UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A multi-model ensemble system for short-range weather prediction in South Africa

by

Stephanie Landman

Submitted in partial fulfilment of the requirements
for the degree of

MASTER OF SCIENCE

in the
Faculty of Natural and Agricultural Sciences
University of Pretoria

January 2012

A multi-model ensemble system for short-range weather prediction in South Africa

Stephanie Landman

Supervisor: Ms. L. L. Dyson
Co-supervisor: Dr. F.A. Engelbrecht
Department: Department of Geography, Geoinformatics and Meteorology
Faculty: Faculty of Natural and Agricultural Sciences
University: University of Pretoria
Degree: Master of Science

ABSTRACT

Predicting the location and timing of rainfall events has important social and economic impacts. It is also important to have the ability to predict the amount of rainfall accurately. In operational centres forecasters use deterministic model output data as guidance for a subjective probabilistic rainfall forecast. The aim of this research is to determine the skill in an objective multi-model, multi-institute probabilistic forecast system. This was done by obtaining the rainfall forecast of two high-resolution regional models operational in South Africa. The first model is the Unified Model (UM) which is operational at the South African Weather Service. The UM contributed three members which differ in physics, data assimilation techniques and horisontal resolution. The second model is the Conformal-Cubic Atmospheric Model (CCAM) which is operational at the Council for Scientific and Industrial Research which in turn contributed two members to the ensemble system differing in horisontal resolution. A single-model ensemble was constructed for the UM and CCAM models respectively with each of the individual members having equal weights. The UM and CCAM single-model ensemble prediction models have been used in turn to construct a multi-model ensemble prediction system, using simple un-weighted averaging. The multi-model system was used to predict the 24-hour rainfall totals for three austral summer half-year seasons of 2006/07 to 2008/09. The forecast of this system was rigorously tested using observed rainfall data for the same period. From the multi-model system it has been found that the probabilistic forecast has skill in predicting rainfall. The multi-model system proved to have skill and shows discrimination between events and non-events. This study has shown that it is possible to make an objective probabilistic rainfall forecast by constructing a multi-model, multi-institute system with high resolution regional models currently operational in South Africa. Thus, probabilistic rainfall forecasts with usable skill can be made with the use of a multi-model short-range ensemble prediction system over the South African domain. Such a system is not currently operational in South Africa.

I declare that the thesis that I hereby submit for the degree Master of Science at the University of Pretoria is my own work and has not previously been submitted by me for degree purposes at any other university or institution.

SIGNATURE DATE

University of Pretoria, 2012

PREFACE

Precipitation is of high relevance to users in South Africa, but it is also highly variable. In short-range weather forecasting (from 12-hours up to 3 days ahead) predicting the location of a precipitation event in general has a greater error than the prediction of the pattern and amount of precipitation. It is therefore essential to have the ability to predict the timing and location of rainfall as well as the spatial extent of rain bearing synoptic events over a region, as skillful as possible.

Weather forecasting is dependent on the accuracy of initial conditions due to the chaotic and non-periodic characteristics of the atmosphere. An inherent characteristic of deterministic forecasts is that the future state of the atmosphere is completely conditional on the present state of the system and the atmospheric evolution is governed by deterministic equations. An alternative approach is to address the limits in deterministic forecasting by means of ensemble forecasting.

Skillful forecasts may aid to timely issue of floods and thereby reducing the risk of property damage. For these reasons it is important for the forecasts to be as skillful as possible in predicting the timing and location of rain-bearing systems. Short-range ensemble prediction is a relatively unexplored field of numerical weather prediction and offers an opportunity to improve upon the skill of forecasting the probability of precipitation events. A probabilistic precipitation forecast reduces the occurrence of missed events within a forecast system and is generally more successful in identifying the area over which precipitation occurs.

In this study a short-range multi-model ensemble system is constructed using two regional numerical weather prediction models (contributing three and two members respectively) which will aim to produce an objective probability precipitation forecast for South Africa.

The hypotheses that will be tested are:

a. that the multi-model ensemble has skill in predicting 24-hour probabilistic precipitation for South Africa;

b. that the multi-model ensemble can produce forecasts with greater skill levels to that of the target single-model numerical weather prediction regional model;

c. That the multi-model outperforms a coarse-resolution model of which the output is freely available on the internet

The necessary steps to be taken in order to test these hypotheses are to:

a. calculate the 24-hour precipitation totals for each of the five members of the multi-model ensemble;

b. rescale each of the five members to a resolution consistent with the analysed rainfall data;

c. calculate the precipitation totals for each of the two single-model ensembles as well as for the multi-model ensemble;

d. calculate the probability forecast for four daily threshold values;

e. rescale the course-resolution single-model ensemble forecast to the same resolution as the single- and multi-models.

f. verify the four ensembles against observational data and compare the multi-model ensemble against the skill of the coarse-resolution ensemble as well as the two single-model ensemble systems.

This dissertation consists of six chapters. Chapter 1 describes the summer rainfall characteristics and dynamics over South Africa, numerical predictability of rainfall on the short-range timescale, NWP models' weakness in predicting rainfall and ensemble and multi-model ensemble prediction methods. The data and methodology used in this study for the design, as well as the methods used for the verification of the multi-model ensemble are detailed in Chapter 2. The multi-model verification results are explained in Chapter 3 and the verification results of the low-resolution single-model ensemble system in Chapter 4. The comparison in skill between the multi-model and the low resolution model is also discussed in this chapter. The results are summarized and conclusions are made in Chapter 5.

ACKNOWLEDGEMENTS

I would like to give my thanks and appreciation to:

- Ms. L. L. Dyson (University of Pretoria) for her assistance, guidance and support during the course of this study.
- Dr. F.A. Engelbrecht (CSIR) for his guidance and advice throughout this study. Also for most needed assistance in some of the programming involved. Most importantly I thank you for supplying the CCAM model data required to conduct this study.
- Mrs. Christien Engelbrecht for the quality control and set-up of the observation data fields.
- Mrs. Estelle Marx (SAWS) for her patience and willingness to assists in any programming or data queries I had.
- Ms. Karin Marais and Anastasia Demertzis, the SAWS librarians who always lend a helping hand with the acquisition of articles or books when needed.
- The South African Weather Service who funded this research study. Thank you for the support.
- Dr. Simon Mason (IRI) for his guidance with verification issues.
- My Mother and Stepfather for their selfless support during this period.
- My husband for his patience, guidance and support.

“He replied, “When evening comes, you say, ‘It will be fair weather, for the sky is red’, and in the morning, ‘Today it will be stormy, for the sky is red and overcast.’ You know how to interpret the appearance of the sky, but you cannot interpret the signs of the times.””

Matthew 16:2-3

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Precipitation is perhaps the most relevant to users of meteorological data in South Africa, but it is also the most variable spatially and temporally. The prediction of rainfall extremes has high socioeconomic impacts (Fauchereau *et al*, 2008; Friederichs and Hense, 2008). It is therefore essential to have the ability to predict the timing and location of rainfall events, and heavy rainfall in particular, as accurately as possible (Theis *et al*., 2005). In short-range weather forecasting (12- to 72-hours) predicting the location of a precipitation event in general has a greater error than the prediction of the pattern and amount of precipitation (Ebert and McBride, 2000). Theis *et al* (2005) noted that with the decrease in the spatial scale of the event being forecast, e.g. thunderstorm development, the uncertainties in regional modelling increases. The large spatial and temporal variability in rainfall together with some internal model restrictions also contribute to the uncertainties and low skill associated with rainfall predictions (Ebert, 2001; Theis *et al*, 2005; Roy Bhowmik and Durai, 2010).

