predictands.

In chapter 1 the following hypotheses were made:

a. The application of NN's on model forecasts improves 14-day Tmin and Tmax forecast in comparison to temperature forecasts rendered by the model and by the forecaster.
b. The application of NN's on model forecasts provides 24-hour PoP forecasts which show improved forecast resolution to the forecaster PoP forecast.
c. The application of a NN on model forecasts improves the 24-hour model QPF in terms of bias and variability.

Sections 2 to 5 describe the findings made in the process of testing the hypotheses of this study.

5.2 Temperature Neural Network Forecasts

The Tmin and Tmax NN's were set up using 14-day forecast variables from the control run of the NCEP EPS. The networks were trained on 3 years and 9 months of archived model data and were verified on a year's model data which are independent of the training data set. The model data consist of 8 forecast variables i.e. mean sea level pressure, 700 hPa relative humidity, 10 meter winds, 500 hPa geopotential heights, precipitation amounts, and temperature. The forecast variables as well as the

observations were scaled to a range of [-1;1] to simplify the training process. NN models were trained to forecast Tmin and Tmax forecasts for a coastal and an inland station, at the Cape Town and Pretoria stations, respectively. Different amounts of hidden nodes (0, 2, 4, 8, 16) were used during training. No tendency was found on what amount of hidden nodes to use to train the optimal network. There were a lot of linear networks (networks with no hidden nodes) that proved to be the best networks, although there were a few networks performing well with 2, 4, and 8 hidden nodes.

According to the RMSE values and the correlation coefficients the forecaster's Tmin and Tmax forecasts for forecasting day 1 are still irreplaceable by any of the modeling systems. Except for the day-1 Tmax forecasts at the Pretoria station where the NN showed quite similar performance to the forecaster. The forecasters and the NN day-2 Tmin and Tmax forecasts at the Cape Town station showed similar performance. At the Pretoria station the forecaster was still the best performer for the day-2 Tmin temperature forecasts while the NN showed major performance for the day-2 Tmax forecasts.

The benefit to use a NN to enhance modeled temperature forecasts starts to show from approximately the day-2 forecasts. The performance of the NN and the NCEP ENS for the day-2 to day-7 forecasts at the Pretoria station is very close to each other but the NN proves to be the best in terms of the correlation coefficient for the Tmin's and in terms of the RMSE for the Tmax's. The NN forecasts at the Cape Town station shows a large improvement over the NCEP forecasts in terms of the RMSE although the correlation coefficients are quite similar. The NN also shows improved performance over the forecaster for the day-3 to day-7 temperature forecasts except for the Tmin forecasts at the Cape Town station where the performance are rather similar.

5.3 Best Predictors for Precipitation Forecasts

A bivariate analysis was made on the different forecast variables of the Eta model to find the best predictors. Table 5.1 shows the best predictors which recurred at 5, 6, 7, or 8 out of the 8 stations.

Summer (Oct-Mar)		Winter (Apr-Sep)	
Predictors	**Recurred at #stations**	**Predictors**	**Recurred at #stations**
Mid-level cloud cover (%) : 12z, 18z	5, 8	Mid-level cloud cover (%) : 6z, 12z, 18z, 24z	7, 6, 8, 6
Accum. Total Precipitation (mm) : 6z - 30z	7	Accum. Total Precipitation (mm) : 6z – 30z	7
700 hPa Relative Humidity (%) : 12z, 30z	6, 5	700 hPa Relative Humidity (%) : 24z	7
Total cloud cover (%) : 6z, 24z	5	Total cloud cover (%) : 6z, 12z, 18z, 24z	7, 8, 7, 8
500 hPa Relative Humidity (%) : 18z	5	Soil Water (mm) : 30z	6
Low level cloud cover (%) : 18z	5	Precipitable Water (mm) : 12z	5

Table 5.1 **The best predictors which recurred at 5, 6, 7, or 8 out of the 8 selected stations using a bivariate analysis.**

According to the analysis mid-level cloud cover, 24-hour accumulated total precipitation, 700hPa relative humidity and total cloud cover were found to be important predictors for both summer and winter precipitation. Mid-level cloud cover at 18z and accumulated total precipitation are the prominent predictors for both seasons. Clouds types at mid-levels are classified as altocumulus and altostratus which are at heights of approximately 2 – 6 km. Altocumulus are formed by convection or waves near mountains and produce virga (inclined trails of precipitation) or occasional light showers. Altostratus are also accompanied by rainfall and are formed by the presence of widespread ascent. Cumulonimbus clouds also contribute to the percentage of cloud cover at the mid-levels and are formed during the incidence of surface heating, instability and convection. These clouds are mostly accompanied by showers and thundershowers which can develop into heavy precipitation events.

Best predictors for summer precipitation tend to be concentrated during the afternoon hours as precipitation at the 8 stations are mostly during these hours. The 500 hPa relative humidity and low level cloud cover are additional predictors to summer months which are not present as predictors for winter precipitation.

Winter precipitation is dependent on cloud cover from the 6z hour to the 24z hour, especially at mid-levels. Additional predictors to this season are the amount of precipitable water at 12z and the amount of soil water at 30z. Therefore the information on the availability of water in the atmosphere during the afternoon and the amount of deposited water on the ground at the end of the 24-hour forecast period indicates the occurrence of winter precipitation. Precipitable water is defined as the

total atmospheric water vapour contained in a vertical column of unit cross-sectional area extending between two specified levels. Precipitation amount usually exceeds the amount of precipitable water due to convergence of water vapour from surrounding areas but are yet correlated with each other (AMS Glossary).

5.4 Probability of Precipitation Neural Network Forecasts

The NN PoP forecasts were set up using variables from the 06Z to 30Z NCEP Eta model forecasts to render the probability of precipitation occurring during a 24-hour period. The networks were trained on 5 years and 5 months of archived model data and verified on different set of model data stretching over a period of 1 year. The Eta model has a large amount of different forecast variables and therefore a bivariate method was applied on the variables to select the best predictors. The best predictors for this model are discussed in section 3 of this chapter. Precipitation observations at eight synoptic stations in the vicinity of the Vaal Dam catchments were used as predictands in the networks and were scaled to binary events (1) and no-events (0). The output of the NN's are in the continuous range [0.0;1.0] which represents the probability of precipitation. The optimal amount of hidden nodes to use from the set (0, 2, 4, 8, 16) to get the best performing network are not clear and vary according to station and season.

According to the Bss the NN PoP forecasts have positive skill over sample climatology for both summer and winter months. The summer PoP forecasts by the NN's prove to be more reliable than the winter PoP forecasts which contain more resolution than the summer PoP forecasts but to the expense of reliability. The reliability aspect of the PoP forecasts for both seasons seems to be the main contributor to the skill of the NN model. The resolution aspect of the PoP NN forecasts are much smaller in amount but are still a positive contributor to skill over climatology. The resolution of the PoP NN forecasts was further investigated by means of the area below the ROC curve. ROC areas for both seasons were higher than the 0.5 threshold which indicates that forecasts have skill. The results again show that the winter PoP forecasts have slightly more resolution. Therefore the PoP forecasts by

the NN's do have the ability to discriminate between no-precipitation and precipitation events. The PoP NN forecasts are therefore potentially useful.

The performance of the PoP NN forecast relative to the Eta model and the human forecaster's forecasts of rainfall occurrence where investigated by looking at the frequency distributions over different probability categories. By stratifying the 0 to 10% NN PoP forecasts as the no-precipitation category it was found that the NN's showed no improvement over the Eta model for no-precipitation events and significant improvement over the Eta model for precipitation events. To equalize the instability between the performances of the different events the no-precipitation category of the NN forecasts were re-stratified to include probabilities from 0% to 20%. Changing the stratification boundaries improved the performance of the PoP NN forecasts for summer months by so much to exceed the performance of the Eta model as well as the PoP forecasts of the human forecaster, except during precipitation events at the Standerton station. During the no-precipitation events for the winter months the PoP NN forecasts performed better than the Eta model forecasts after restratifying the categories, but equal performance were detected between the NN forecasts and the human forecaster's forecasts. The performance of the NN for the few precipitation events during the winter months showed to be better than the forecaster's forecasts and approximately similar for the Eta model forecasts for two of the three available stations, before changing the stratification. Re-stratifying the NN forecasts during the winter precipitation events worsened the performance.