1.1 RAINFALL OVER SOUTH AFRICA

South Africa is considered to be mostly a semi-arid area, with a hot and dry climate. The leading cause of South Africa to be a semi-arid region is due to its geographic location within the subtropical high pressure belt (Tyson and Preston-Whyte, 1993). The average annual rainfall of South Africa is less than 500 mm, with only a relatively small part of the country receiving more than 500 mm per annum. In Figure 1.1 it is seen that the eastern parts of the country experience annual rainfall in excess of 800 mm in some regions. In contrast, the western parts experience desert like conditions with less than 200 mm per year. The maximum annual rainfall of about 1200 mm occurs over the eastern escarpment of the country and the minimum rain over the extreme north-western parts (less than 100 mm).

Rainfall over South Africa is characterized by three rainfall regions, namely a summer rainfall region (that receives the bulk of its rainfall from October to March), the winter rainfall region of the southwestern Cape (that receives the bulk of its rainfall from May to August) and all-year rainfall along the Cape south coast (Figure 1.2). The eastern

parts of the country are mostly associated with summer rainfall; with most rainfall events caused by heat induced convective storms (Tyson and Preston-Whyte, 1993; Fauchereau *et al*, 2008). Along the eastern escarpment of South Africa, the orographic lift of moist air advected from the east contributes to the relatively high annual rainfall of the escarpment areas and the east coast. This advection of moist air is caused by the ridging high pressure system over eastern South Africa (van Heerden and Hurry, 1987). Todd and Washington (1998) and Fauchereau *et al* (2008) confirmed the earlier analysis of Harrison (1984) that tropical temperate troughs (TTT's) are synoptic systems contributing significantly towards daily rainfall variability and annual rainfall totals over the areas that receive summer rainfall. TTT's result in linkages forming between the colder mid-latitude regimes and the warm, moist air from the tropics. TTT's result in the formation of synoptic-scale northwest to southeast orientated cloud bands over southern Africa, with 30% to 60% of the austral summer half-year rainfall accounted for by these systems (Fauchereau *et al*., 2008). Other rain producing systems of summer rainfall over the region are upper-air troughs of the westerlies, cut-off lows, tropical lows and ridging highs (Taljaard, 1996).

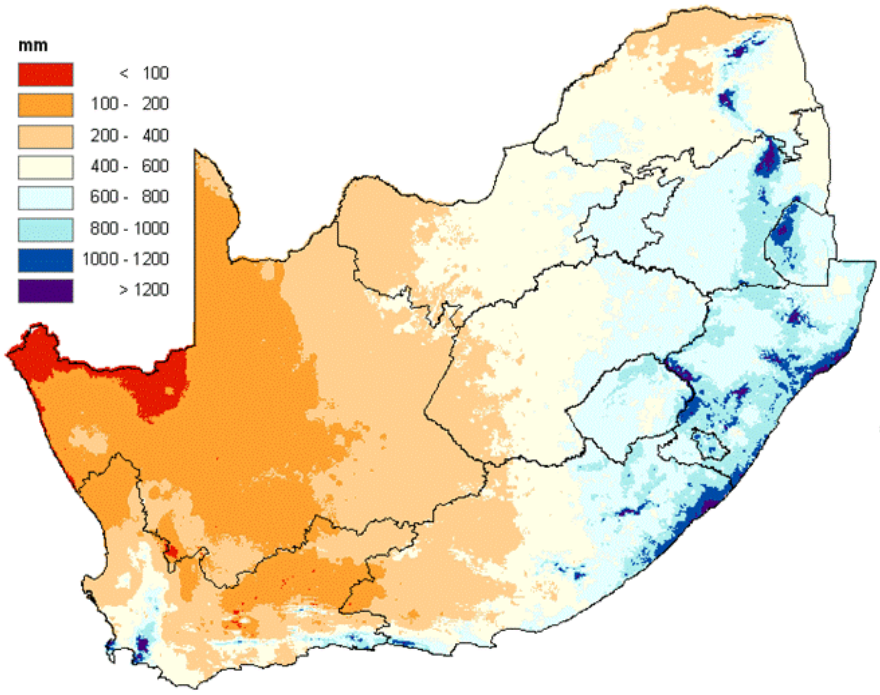


FIGURE 1.1: The mean annual rainfall over South Africa (adapted from Schulze *et al*, 2008)

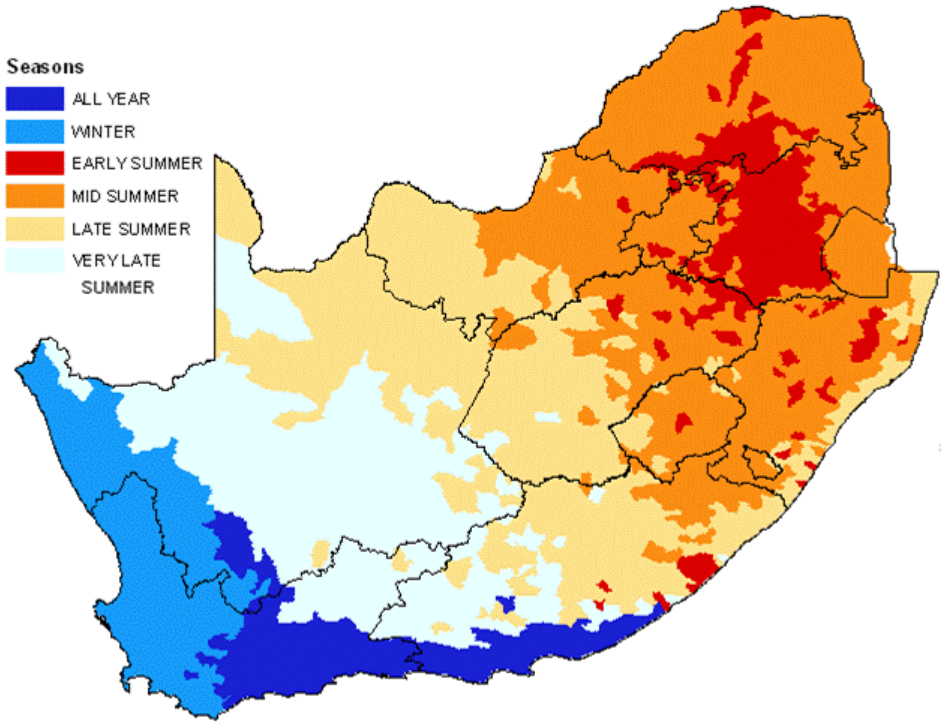


FIGURE 1.2: The seasonality of rainfall over South Africa per quaternary catchment (adapted from Schulze *et al*, 2008)

The Western Cape Province receives mostly winter rainfall (Schulze and Maharaj, 2007). The common cause of the winter rainfall is mostly due to frontal rain, caused by the mid-latitude low pressure systems moving over the sub-continent during the winter months (van Heerden and Hurry, 1987, Tyson and Preston-Whyte, 1993). The all-year rainfall region is situated over the southern and south-eastern coastal areas (Schulze and Maharaj, 2007).

Figure 1.3 illustrates this by indicating the rainfall seasonality per quaternary catchment for each of the months in the summer half-year (Schulze *et al*, 2008). The north-eastern and central interior regions receive rainfall from early summer (October) through to late summer (March) with the maximum number of rain-days during January. Taljaard (1996) indicated that the 50 mm/day isohyet gradually moves to the west beginning October until March/April after which the isohyets rapidly moves back to the east for the beginning of the winter season (May). This transition of rainfall from east to west is also visible in Figure 1.3 where the number of rain-days increases to the west resulting in only a small region over the far western parts of the country having less than 1 day per month with rainfall.