Changing the stratification of the Pop categories according to which PoP forecasts can be regarded as Yes or No forecasts is explained by the Signal Detection Theory (SDT). Mason (1982) introduced this verification procedure to meteorology. It falls in a group of measures which require the stratification of observations. SDT is also used by other research fields which need to make use of decision-making techniques, such as diagnosing cancer (Swets and Pickett, 1982) and analyzing polygraph tests (Szucko and Kleinmuntz, 1981). SDT verifies the ability of forecasts to discriminate between two distinct events while assessing nothing on the reliability of the forecasts. It also has the ability to portray the cost of increased false alarms when decision thresholds are relaxed for severe weather forecasting. The decision threshold is the probability value which separates the forecasts in Yes and No events. The decision on a threshold

value depends on the apparent costs associated with missed events. SDT makes use of two functions namely the hit rate and the false alarm rate which are determined through a contingency table by the functions, [hits / hits + misses] and [false alarms / false alarms + correct rejections], respectively. The ROC curve contains both the hit rate and false rate values for different decision thresholds. Figures 5.1 to 5.6 show the ROC curves of the summer and winter PoP's of the NN and the forecaster for the Ermelo, Standerton and Bethlehem stations.

The NN PoP's and forecaster PoP's prove to have positive skill at all three stations and for both seasons. The NN PoP's also proved to have more resolution or the ability to discriminate between rain and no rain events by the areas under the ROC curves. The 30% PoP forecasts for summer months (Figure 5.1, 5.3, 5.5) by the forecaster have high hit rates (80-90%), but also relatively high false alarm rates (40-60%). But then moving just to the next forecaster PoP category of 60%, there is a great reduction in false alarms (10%) and hits (30-50%) during summer. This show that the forecaster PoP lack some discriminating categories between the 30% and 60% category. In most of the cases the 60% forecaster PoP and the 70% NN PoP are close to each other on the ROC graphs, showing that they are relatively similar in skill. Therefore the NN PoP's succeeded in adding more discriminating categories in comparison to the forecaster PoP's. A similar trend comes about for the winter PoP's, with a large discrepancy between the 30% and 60% forecaster PoP categories regarding the amount of hits and false alarms. Very few warnings of 80% and above chance of rainfall are issued by the forecaster during summer, and for winter months there are very few warnings of 60% chance of rainfall and above. For NN PoP's the 0% and 90-100% chance of rainfall categories are almost never issued, while only a few warnings occurred in the 70-80% categories.

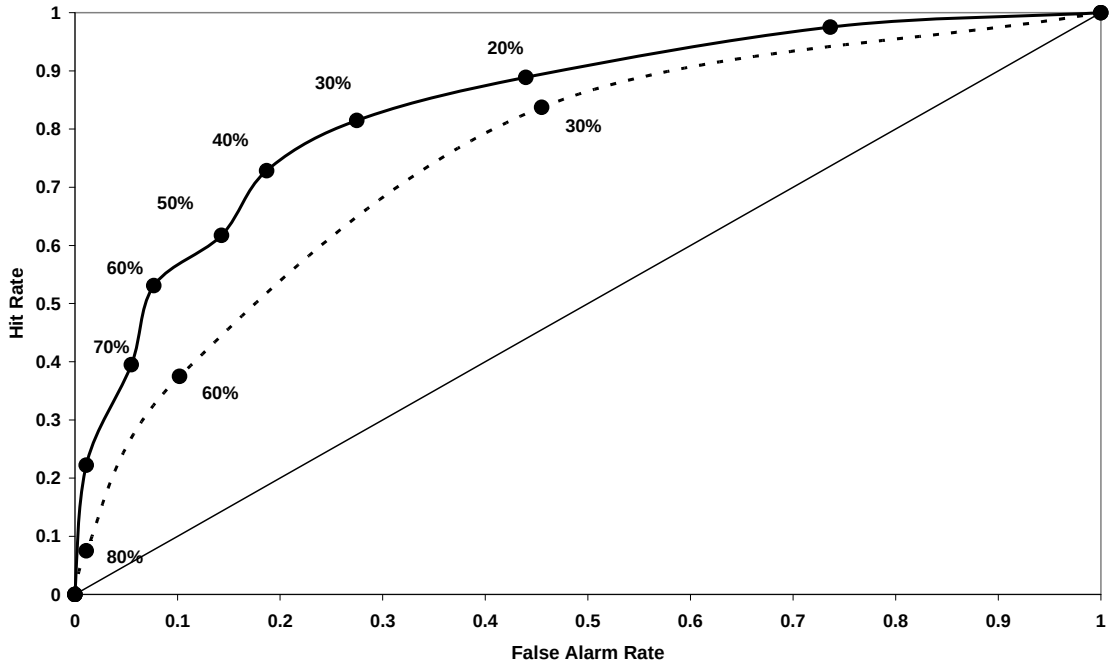

Figure 5.1 **The ROC curves of the summer PoP's by the neural network (solid line) and the forecaster (dashed line) for the Ermelo station.**

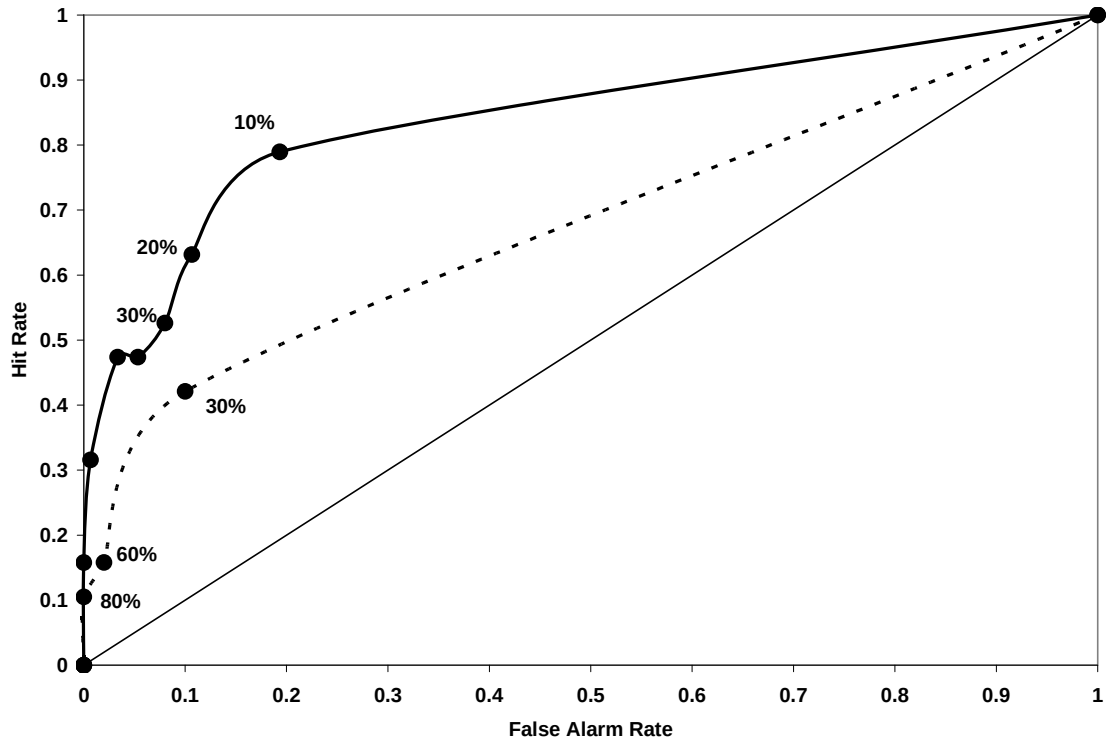

Figure 5.2 **The ROC curves of the winter PoP's by the neural network (solid line) and the forecaster (dashed line) for the Ermelo station.**

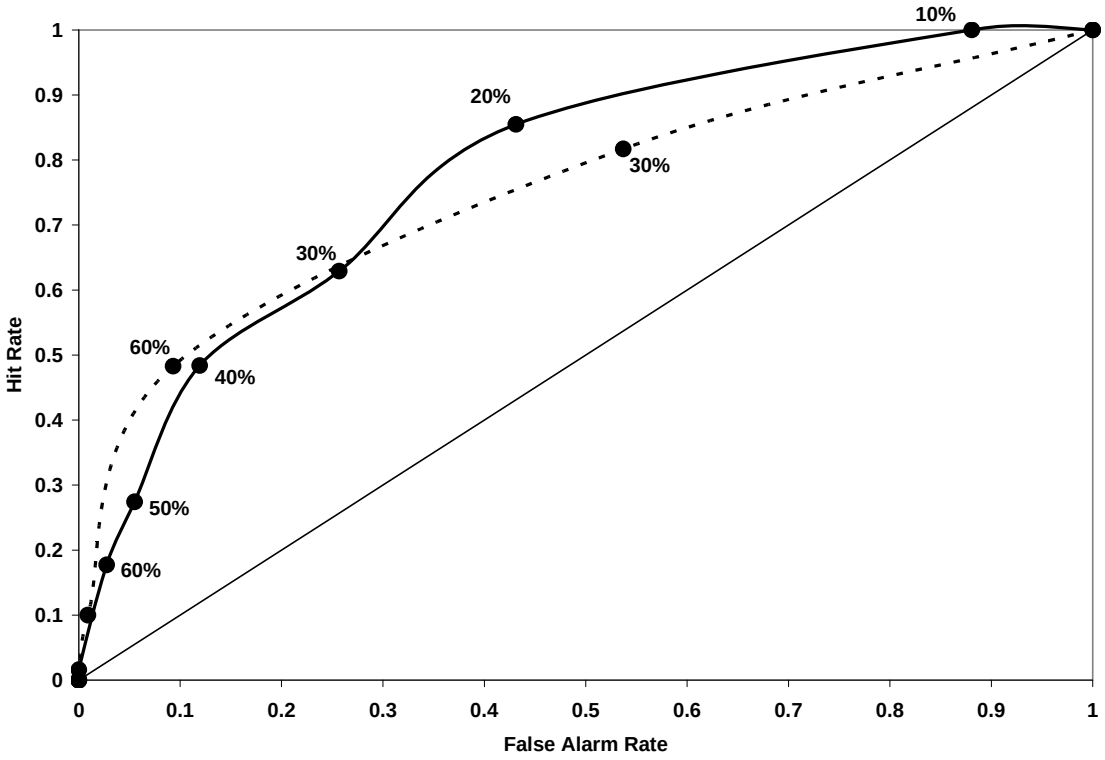

Figure 5.3 **The ROC curves of the summer PoP's by the neural network (solid line) and the forecaster (dashed line) for the Standerton station.**

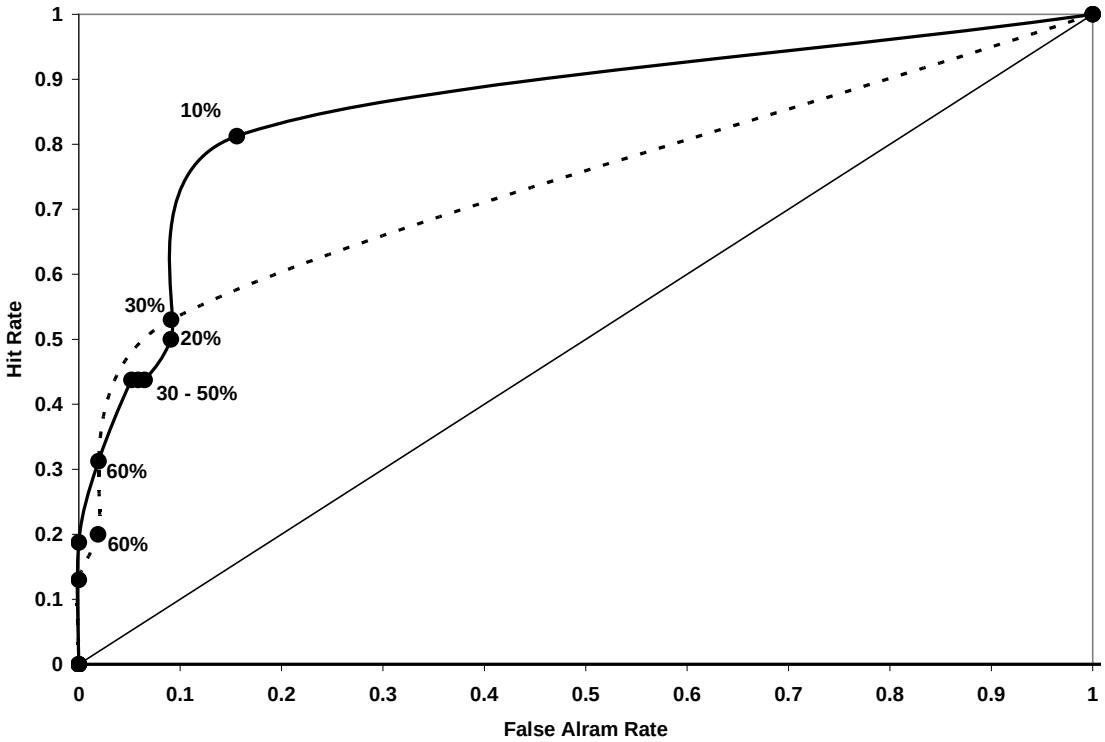

Figure 5.4 **The ROC curves of the winter PoP's by the neural network (solid line) and the forecaster (dashed line) for the Standerton station.**

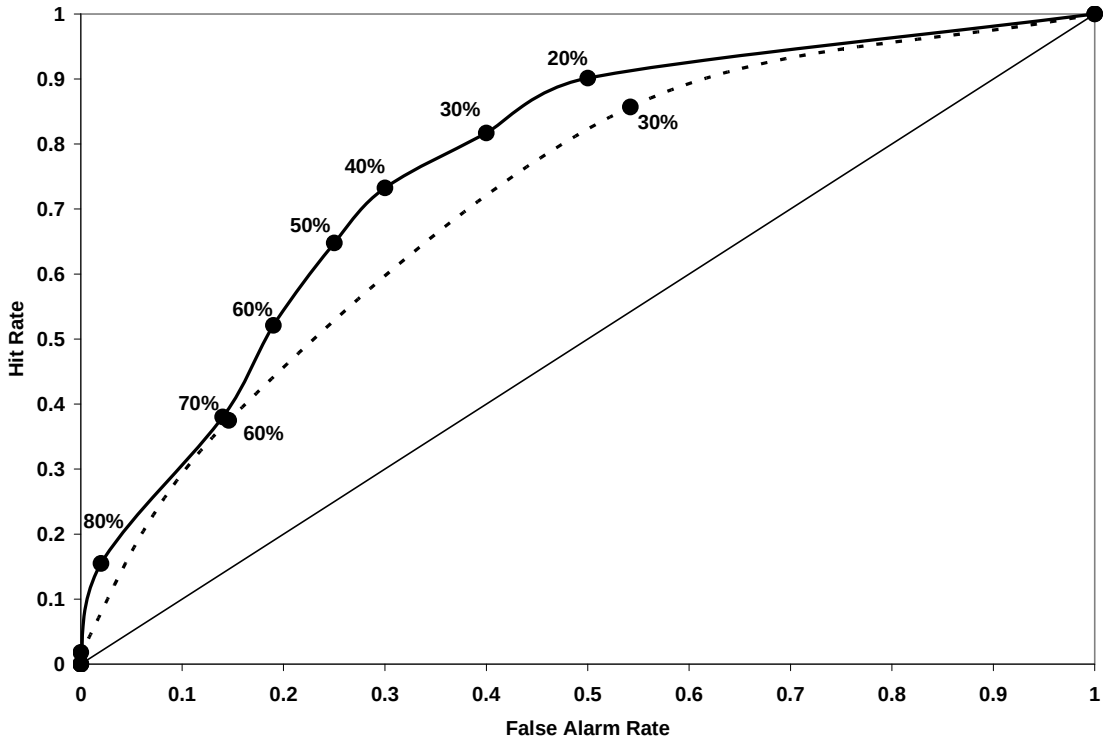

Figure 5.5 **The ROC curves of the summer PoP's by the neural network (solid line) and the forecaster (dashed line) for the Bethlehem station.**