In this research the focus is on the austral summer half-year rainfall (October to March) when the eastern half of the country receives most of its rainfall.

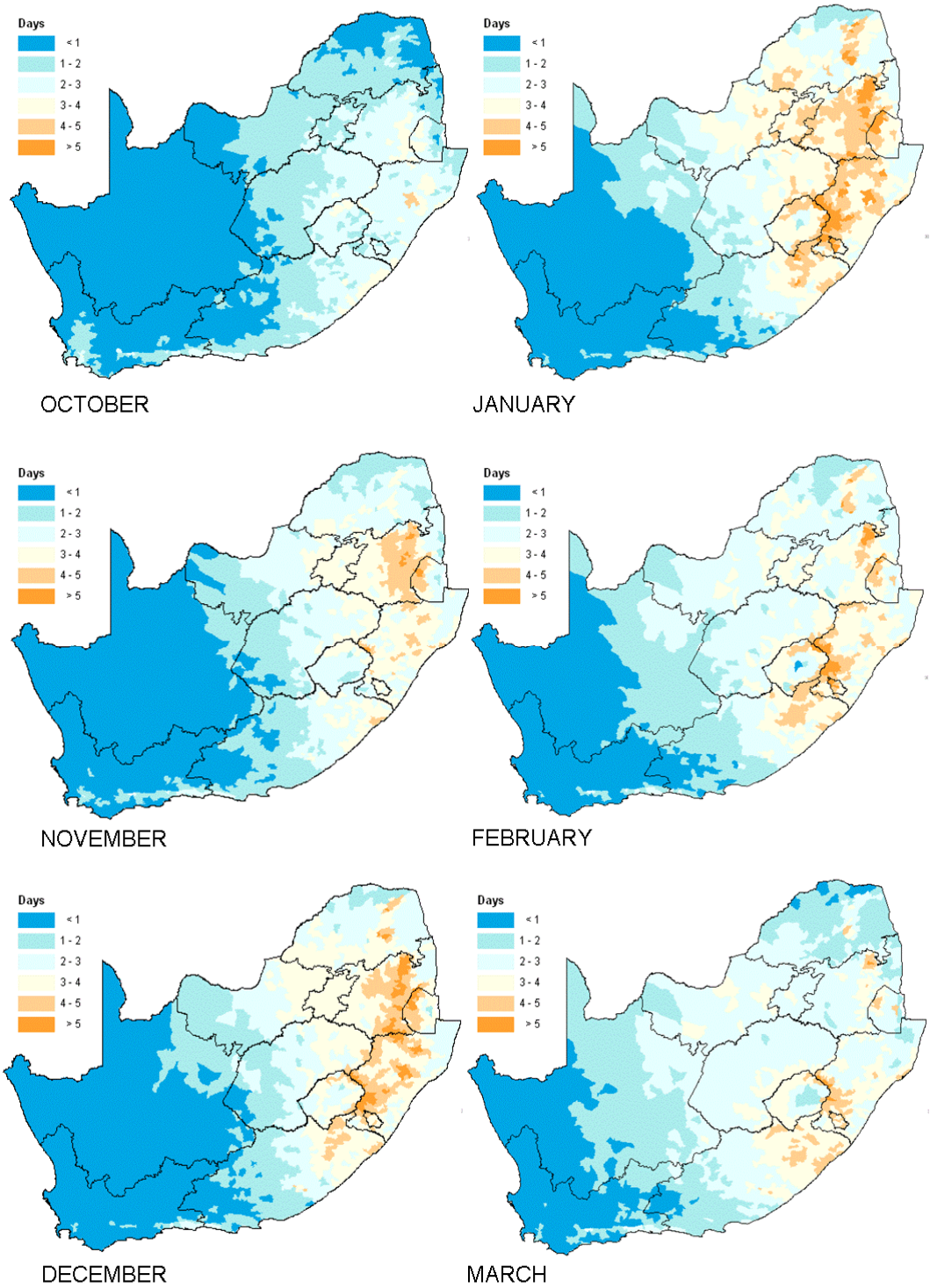


FIGURE 1.3: Mean number of days per month with rainfall greater than 10 mm/day (adapted from Schulze *et al*, 2008)

1.2 PREDICTABILITY OF RAINFALL OVER THE SHORT-RANGE TIMESCALE

Most operational centers rely on limited-area numerical weather prediction in order to generate reliable and accurate weather forecasts (Stensrud *et al*., 1999; Toth *et al*., 2001). Traditionally, these forecasts have been presented in a deterministic manner. An inherent characteristic of deterministic forecasts is that the future state of the atmosphere is completely conditional on the present state of the system, and the atmospheric evolution is governed by deterministic equations (Lewis, 2005). Accurate short-range numerical forecasting is therefore dependent on accurately describing the state of the atmosphere at the start of the forecast; this state is known as the "initial conditions" (Kalnay, 2003). The reason for this dependency on accurate initial conditions stems from the chaotic and non-periodic characteristics of the atmosphere (Lorenz, 1963). Due to this dependency, forecasts that are initialized with only slightly different initial states diverge increasingly as a function of model integration time. However, atmospheric observations from weather stations are scarce and widely distributed in three-dimensional space, resulting in an incorrect representation of the atmospheric state and erroneous initial conditions (Du and Mullen, 1997). Although remotely sensed data are being used increasingly to describe the atmospheric initial state provided to atmospheric models, the quality and number of variables available from this data resource restricts its use for model initialization. Deterministic or best-guess forecasts are therefore not considered reliable, at least at model integration times of a few days or longer into the future, due to the uncertainties that exist in the initial conditions as well as the internal error (physics and dynamics) of the numerical models (Lorenz, 1963; Ebert, 2001; Stensrud *et al*., 2005; Theis *et al*, 2005). Therefore, even though a single numerical model can produce skillful deterministic forecasts a number of days into the future, the confidence in deterministic forecasts decreases as a function of integration time (e.g. Santos-Muños *et al*, 2010). Operational deterministic forecasts are historically performed at high spatial resolution over limited areas, and are usually constructed from nested limited-area models (Roy Bhowmik and Durai, 2010). The advantage of using regional models for issuing short-range forecasts is that the forecasts are superior in quality and value to end-users on a higher resolution as oppose to coarser resolution global model forecasts (Wandishin *et al.,* 2001).

The issuing of probability forecasts at the short-range time-scale may address to some extent the uncertainties associated with describing the initial state of the atmosphere, and will additionally allow for more flexibility in the usage of the forecast (Theis *et al*, 2005).

Indeed, most National Meteorological Services (NMS) issue precipitation in terms of probabilities whereby the user can get additional information regarding the uncertainty pertaining to the specific forecast (Staël von Holstein, 1971). Forecasters are able to issue these probability forecasts by means of a subjective "ensemble forecast" due to the fact that they use multiple information sources and models to issue forecasts (Ebert, 2001). Forecasters have long been aware of the fact that a variety in models can produce a variety in the outcome of the predicted weather (Ebert, 2001). These probability forecasts issued by forecasters remain subjective because they are still based on the forecaster's own insights and experience (Staël von Holstein, 1971). The forecaster's judgment is applied when combining the objective information systems to his/her disposal.