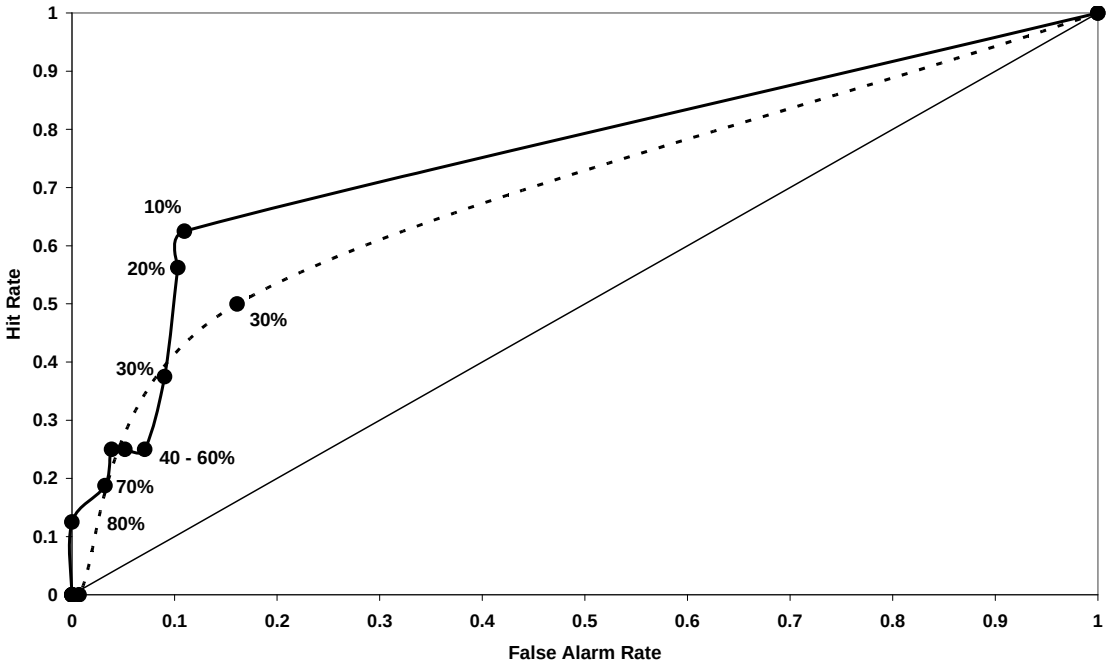

Figure 5.6 **The ROC curves of the winter PoP's by the neural network (solid line) and the forecaster (dashed line) for the Bethlehem station.**

PoP categories close to the top right corner of the ROC curve are called relaxed warnings, because a large number of warnings are issued in these categories which coincide with a large amount of hits and false alarms. If it was decided that an

acceptable forecast category has a hit rate larger than approximately 60% and a false alarm rate less than approximately 40%, then only the neural net PoP's of 30-50% in the summer, and the NN PoP's of 10% in the winter, would have been acceptable. In the decision-making process it doesn't quite work this way. Different forecast users have different cost and losses regarding the amount of false alarms and misses a specific category has. A user will decide to put weight on warnings issued at low probabilities if the cost of a missed event is higher than the cost of a false alarm.

Forecast quality determines the performance of a forecast relative to the observed event, but forecast value determines the ability of a forecast to be useful in making decisions on which actions to follow regarding the probability of an influential event. If it cost an amount of C to safeguard against an event and the loss is an amount of L due to a missed event, then the total cost by a user due to forecast error and correct forecast is,

$$Total\ cost = (misses * L) + (false\ alarms * C) + (hits * C) \tag{5.1}$$

Thornes and Stephenson (2001) develop a Value index, V (Equation 5.2) which varies between 0 and a perfect score of 1.

$$V = \frac{E(S) - E(A)}{E(S) - E(P)} \tag{5.2}$$

E(S) is the expense made when safeguarding actions are taken regardless of the forecast. E(A) is the expense made when safeguarding actions are taken regarding the forecast suggestion. When a forecast system is perfect or correct every time an expense of E(P) are made. It is known that E(S) = (hits + false alarms + misses + correct rejections) * C, and E(A) = Total cost (Equation 5.1), and E(P) = (hits + misses) * C.

The Relative value (RV) or Value score (Richardson, 2000, Wilks (2001) determines the value of a set of forecasts. It is a skill score of expected expense, with climatology as the reference forecast, and described by,

$$V = KSS - (1 - POD)\frac{\left(P_{c\lim} - \frac{C}{L}\right)}{\frac{C}{L}\left(1 - P_{c\lim}\right)} \qquad \text{if} \quad \frac{C}{L} < P_{c\lim}$$

$$V = KSS \qquad \text{if} \quad \frac{C}{L} = P_{c\lim}$$

$$V = KSS - PFD\frac{\left(\frac{C}{L} - P_{c\lim}\right)}{P_{c\lim}\left(1 - \frac{C}{L}\right)} \qquad \text{if} \quad \frac{C}{L} > P_{c\lim}$$

with

$$KSS = \frac{(hits * correct\_rejections - false\_alarms * misses)}{(hits + misses)(false\_alarms + correct\_rejections)}$$

$$POD = \frac{hits}{(hits + misses)}$$

$$POFD = \frac{false\_alarms}{(false\_alarms + correct\_rejections)}$$

(5.3)

KSS is the Kuiper Skill Score (Hansen and Kuipers 1965; Murphy 1996). Buizza (2001) found from his results that the RV of a forecast system is sensitive to imposed random and systematic errors.

The RV's for each probability category are plotted against a range of cost/loss ratios (C/L) as every different user has a unique cost/loss ratio when taking action regarding a particular event. The envelope of RV curves gives an indication of the potential value of a forecasting system for different cost/loss ratios and is related to the resolution of the forecast. Each user can choose an optimal probability threshold for taking action according to their unique cost/loss ratio, if the forecast probability is larger or equal to the threshold. The maximum forecast value are reached when the cost/loss ratio are equal to the climatological probability of an event.

Figures 5.7 to 5.10 show RV plots for PoP forecasts at the Ermelo station. The summer PoP forecasts by the forecasters (Figure 5.7) reached a maximum RV of 0.38 by the 30% PoP forecasts. The range of user cost/loss ratios covered by these forecast

are from approximately 0.21 to 0.84. Therefore users with lowest cost/loss ratios (close to 0) or the highest cost/loss ratios (close to 1) will not find use in these forecasts. The summer PoP forecasts by the NN (Figure 5.8) reached a maximum RV of 0.50 by the 40% PoP forecasts. The range of user cost/loss forecasts covered by these forecasts is from approximately 0.11 to 0.91. Clearly there is an improvement in forecast value and cost/loss range of users over those reached by the forecaster PoP's.

For the winter months the forecaster PoP's of 30% (Figure 5.9) reached a maximum RV of 0.33. Only users with cost/loss ratios close to 0 and 1 will find no use in these forecasts. The winter PoP's of the NN (Figure 5.10) reached a maximum RV of 0.63 and the whole range of the user cost/loss ratios are useful except at ratios of 0 and 1. Again the NN PoP forecasts for the winter months have more economic value than the forecaster's PoP forecasts. The highest RV's for both POP forecast types are concentrated at the lower cost/loss ratios due to low frequency of precipitation events during the winter months. There was a larger improvement of NN PoP's over forecaster PoP's during winter months than what was found for the summer months.