Due to the increase in computer power, more NMS's are today able to run regional models in-house. These models are primarily used for predicting weather patterns on smaller scales than those simulated by global circulation models (de Elia *et al*, 2001). These regional models run at high horisontal resolution and over a limited domain and receive lateral boundary and initial conditions from coarser global circulation models (de Elia *et al*, 2001). This increase in model resolution implies an improvement in the representation of model physical processes as well as dynamics, and therefore enhances the simulation quality. However, the predictability of smaller spatial scale features is less than those of synoptic scale systems. In order to correct coarse as well as fine resolution numerical forecast output to best capture small-scale events, Friededrichs and Hense (2008) noted that post-processing techniques are used. The two most widely used techniques are the perfect prognosis method and model output statistic (MOS).

In order for a model to forecast precipitation accurately, thorough knowledge and data of atmospheric moisture and vertical motion fields must exist (Ebert, 2001). The forecast and simulation of rainfall is particularly challenging over those regions of the world where rainfall is primarily convective such as over the summer rainfall area of South Africa. Roy Bhowmik and Durai (2010) as well as Friederichs and Hense (2008) also noted that even with the recent improvements in numerical prediction, the

skill in predicting rainfall on the small scale is inadequate. Even at the largest international forecasting centers, operational weather prediction is limited by computational restrictions to spatial resolutions where convection can at best be partially resolved (Engelbrecht *et al.*, 2007). This necessitates the use of convective parameterization schemes within numerical models, in order to simulate the effects of convection on atmospheric stability and the vertical distribution of momentum, heat and moisture. That is, the convective parameterization included in model formulations need to represent the intricate processes of cloud dynamics and microphysics (Engelbrecht *et al*, 2007). This statistical treatment of convection is regarded to be one of the greatest sources of uncertainty in both weather prediction and climate simulation (Stensrud *et al.,* 2005; Tadross *et al.*, 2006; Engelbrecht *et al.*, 2007). Stensrud *et al*. (2005) also points out that uncertainty and assumptions made on the model physics and parameterization may lead to incorrect spatial and temporal placement of precipitation occurrence within forecasts. This conclusion was also substantiated by a number of authors who found that changes in parameterization of convection lead to different simulations of convective rainfall (Wang and Seaman, 1997; Tadross *et al.*, 2006). Tropical- and southern Africa are indeed two regions that are prone to convection, and the uncertainties associated with cumulus parameterization within models are highly relevant to these regions (Tadross *et al.*, 2006).

Parameterization schemes, such as those needed to simulate convective precipitation, are a limiting factor in numerical weather prediction. These schemes are generally simplified for computational purposes (Bowler *et al*, 2008a) and therefore do not simulate a small-scale weather phenomenon as realistically as it otherwise could. Parameterization schemes are also limited by the understanding of the physical processes to be parameterized. This simplification of the physical schemes contributes to the internal model error and hence also forecast uncertainty. Richardson (2001) however stated that the error in initial conditions has a greater impact on weather prediction on the short-range than the error in the physical schemes. This impact of different input data and initial conditions were also classified by He *et al* (2009) with a multi-model ensemble system over China.

In the 1960's, Edward Lorenz showed that the chaotic nature of the atmosphere pose a limit on the predictability of the weather (Kalnay, 2003; Bishop *et al*, 2009). This limit was said to be about fourteen days, and stems from the uncertainty in describing the initial state of the atmosphere. But Bishop *et al* (2009) also stated that even if we

could accurately solve primitive equations and perfectly observe the atmosphere, the limit of predictability of flows still remains. Errors in the initial conditions amplify with increase lead time, leading to differences in forecast as well as decreasing user's confidence (Santos-Muñoz *et al*, 2010).

Owing to the dependency of numerical weather prediction models on initial conditions and the non-linear evolution of the atmosphere, ensemble forecasting, where a probabilistic forecast of the future state of the atmosphere is issued rather than a single deterministic forecast, was proposed by Leith in 1974. The ensemble of forecast outputs from single or multiple numerical weather prediction models provide detail of the forecast regarding the confidence, possible errors and probability outcomes (Bakhshaii and Stull, 2009). Moreover, an ensemble to some extent describes the uncertainties pertained in single-model forecasts (Zongjian, 2008; Kalnay, 2003). However, Clark *et al* (2008) noted that with all the advances in rainfall prediction skill-improvement techniques, the best way to configure a numerical forecast system is still being investigated. Therefore, running models with different configurations provide an additional way of constructing ensembles of forecasts, in addition to running a model with one physical configuration but initializing from slightly different initial states.

1.3 ENSEMBLE AND MULTI-MODEL ENSEMBLE PREDICTION

Ensemble prediction systems (EPS) represent a stochastic approach which couples probability with determinism (Lewis, 2005), and which has the specific aim of predicting the probability of future weather events to occur; in turn addressing the uncertainty of a deterministic forecast (Stensrud *et al*., 1999). Theis *et al* (2005) concluded that precipitation forecasts should be addressed in a probabilistic manner in order to account for the chaotic atmosphere as well as the afore-mentioned uncertainties in the models and observations. An important goal of ensemble prediction is to provide an estimation in the reliability of the forecast being made (Kalnay, 2003; Grimit and Mass, 2005). Hoffman and Kalnay (1982) did however also mention that the advantages to be gained from constructing an ensemble may not be that great when comparing to an already optimally designed deterministic forecast. Atger (2001) also found that a high resolution single-model forecast can under certain conditions perform better than a lower resolution ensemble.

Ensemble forecasts may be constructed in various ways (e.g. Kalnay, 2003). The traditional approach is to perform multiple model runs using the same model by initializing each run from differently constructed initial conditions. This can be achieved for example by introducing small perturbations to the control analysis (Monte Carlo) or by using different analysis times to initialize the forecasts (Lagged Average Forecasting (Hoffman and Kalnay, 1982)). On average the forecasts obtained from these perturbed initial conditions will be less skillful than the control forecast (Kalnay, 2003; Grimit and Mass, 2002). Single-model ensemble systems effectively prevent the description of the forecast uncertainty associated with model error, and this might lead to underestimation of the forecast error (Clark *et al*, 2008). Hence an alternative ensemble construction is to create forecasts initialized in a similar way as mentioned before but obtained with different physical settings (such as convection and cloud scheme settings). In this manner, credible assimilations of model errors can be addressed by changing any known model shortcomings such as sub-grid parameterization schemes (Clark *et al*, 2008). The resulting collection of forecasts may be used to make probability forecasts. Ebert (2001) and Eckel and Mass (2005) noted that with single-model EPS the main uncertainty is with the initial conditions and any systematic bias within the model itself will inherently be present within the EPS itself. Hamill and Coucci (1997) wrote that in such a case further calibration may be required.

Another way of constructing ensemble forecasts is to combine forecasts of different atmospheric models that are valid for the same period. This is referred to as multi-model ensemble prediction (Kalnay, 2003). A multi-model ensemble forecast has the advantage above single-model ensembles in that the different models take into account the uncertainties in initial conditions and model configurations (physics and dynamics) and is therefore also less likely influenced by systematic errors (Ebert, 2001; Zongjian *et al*, 2008; Bowler *et al*, 2008b). Bright and Mullen (2002) also indicated that an ensemble forecast on the short-range timescale might be able to signal the possible occurrence of high-impact events. Du *et al* (1997) and Atger (2001) found an improved quality in the ensemble mean rainfall forecast in terms of both amount and distribution.

Multi-model ensemble forecasting succeeds in addressing the uncertainties that exists in the systematic errors of each numerical model as well as the uncertainties within the initial conditions (Ebert, 2001). Clark *et al* (2008) noted that in addition to addressing uncertainties in initial conditions, ensemble forecasting, more specifically

multi-model systems, will also inadvertently address errors related to lateral boundary conditions. In the short-range time scale, synoptic and mesoscale features, such as surface temperature and precipitation, are less predictable due to their more chaotic features than those features on the planetary scale (Hamill and Colucci, 1997; Friederichs and Hense, 2008; Roy Bhowmik and Durai, 2010). For this reason, ensemble methodologies will improve on the uncertainty and model error that exists on this time-scale. The model uncertainty can be accounted for by running the same model with different physical parameterizations or analysis times or by using model runs from different numerical models (Bowler *et al*, 2008b, Wandishin *et al.,* 2001). The uncertainty in initial conditions can be addressed by running the same model a number of times by adding and/or subtracting perturbations from the original atmospheric state. This will result in equally likely forecast solutions (Wandishin *et al*., 2001; Kalnay, 2003). The errors and uncertainties in each individual member of the ensemble are averaged out, making the ensemble mean smoother (Bowler *et al*., 2008b; Kalnay, 2003). However, Clark *et al* (2008) noted that rainfall forecasts from ensemble members are more often than not under-dispersive; with the observed field falling outside the forecast distribution, resulting in an insufficient ensemble (Kalnay, 2003).

1.3.1 ENSEMBLE PREDICTION ON THE SHORT-RANGE TIMESCALE

Even though the operational use of short-range ensemble systems has lagged behind that of long-range forecasting or even medium-range (Eckel and Mass, 2005), there are a number of NMSs that uses short-range ensemble prediction systems operationally or quasi-operationally. For short-range ensemble forecasting the errors in predicting small-scale features grow quickly, model error has a large impact and the limited area models can inhibit ensemble dispersion (Eckel and Mass, 2005). The authors have also shown that the influence of initial conditions and model uncertainty relies greatly in the parameter and its scale to be forecast.

Most of the short-range ensemble prediction systems run by NMS's are situated in Europe or North-America. These NMSs include NCEP (USA), INM (Spain), NMI (Norway) and Met Office (UK). Of these systems, the NCEP-SREF, Spain's SREPS and the Deutscher Wetterdienst's ensemble make use of multi-model ensembles. The Bureau of Meteorology (BoM, AUS) also has an operational multi-model ensemble prediction system.

The multi-model short-range ensemble system at NCEP (since 2008) comprises of 21 members obtained from two regional models. The regional models use the output from the NCEP medium-range ensemble as boundary conditions (Stensrud *et al*, 1999) and the breeding vectors technique is applied for the creation of initial conditions (Toth and Kalnay, 1993). Different horisontal resolutions and parameterization schemes are used in the different model runs in order to address for model uncertainties (Tracton *et al*, 1998).

The ensemble mean on the short-range timescale showed similar characteristics as the ensemble mean on the medium range (Hamill and Colucci, 1997). This result proved to be true for various weather elements such as cloud cover, temperature and precipitation amount. In fact, research has shown that the ensemble mean forecast out performs the single deterministic forecast (Ebert, 2001). Ensembles produce probability forecasts with skill better than a single deterministic model run for forecasts of precipitation, and are easily calibrated to make them very reliable (Wandishin *et al.*, 2005 and Toth *et al.*, 2005). The most notable attributes of EPS are their reliability and resolution; their ability to predict events from non-events (Toth *et al.*, 2005).

Studies have shown that bias correction has to be performed on short-range EPS because only then can model diversity add good spread (Eckel and Mass, 2005). It has been proven that mesocsale models have a considerable bias and with model diversity the negative impact of the bias increases. It was also found that statistical post-processing of different model fields outperformed the technique of averaging from different models (i.e. ensemble mean) (Krishnamurti *et al.*, 2000).

The applicability of multi-model EPS on the short-range was proven by previous studies (Ebert, 2001, Arribas *et al.*, 2005). These studies showed that an ensemble outperforms each of the individual members on this timescale.

1.3.2 ENSEMBLE PREDICTION ON THE MEDIUM-RANGE TIMESCALE

The medium-range forecast timescale is defined by the World Meteorological Organization as time scale from 3 to 10 days. Ensemble prediction systems for medium-range forecasting are generally designed to focus on uncertainties due to synoptic-scale baroclinic instabilities (Bowler *et al*, 2008a). Medium-range ensemble forecasts are produced operationally at institutions such as the European Centre for

Medium-range Weather Forecasts (ECMWF), the National Centre for Environmental Prediction (NCEP) and the Canadian Meteorological Centre (NMC) who produces medium-range forecasts (Arribas *et al.*, 2005). These ensemble systems are considered to be mature since they have been in operations for more than a decade (Bowler *et al*, 2008a). The ECMWF uses the singular vector (SV) technique to grow perturbations over the early part of the forecast period whereas the NCEP implements the ensemble transform bred vector technique (Wei *et al*, 2008).

1.3.3 ENSEMBLE PREDICTION ON THE LONG-RANGE TIMESCALE

Ensemble forecasting has been operational in the long-range timescale for a number of years. Two well-known projects done in Europe, the PROVOST (Prediction of Climate Variations on Seasonal to Interannual Timescales) project and the DEMETER (Development of a European Multi-Model Ensemble system for season to inTERannual prediction) project showed that the multi-model ensemble had more skill over that of the single-model ensembles (Palmer *et al*, 2004). The same result was found over southern Africa for mid-summer seasonal rainfall (Landman and Beraki, 2010).

1.4 AIMS AND APPROACH OF RESEARCH

The main aim of this dissertation is to investigate the skill of a high-resolution short-range multi-model ensemble prediction system for South Africa. To achieve this aim the following objectives are relevant:

- To test the daily rainfall forecast performance of a number of individual short-range forecast systems
- To test the skill when the forecasts of the individual systems are combined into a multi-model system
- To compare the skill levels of the multi-model system with that of a course-resolution global model

These objectives will be addressed by:

- Constructing daily rainfall single-model ensemble systems by combining the respective ensemble members produced by models being run operationally in South Africa. Two ensembles are developed by taking into considerations the

models' different physics and horisontal resolutions. The models are also initialized differently.

- Constructing a multi-model daily rainfall ensemble system by combining the two single-model ensemble systems into a multi-model system with each single-model ensemble being weighted equally.
- Testing the skill of the forecast systems through elaborate verification procedures.
- Comparing the skill levels of the South African based multi-model system with the skill levels of course-resolution model output obtained from an international centre.

1.5 SUMMARY

South Africa is a region that displays a great deal of natural variability, with a wide variety of different weather systems that bring rainfall to the summer rainfall region. This dissertation is to investigate the accuracy and skill of a high-resolution, multi-model ensemble forecasting system in forecasting rainfall over South Africa at the short-range time scale. The investigation takes place against the background of internal model and initial conditions errors on numerical weather prediction bringing uncertainty to deterministic weather prediction, thereby necessitating the issue of probabilistic forecasts.

This chapter concluded with the aim and approach of the research of this dissertation. In the next chapter, the observational data, numerical models and methodology used in this study are described in detail

CHAPTER 2

DATA AND METHODOLOGY

In this chapter the methodology applied and the data used are described. The chapter is divided into seven sections in order to explain how the ensemble was obtained and the results verified. The first two sections outline the period and area of interest, followed by the description of the observational data and the construction of the gridded data set (sections three and four). Sections five and six describe the numerical weather prediction model data used, as well as the design of the different ensemble forecasts. The final section will provide detail on the statistical scores used to verify the ensemble prediction system.