For example, if a certain user has a cost/loss ratio of 0.4 then the optimal decision threshold for summer precipitation will be the 30% forecaster PoP (RV = 0.32) and the 30% NN PoP (RV = 0.43). For winter precipitation the optimal decision threshold will be the 80% forecaster PoP (RV = 0.11) and the 60% NN PoP (RV = 0.33).

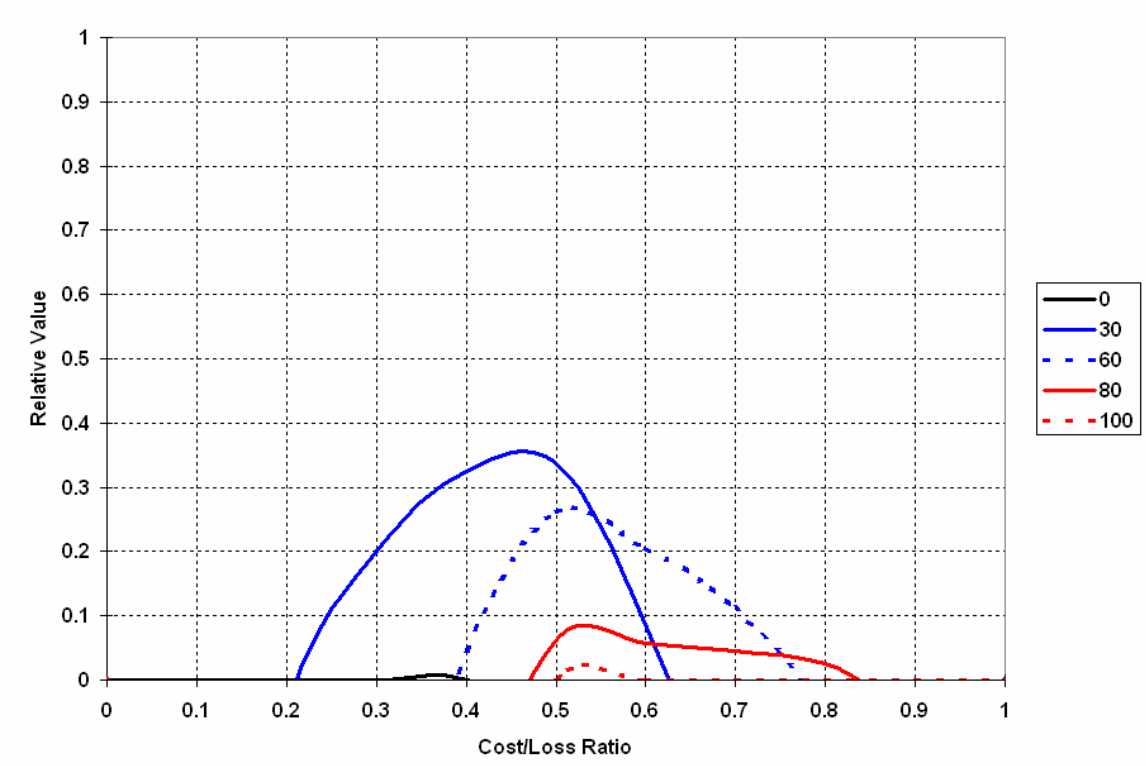

Figure 5.7 **The Relative Value (RV) plot for PoP forecasts at the Ermelo station made by the forecasters during summer months.**

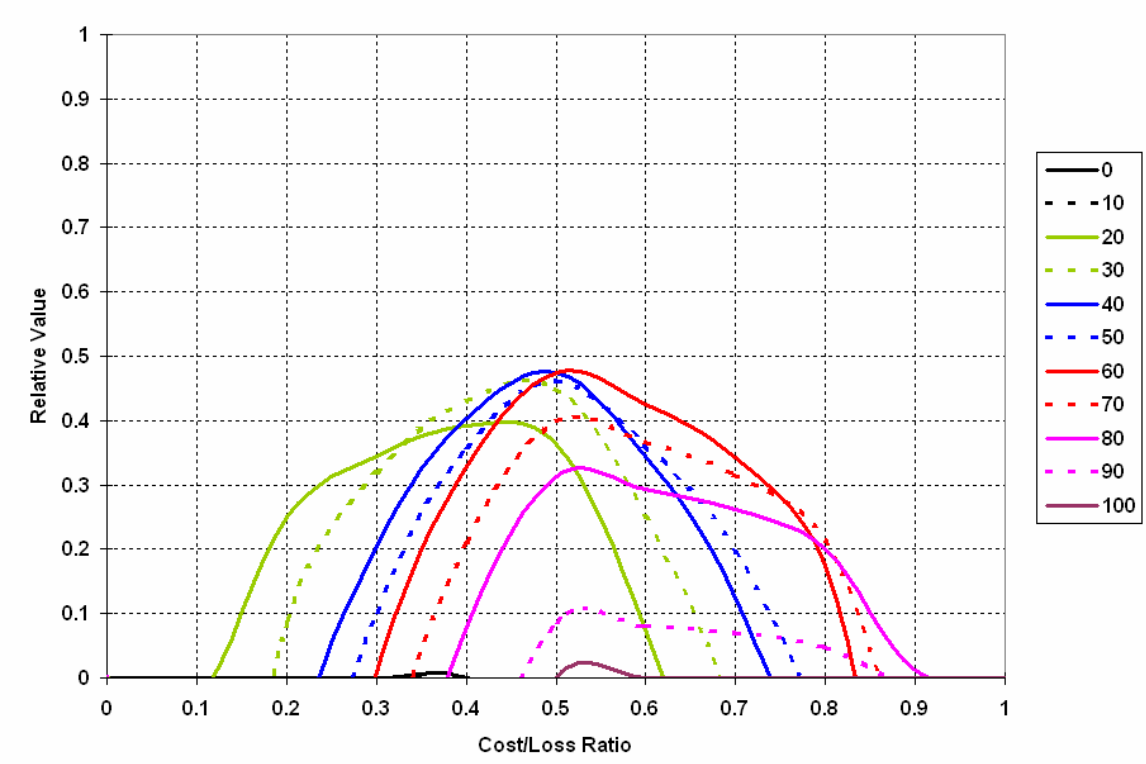

Figure 5.8 **The Relative Value (RV) plot for PoP forecasts at the Ermelo station made by the neural network during summer months.**