2.1 PERIOD AND AREA OF INTEREST

2.1.1 PERIOD OF INTEREST

In this study, three austral summer half-years were selected for the purpose of performing hindcasts; but only days where all of the ensemble members were available were used. These are the months of October through to March of the years 2006/07, 2007/08 and 2008/09. For these periods, 24-hour rainfall totals were calculated from the numerical weather prediction data, as well as from observations stemming from automatic weather stations. These totals were accumulated over the 24-hour periods from 06:00 UTC on a given day, to 06:00 UTC on the next day, in correspondence to the time of observation of the numerous manual weather stations from the South African Weather Service (SAWS). Therefore, the corresponding forecast hours of the numerical weather prediction data for this accumulation were hours 6 to 30. Details of the observational data used are provided in section 2.3.

2.1.2 AREA OF INTEREST

The area of interest is the South African domain, located at the southern tip of Africa between 22° to 35°S and 16° to 33°E (Figure 2.1). South Africa is primarily a summer rainfall region, with only the southwestern Cape of South Africa being a winter rainfall region (Tyson, 1986). The accuracy and skill of the multi-model forecasts and the various ensemble members are evaluated over the full South African domain for the summer half-years under consideration.

2.2 RAINFALL DATA

Daily rainfall data for the summer half-years (October to March) for the 2006/07, 2007/08 and 2008/09 seasons were obtained from weather stations managed by SAWS and the Agricultural Research Council (ARC). SAWS has a dense network of manual and automatic weather stations distributed across the country. Figure 2.2 indicates the distribution of the SAWS and ARC observational stations. The manual stations report 24-hour accumulated rainfall in the morning (06:00 UTC) and the automatic stations report hourly accumulated rainfall totals. For the automatic stations, hourly observations were accumulated to the same 24-hour period as the manual stations. Hourly data from the ARC automatic stations were accumulated in a similar way, but with some additional quality control measures as detailed in the section below.

2.2.1 QUALITY CONTROL OF RAINFALL DATA

2.2.1.1 South African Weather Service data

The SAWS performs qualify control on their daily rainfall data, which results in every rainfall value to be accompanied by a quality flag, e.g. “Accumulated' or “Normal”. Only daily rainfall values accompanied by the flag ‘Normal’ (indication that rainfall amount is representative of previous 24-hour period) were considered in the development of the daily rainfall grids. The “Accumulated” flag is used when rainfall was observed at a station for a period greater than 24-hours.

2.2.1.2 Agricultural Research Council – Institute for Soil, Climate and Water data

In order to correspond to the period of availability of SAWS daily rainfall data, hourly rainfall from the ARC stations were accumulated to the 24-hour period ending at 06:00 UTC, on condition that each of the hourly values for a particular station were present. A 24-hour accumulation was not performed for the last day of each March, as data for the last 6-hour period of the day (ending at 06:00 UTC on 1 April) was not available. Prior to the accumulation of the hourly values to daily values, a station location test were performed to ensure that all the stations utilized had unique locations. Stations with shared geographical positions were not included in the derived daily rainfall dataset.

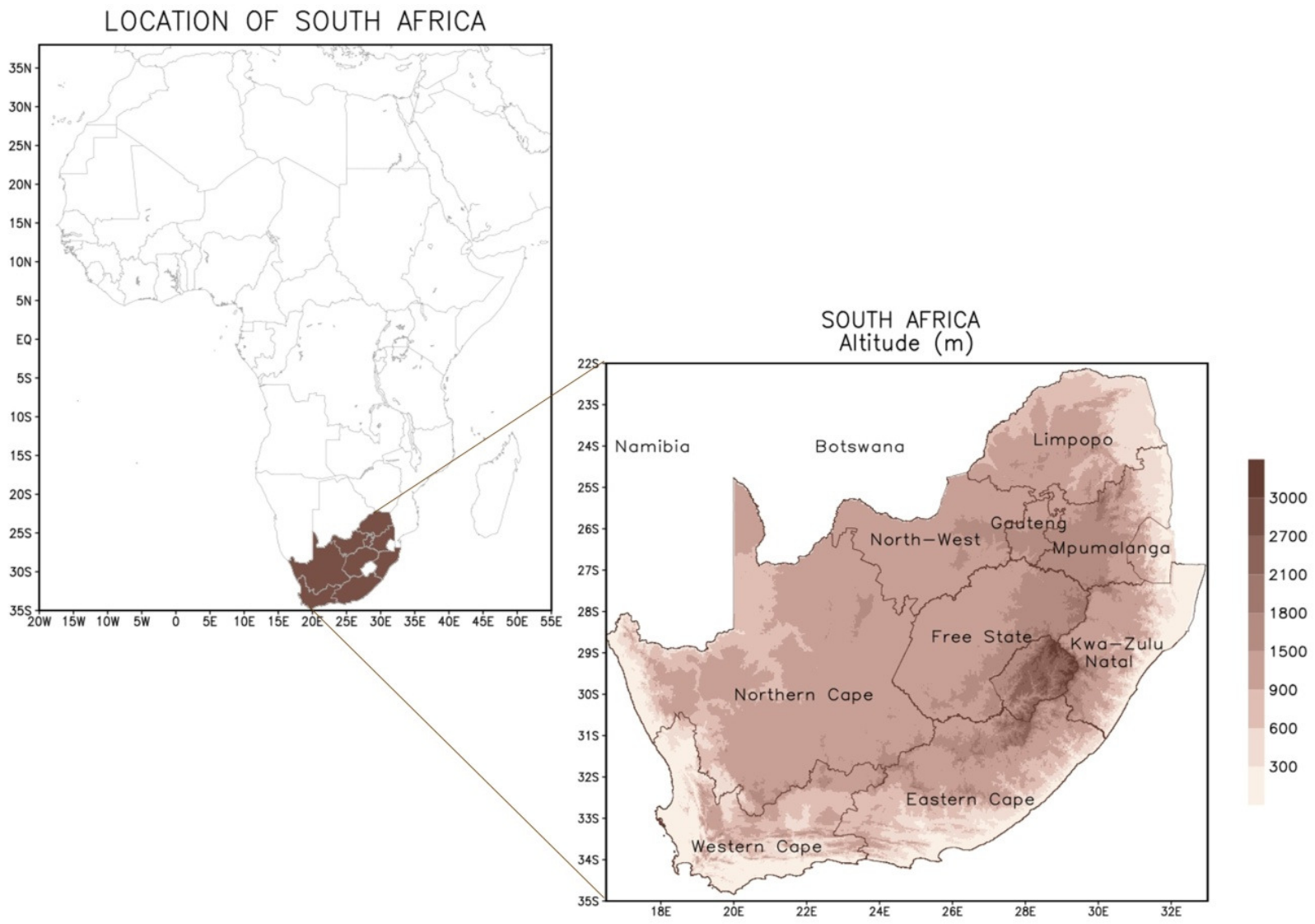


FIGURE 2.1: Location and topographical map of South Africa

2.2.1.3 Combined rainfall data set

All the daily rainfall values in this combined data set were subjected to three basic quality control measures, prior to the construction of the rainfall grid. These quality control measures are the following:

i. Extreme value test

Daily values that exceeded 597 mm were set to missing values. The threshold of 597 mm was chosen as it is the current validated 24-hour rainfall record (Lynch, 2003) over South Africa. On 31 January 1984, 597 mm was recorded at St Lucia when cyclone Demoina invaded the northeastern part of South Africa.

ii. Internal consistency test

This quality control measure tests for non-logical values. All negative rainfall values (not equal to the negative value used to indicate missing data) are set to missing values.

iii. Continuous no-change test

This quality control measure identifies all cases where data in the station time series display the same value for a chosen number of consecutive days (with the exception of missing values and 0 mm of rainfall). For such cases, rainfall values that are identical for at least three consecutive days are excluded from the data set.

After all the quality control measures had been preformed, data from approximately 2050 rainfall stations were available for analysis in the combined data set. Figure 2.2 shows the spatial distribution of the rainfall stations. It can be seen from the figure that the stations are concentrated over the urban areas of the country and that in some areas; especially the Northern Cape Province the station density is very low. It should be noted that not all of the rainfall stations were operational every day during the study period, and consequently the number of rainfall stations for which data were available for analysis varied between approximately 2000 in the 2008/09 season and 2060 in the 2006/07 season (Figure 2.3).

In Figure 2.3 it can be seen that there was a steady decline in the number of stations from February 2007, with a minimum total number of stations recorded on the 14th of February and 2nd of March (2007). Minimum AWS numbers (during the study period) were recorded twice during December 2007. Figure 2.3 shows that the availability of rainfall stations was more constant during the last summer half-year of the study

period. However, the average total number of stations available is lower than that during the 2006/07 summer season.

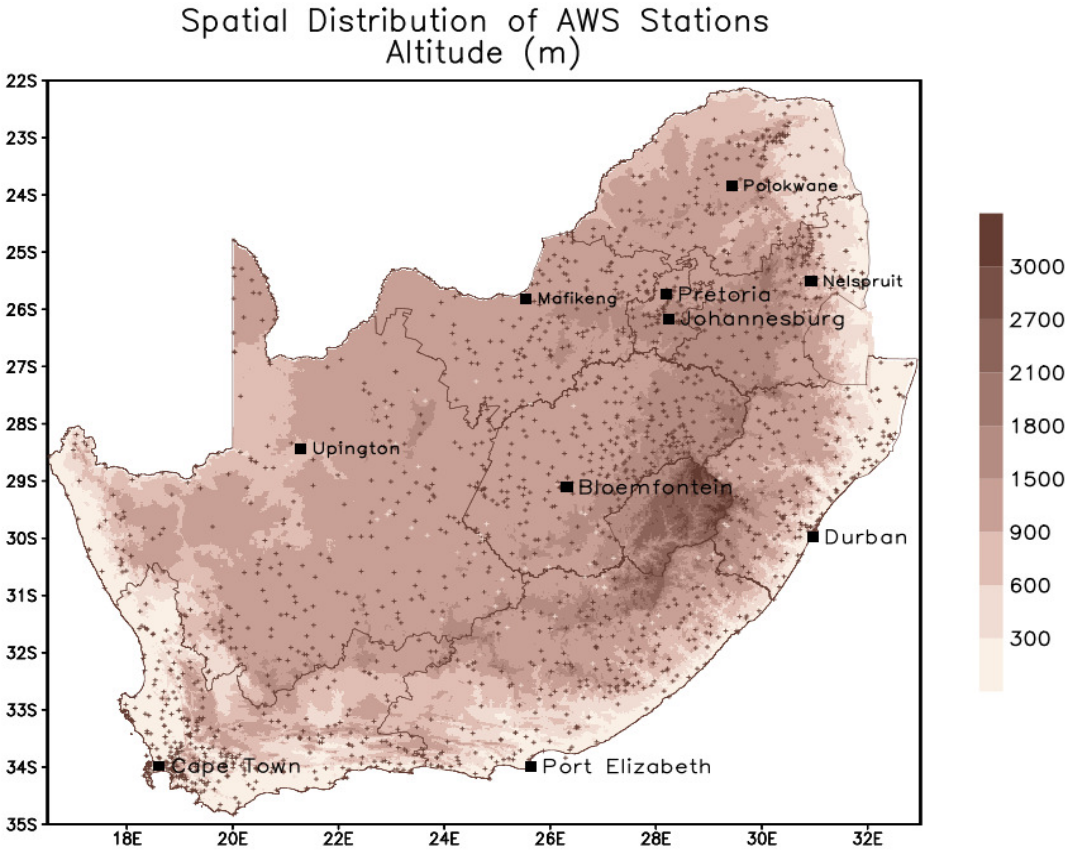


FIGURE 2.2: Location of rainfall stations in the combined SAWS-ARC data set.

For the 2006/07 summer half-year, a total of 2074 stations are available for the construction of the daily rainfall grids (i.e. Figure 2.4) while there are 2065 and 2057 stations available for the 2007/08 and 2008/09 summer half-years, respectively.

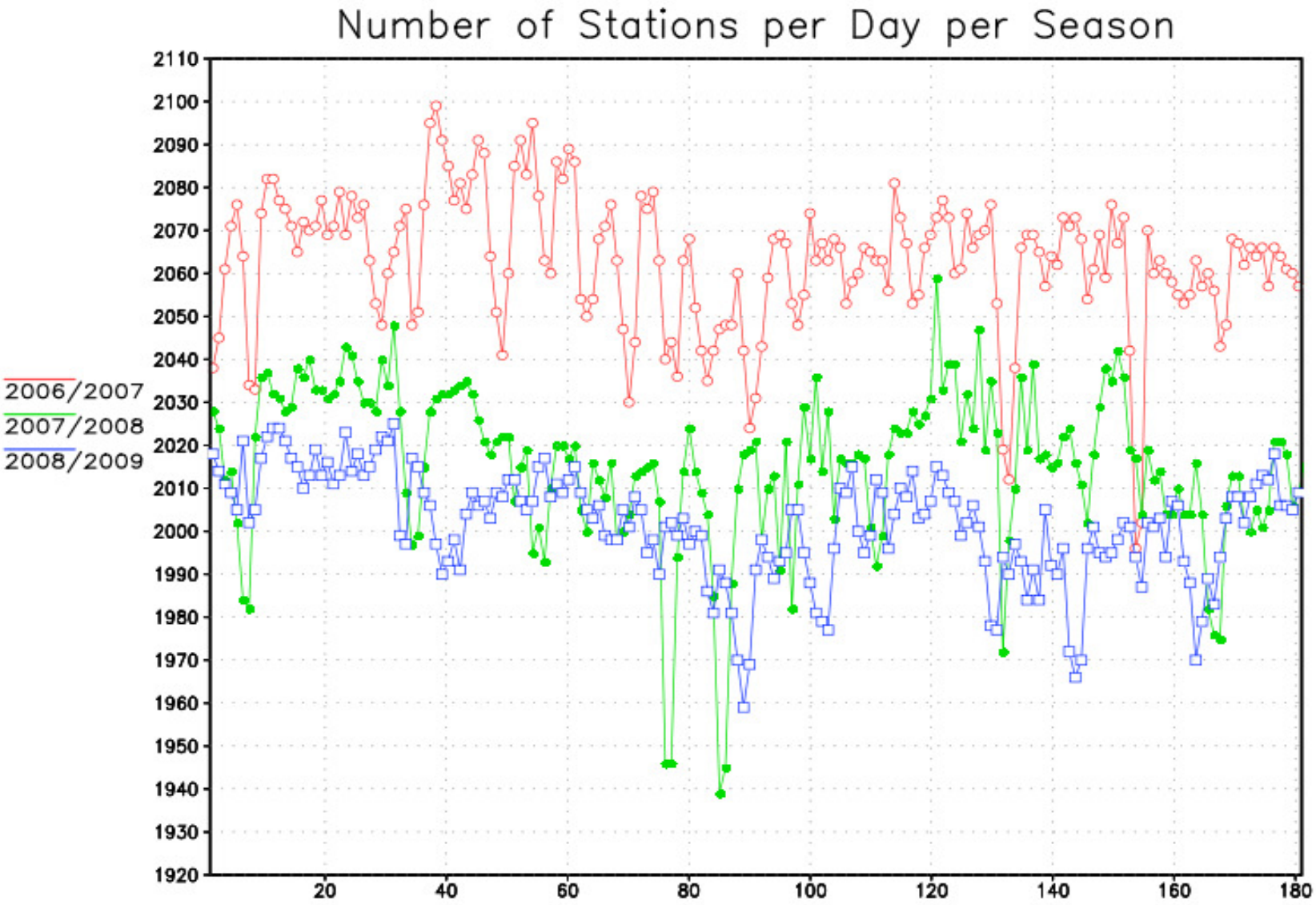


FIGURE 2.3: Number of daily rainfall stations in the combined data set per season