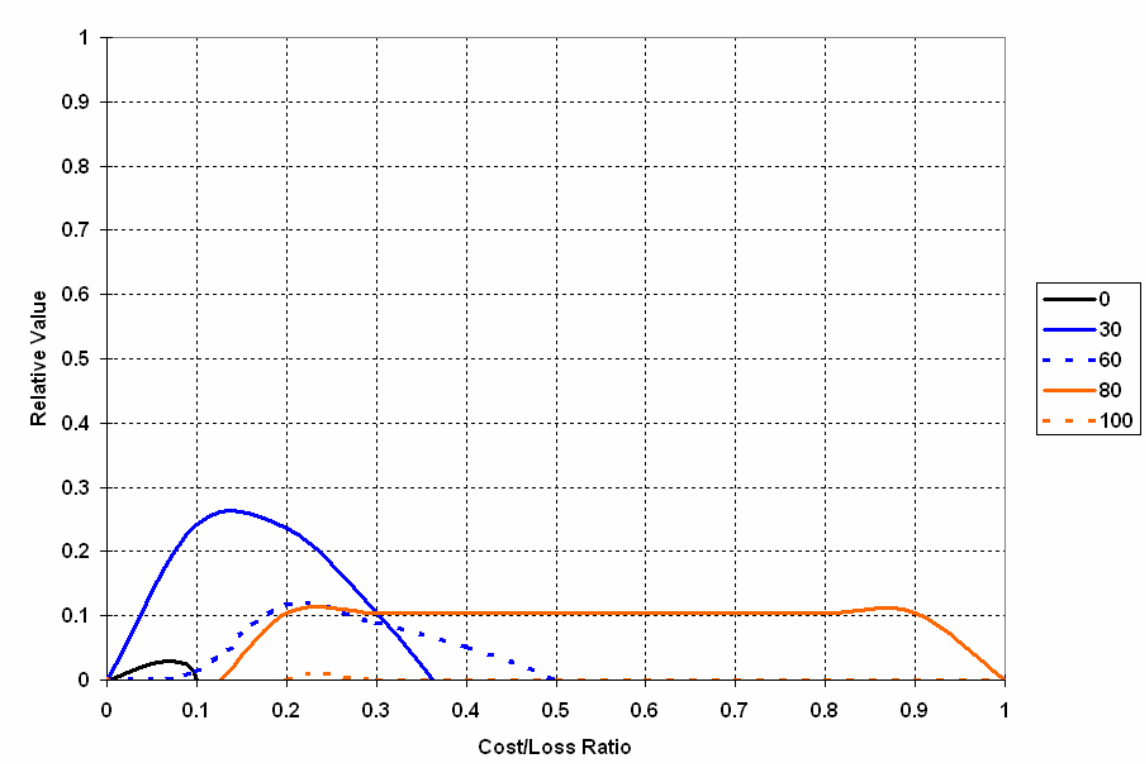

Figure 5.9 **The Relative Value (RV) plot for PoP forecasts at the Ermelo station made by the forecasters during winter months.**

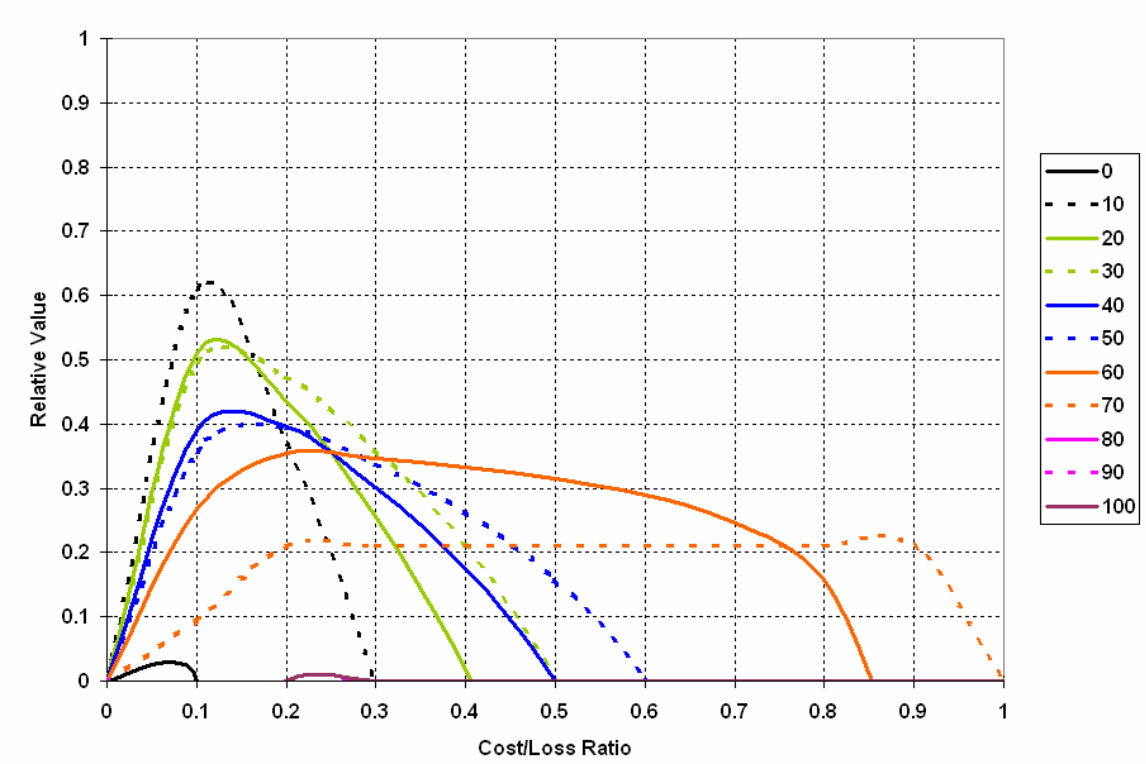

Figure 5.10 **The Relative Value (RV) plot for PoP forecasts at the Ermelo station made by the neural network during winter months.**

It was found during the assessment that the human forecaster has the tendency to be conservative regarding the forecasting of a precipitation event. The fraction of approximately 40% cases of the PoP forecasts by the human forecaster fell in the 30% probability of precipitation category, for both events of no-precipitation and precipitation. Doswell (2004) stated that there is a need to study the different cognitive approaches between human forecasters which perform well time after time and those that perform poorly, to be able to improve human forecaster products. He also mentioned that it is unclear how to merge analytical methods and products effectively with human intuitive approaches. Therefore it is a much more complex task to elevate human forecaster performance than it is to improve the performance of objective methods. Doswell also suggested that to achieve improved human forecaster performance there need to be constant teamwork between meteorologists, cognitive psychologists, and others concerned with decision-making processes.

5.5 Precipitation Intensity Neural Network Forecasts

Precipitation intensity forecasts by the NN were set up similarly to the PoP NN discussed in section 4 of this chapter, except that the predictors and the predictand are scaled to the range [-1; 1] to render 24-hour accumulated precipitation forecasts in amounts of millimeters when scaled back to the original scale.

Results show that the NN model was not able to enhance the Eta model forecasts in terms of precipitation intensity. The decrease in forecast error was insignificant for both summer and winter months. The NN was able to decrease the bias between forecasts and observations for some of the stations. The standard deviation between the forecasts and the observations proved to be the main contributor to the RMSE. The NN was able to decrease the standard deviation as well but the amount of improvement was insignificant. Therefore the NN performed well in improving the reliability of the precipitation forecast, but it could not improve on the resolution aspect.

Lack of resolution in a model cannot be improved by calibration of forecasts. The model needs to be redesigned in terms of input data and network architecture or in the case of a NWP model, the initial conditions and model physics needs to be redesigned. The outcome of these results should in actual fact not be a surprise as it can't be expected to enhance the precipitation intensity forecasts of the Eta model which does not initially have the skill in resolution to forecasts precipitation in terms of amounts. According to the verification results described on the Eta precipitation amount forecasts in the previous chapter this model tends to forecast precipitation amounts in the lower range of thresholds regardless of the type of precipitation event.

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Outcome of research

The outcome of applying a NN system on NWP model data has proved to be instructive. The hypotheses which were mentioned in chapter 1 of this dissertation proved to be fairly true, except for hypothesis c). The outcome of hypotheses a) to c) is the following.

a. Temperature forecasts by the NN's shows less error than the NCEP ensemble temperature forecasts and even exceed the performance of the forecaster from day-3 forecasts and onwards. Diurnal variability at coastal stations was much better presented by the NN than by the NCEP ensemble model due to the coarse resolution of the model.
b. NN PoP's have more resolution than those of the forecaster. The relative economic value of the NN PoP to users is higher than the value of the forecaster PoP's.
c. NN's forecasting precipitation amount are not able to enhance modeled precipitation amounts. The reduction of errors is insignificantly small. The reliability of the forecasts improved slightly but there is still a large error present due to the inability of the NN to correctly represent the variability of precipitation amounts.

6.2 Implications

1. Neural network forecasting improves temperatures forecasts at the longer lead time range and gives more detail and value to PoP forecasts.
2. The forecaster will benefit in the use of NN temperature forecasts as a guide from day 3 and onwards. The forecaster can also use the NN PoP forecasts as a guide to add more probability categories to their PoP forecast.

3. The general public can have more confidence in their decisions when planning ahead for a few days on events affected by the day's temperature. Businesses can add more value to their cost/loss management by using the PoP forecasts generated from a NN system.
4. The complexity and difficulty to forecast realistic precipitation amounts has spurred the need to use a different approach to forecasting precipitation. Issuing PoP forecasts, especially through a NN system, already proved to be beneficial. This inspires the need to search for more different approaches to forecasting precipitation, like for example generating PQPF's using NN's or other statistical forecasting techniques.
5. Statistical post-processing of model forecasts is an inevitable task and should form part of the operational NWP environment. Possible advantages from other existing and new MOS techniques need to be explored by the researcher.

6.3 Caveats

1. A definite disadvantage of applying a NN system is the requirement of such a system to use data from a frozen model. A frozen model is a model which didn't have any modifications or upgrades done on it since the date the archiving of the model data started. This is not always feasible for an organization to have a frozen model additional to an operational model which needs constant upgrading to stay abreast with new developments.
2. Predictor-predictand relationships are not stationary due to changes in the climate that come about in time. Assuming stationarity in the development of a NN system might be successful in generating forecast in present time but does not reassure a successful system for future climate conditions. Transfer functions become invalid and weights attached to units change. Including suitable low-frequency predictors such as geopotential height might partially solve the problem. Hewitson (1999) declared that atmospheric moisture content has control over present precipitation occurrence but will exert even more control over future precipitation occurrences. Hewitson (2004) also found that assuming stationarity for geopotential heights is acceptable when generating forecast for the 2080's using 1961-1990 climatology, but collapse

6.4 Recommended future research

1. A NN system needs to be developed which can adapt to model changes to address the disadvantage of using a frozen NWP model.
2. NN forecasting need to be applied on severe weather forecasting in terms of thunderstorm, tornado and hail probabilities.
3. A NN system needs to be developed to forecast categorical precipitation amounts and to generate PQPF's. This can be done by selecting certain significant thresholds to divide precipitation intensity forecasts into categories for example, light, moderate and heavy rainfall. This method can even be applied on temperature forecasts to forecast the level of temperature intensity rather than the exact measurement in °C.
4. The effect of increasing model resolution on the accuracy of NN forecasts need to be explored as the usual parameterized precipitation can be calculated explicitly. Increased resolution might increase the realism of the forecast, but verification scores mostly fail to proof improvement due to insufficient observations. High resolution forecasts also has a limited predictability to give the exact location of the event (Stein et. al., 2000, Mass et. al., 2002).

A NN system accounts for the systematic model biases and errors of the underlying model it was trained on when producing forecasts. Changes to the current model will reduce the accuracy of the NN forecasts. An operational forecasting organization will therefore benefit from a NN system if it is feasible to carry out historical reruns of the model every time it changes to build a data sample containing several years of data.

REFERENCES

1. Applequist S., G. E. Gahrs, R. L. Pfeffer, and X.-F. Niu, 2002: Comparison of methodologies for probabilistic quantitative precipitation forecasting. *Wea. Forecasting,* **17,** 783–799.
2. Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment. *J. Atmos. Sci.,* **31,** 674–701.
3. Atger F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev,* **127,** 1941–1953.
4. Banitz E., 2001: Evaluation of short-term weather forecasts in South Africa. *Water SA,* **27,** 489-498.
5. Bernardet, L. R., 2002: Verification of high-resolution precipitation forecasts for the 1996 Atlanta Olympic Games. *Natl. Weather Dig.,* **26,** 19-36.
6. Betts A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc,* **112,** 677–692.
7. Betts A. K., and M. J. Miller, 1993: The Betts–Miller scheme. The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr., No. 46, *Amer. Meteor. Soc.,* 159–164.
8. Bishop, C.M. (1995), Neural Networks for Pattern Recognition, Oxford: Oxford University Press.
9. Bland, M., 2000: *An introduction to medical statistics,* 3rd ed. Oxford: Oxford University Press.
10. Buizza R., 2001: Accuracy and Potential Economic Value of Categorical and Probabilistic Forecasts of Discrete Events. *Mon. Wea. Rev ,* **129,** 2329-2345.
11. Colle B. A., M. Jones and J. S. Tonque, 2005: Evaluation of a mesoscale short-range ensemble forecast system in the Northeast United States. *AMS Conference on Weather Analysis and Forecasting,* **21.**
12. Coulibaly, P., Y.B. Dibike, and F. Anctil, 2004: Downscaling Precipitation and Temperature with Temporal Neural Networks. *J. Hydrometeorol,* **6(4),** 483-496.
13. Croxton, F. E., and D. J. Crowden, 1955: *Applied General Statistics.* Prentice-Hall, 843 pp.
14. Doswell C. A., 2004: Weather Forecasting by Humans—Heuristics and Decision Making. *Wea. Forecasting,* **19,** 1115–1126.
15. Doswell, C. A., III, and H. E. Brooks, 1998: Budgetcutting and the value of weather services. *Wea. Forecasting,* **13,** 206–212.
16. Efron, B., 1982: *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: SIAM.
17. Efron, B. and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap,* London: Chapman & Hall.
18. Fritsch, J. M. and R. R. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bul l. Amer. Meteor. Soc.,* **85,** 955–965.
19. Gallus, W. A., Jr., and M. Segal, 2004: Does increased predicted warm season rainfall indicate enhanced likelihood of rain occurrence? *Wea. Forecasting,* **19,** 1127-1135.

20. Giorgi, F., and M.R. Marinucci, 1996: An investigation of the sensitivity of simulated precipitation to the model resolution and its implications for climate studies. *Mon. Wea. Rev.,* **124,** 148-166.
21. Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203-1211.
22. Golding, B.W., 2000: Quantitative precipitation forecasting in the UK. *J. Hydrology,* **239,** 286-305.
23. Goutte, C., 1997: Note on free lunches and cross-validation. *Neural Computation,* **9,** 1211-1215.
24. Grams J. S., W. A. Gallus, S. E. Koch, L. S. Wharton, A. Loughe, and E. E. Ebert, 2005: The use of a modified Ebert-McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. *AMS Conference on Weather Analysis and Forecasting / Numerical Weather Prediction,* **21.**
25. Grell, G. A. (1993), Prognostic evaluation of assumptions used by cumulus parameterizations, *Mon. Wea. Rev.*, **121,** 764–787.
26. Grell, G. A., J. Dudhia, and D. R. Stauffer 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5), *NCAR/TN-398+STR*, 122 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
27. Hall T., Brooks H. E., Doswell C. A. 1999: Precipitation Forecasting Using a Neural Network. *Wea. Forecasting,* **14,** 338–345
28. Hamill T. M., Whitaker J. S., Wei X., 2004: Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. *Mon. Wea. Rev.,* **132,** 1434–1447
29. Hanssen A. J. and W. J. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Koninklijk Nederlands Meteorologist Institua Meded. Verhand,* 81-2-15.
30. Haykin, S., 1994: *Neural Networks.* Macmillan College Publishing Company, Inc, New York.
31. Hewitson, B.C., 2004: Empirical downscaling: assumptions, caveats and stationarity. *Paper presented at the Ninth International Meeting on Statistical Climatology,* Cape Town, South Africa, 24-28 May 2004.
32. Hewitson, B.C., 1999: Deriving regional precipitation scenarios from General Circulation Models. Water Research Commission Report 751/1/99, Pretoria, South Africa, 40pp.
33. Hjorth, J.S.U., 1994: *Computer Intensive Statistical Methods Validation, Model Selection, and Bootstrap,* London: Chapman & Hall.
34. Janjic, Z.I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927-945.
35. Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.,* **47,** 2784-2802
36. Kain, J.S., and J.M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain- Fritsch scheme. The representation of cumulus convection in numerical models. *Meteor. Monogr., No. 24, Amer. Meteor. Soc.,* 165-170.
37. Klein W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter, *J. Atmos. Sci.,* **16,** 672-682.
38. Krishnamurti, T. N., S. Low-Nam and R. Pash, 1983: Cumulus parameterization and rainfall rates II. *Mon. Wea. Rev.*, **111,** 816-828.

39. Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307
40. Marzban, C., 1998: Scalar measures of performance in rare-event situations. *Wea. Forecasting*, **13**, 753-763.
41. Marzban, C. 2000: A neural network for tornado diagnosis. *Neural Computing and Applications,* **9 (2),** 133-141.
42. Marzban, C., 2003: A neural network for post-processing model output: ARPS. *Mon. Wea. Rev.,* **131(6),** 1103-1111.
43. Marzban, C., S. Sandgathe, E. Kalnay, 2005: MOS, Perfect Prog, and Reanalysis Data. *Mon. Wea. Rev,* **134,** 657-663.
44. Marzban, C., G. J. Stumpf, 1996: A neural network for tornado prediction ..., *J. Appl. Meteorol.,* **35,** 617.
45. Marzban, C., and G. J. Stumpf, 1998: A neural network for damaging wind prediction. *Wea. Forecasting*, **13**, 151-163.
46. Marzban, C., and A. Witt, 2001: A Bayesian Neural Network for Hail Size Prediction. *Wea. Forecasting,* **16,** 600-610.
47. Marzban, C., E. D. Mitchell, G. Stumpf, 1999: The notion of ``best predictors:" An application to tornado prediction. *Wea. Forecasting,* **14**, 1007-1016.
48. Marzban, C., S. Sandgathe, E. Kalnay, 2005: MOS, Perfect Prog, and Reanalysis Data. *Mon. Wea. Rev.,* **134(2),** 657-663
49. Mason, I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine,* **30,** 291-303.
50. Mass, C.F., D. Ovens, K Westrick, and B.A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts, *Bull. Amer. Meteorol. Soc.,* **83,** 407-430.
51. McCann, D.W., 1992: A Neural Network Short-Term Forecast of Significant Thunderstorms, *Wea. Forecasting*, **7**, 525-534.
52. Mesinger, F., Janjic, Z. I., Nickovic, S., Gavrilov, D., and Deaven, D. G., 1988: The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev.,* **116,** 1493-1518.
53. Mohanty U. C., N. Ravi, and O. P. Madan, 2001: Forecasting precipitation over Delhi during the southwest monsoon season. *Meteor. Appl*, **8**, 11–21.
54. Moller, M. F., 1993: A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, **6,** 525-533.
55. Murphy, A.H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595-600.
56. Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293.
57. Murphy, A. H., 1996: The Finley affair: A signal event in forecast verification. *Wea. Forecasting,* **11,** 3–20.
58. Narasimhan R., J. Keller, and G. Subramaniam, 2000: Ozone modeling using neural networks. *J. Appl. Meteor.*, **39**, 291–296.
59. Neal, R. M., 1996: *Bayesian Learining for Neural Networks.* New York: Springer-Verlag, 204p.
60. Nigrin, A., 1993: *Neural Networks for Pattern Recognition.* The MIT Press, Cambridge, MA.

61. Olson, D. A., N. W. Junker, and B. Korty, 1995: Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Wea. Forecasting,* **10,** 498–511.
62. Olsson J., C. B. Uvo, K. Jinno, A. Kawamura, K. Nishiyama, N. Koreeda, T. Nakashima, and O. Morita, 2004: Neural Networks of Rainfall Forecasting by Atmospheric Downscaling. *J. Hydrol. Eng.,* **9,** 1-12.
63. Orr, G.B., and Mueller, K.-R., eds. 1998: *Neural Networks: Tricks of the Trade,* Berlin: Springer, ISBN 3-540-65311-2.
64. Pankiewicz, G. S., 1995: Pattern recognition techniques for the identification of cloud and cloud systems. *Meteor. Appl.,* **2,** 257–271.
Pankiewicz, G. S., 1997: Neural Network Classification of Convective Airmasses for a Flood Forecasting System. *Int. J. Rem. Sen.*, **18(4),** 887-898.
Pasini A., and S. Potesta', 1995: Short-Range Visibility Forecast by Means of Neural-Network Modelling: a Case-Study, *Il Nuovo Cimento 18C,* 505-516.
65. Pliske R., D. Klinger, R. Hutton, B. Crandall, B. Knight, and G. Klein, 1997: Understanding skilled weather forecasting: Implications for training and the design of forecasting tools. Contractor Rep. AL/HR-CR-1997-003, Material Command, Armstrong Laboratory, U.S. Air Force, 122 pp.
66. Poolman, E., 2004: Probabilities and precipitation terms used in weather forecasts in the South African Weather Service. Internal SAWS report.
67. Qian, Y., and L. R. Leung, 2005: Downscaling Extended Weather Forecasts for Hydrologic Prediction. *Bulletin of the American Meteorological Society,* **86(3),** 332-333.
68. Richardson, D.S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Royal Meteorol. Soc.*, **126**, 649-667.
69. Ripley, B.D. (1996) Pattern Recognition and Neural Networks, Cambridge: Cambridge University Press.
70. Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, 1996: Changes to the operational "early" Eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting*, **11**, 391-413.
71. D.E. Rumelhart, D. E., and J.L. McClelland, 1986: *Parallel Distributed Processing,* volume 1. MIT Press.
72. Sarle, W. S., 1995: Stopped Training and Other Remedies for Overfitting. *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics,* 352-360.
73. Sarle, W.S., ed., 1997: Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: ftp://ftp.sas.com/pub/neural/FAQ.html.
74. Schoof J. T., and S. C. Pryor, 2001: Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, International *J. Clim.,* **21,** 773-790.
75. Silverman, D. and J. Dracup, 2000: Artificial Neural Networks and Long Range Precipitation Prediction in California, *J. Appl. Meteorol.,* **31**, 57-66.
76. Shukla, J., 1985: Predictability, Advances in Geophysics, in: Atmos. And Ocean modelling. Part B: Weather Dynamics, edited by: Manabe, S., Academic Press: 87-123.
77. Snellman, L., 1977: Operational forecasting using automated guidance. *Bull. Amer. Meteor. Soc.,* **58,** 1036–1044.

78. Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. *WWW Tech. Rep. 8,* WMO/TD 358, 114 pp.
79. Stein, J., E. Richard, J.P. Lafore, J.P. Pinty, N. Asencio, and S. Cosma, 2000: High-resolution non-hydrostatic simulations of flash-flood episodes with grid nesting and ice-phase parameterization. *Meteorol. Atmos. Phys.,* **72,** 203-221.
80. Swets, J. A. and R. M. Pickett, 1982: *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory.* New York: Academic Press.
81. Szucko, J. J., and B. Kleinmuntz, 1981: Statistical versus clinical lie detection. *American Psychologist,* **36,** 488-496.
82. Taljaard, J. J., 1986: Change of rainfall distribution and circulation patterns over Southern Africa in summer. *J. Climatol.,* **6,** 579-592.
83. Tennant, W. J., 2006: Personal communication.
84. Thornes, J.E. and D.B. Stephenson, 2001: How to judge the quality and value of weather forecast products, *Meteorol. Appl.,* **8,** 307-314.
85. Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117(8),** 1779-1800.
86. Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breading method. *Mon. Wea. Rev.,* **125,** 3297-3319
87. Tsonis, A.A., 2001: The impact of nonlinear dynamics in the atmosphere sciences, *Int. J. Bifurcation and Chaos*, **11,** 881-902.
88. Tsonis, A.A., and J.B. Elsner, 1989: Chaos, strange attractors and weather, *Bull. Amer. Meteorol. Soc.,* **70,** 14-23.
89. Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, NY. 467 pp.
90. Wilks, D.S., 2001: A skill score based on economic value for probability forecasts. *Meteorol. Appl*., **8**, 209-219.
91. Yuan, H., S. L. Mullen, X. Gao, S. Sorooshian, J. Du, , and H-M. H. Juang, 2005: Verification of Probabilistic Quantitative Precipitation Forecasts over the Southwest United States during Winter 2002/03 by the RSM Ensemble System. *Mon. Wea. Rev,* **133(1),** 279–294.
92. Zhang, H. and Frederiksen, C.S. 2003: Local and non-local impacts of soil moisture initialization on AGCM seasonal forecasts: A model sensitivity study. *J Climate,* **16,** 2117-37.
93. Zurada, J. M., 1992: *An Introduction to Artificial Neural Systems.* St. Paul: West Publishing Company.