2.3 CONSTRUCTION OF A HIGH-RESOLUTION GRID

In order for the numerical precipitation forecasts to be compared to the observed rainfall, the rainfall totals recorded at the weather station locations were grouped in to a grid field with horisontal resolution of 0.25º. An average rainfall value was then calculated for each grid box using the box-average technique (Peel and Wilson, 2008). This procedure has been shown to successfully represent station data on the same grid field as that of the numerical weather prediction output (Peel and Wilson, 2008). A different number of rainfall stations were used in the calculation of the average grid box value, depending on the availability of rainfall stations in the geographical position demarcated by the grid box. Figure 2.4 shows the average number of rainfall stations used in the calculation of the average rainfall values for the different grid boxes. There are areas within the 0.25° grid where no rainfall stations are available (white grids), and several grid boxes which only had one station. The highest number of rainfall stations per grid box is present over the south-west coast and surrounding areas, where as many as 25 stations were used in the calculation of the average grid box rainfall.

If no stations were present within a specific grid box, the grid box was excluded from the subsequent analysis of forecast statistics. The results of the verification will be sensitive to the number of stations per grid box. However, for this study the minimum number of stations per grid box was chosen to be one. If the minimum number with at least two stations per grid were chosen, then the observation grid would have fewer samples, particularly in sparsely covered regions, and the results would be skewed toward more populous regions of South Africa. Even with a grid box with only one station which has the characteristics of a point measurement, making comparison with a model grid box average a bit more problematic, the greater number of observational grid boxes was chosen to be the better option for the purpose of this study.

Two regions are prominent with regards to station density, namely the extreme southwestern part of the country as well as the central parts of Gauteng and the northeastern Free State. The Northern Cape Province has the lowest number of rainfall stations. These regions are highly correlated to the population size of the region.

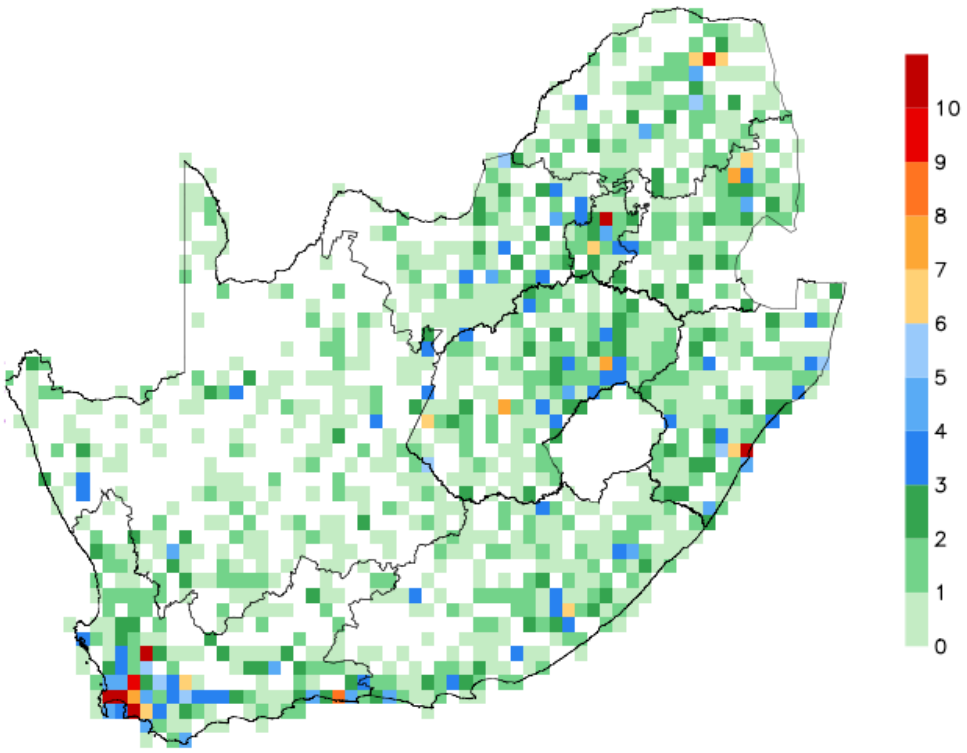


FIGURE 2.4: The average number of stations per grid box for 0.25° horisontal resolution

2.4 NUMERICAL WEATHER PREDICTION DATA

2.4.1 UNIFIED MODEL

The Unified Model (UM) is a non-hydrostatic model developed at the UK Met Office. Its vertical coordinate is based on geometric height. The UM can in principle be applied at time-scales ranging from weather forecasting to climate projection, and at resolutions ranging from relatively low to very high resolutions beyond the validity of the hydrostatic assumption (Davies *et al*, 2005). The UK Met Office runs the UM at global scale with horisontal resolution of 40 km, providing initial and boundary conditions for a regional version of the UM. Since May 2006, the UM version 6.1 has been running operationally at the SAWS with different configurations, including various horisontal resolutions, parameterizations schemes and data assimilation processes (Tennant, 2007). The three configurations used in this study are a 12km resolution run without data assimilation, a 12km resolution with continuous 3D-Var (three-dimensional variational assimilation; Kalnay, 2003) data assimilation and a 15km resolution run (also without data assimilation). Data assimilation is a statistical method of combining the latest observational data and the first guess field from the previous short-range forecast for the same period (Kalnay, 2003). All three of the configurations run in-house at the SAWS on a NEC SX-8 supercomputer.

The cloud scheme used by the UM at the SAWS is the 2A scheme, the Standard Smith Scheme, which diagnoses the liquid cloud fraction and liquid cloud water. The frozen cloud fraction is diagnosed from input frozen cloud condensate, assuming an inverted Smith scheme relationship between the two. Liquid and frozen cloud fractions are combined to produce a total bulk cloud fraction (Wilson and Bushell, 2007).

The SAWS UM applies the Prognostic Cloud 2 (PC2) scheme for convective systems, which in turns solves the 4A mass-flux convection scheme for CAPE scaling. Environmental compensation for initial parcel excesses ensures the environment is cooled and dried to account for the excess heat and moisture attributed to the parcel as convection is initiated (Wilson and Bushell, 2007).

2.4.1.1 12 km no Data Assimilation

The 12 km no-DA UM run covers the sub-continent of southern Africa as well as large areas of the surrounding oceans (Figure 2.5). This configuration runs once a day with 38 levels in the vertical, and produces forecasts 48 hours ahead from the initialized field at 00:00 UTC (Tennant, 2007).The forecast output fields are written every hour. This run uses the 18:00 UTC forecast from the UM Global Model to provide initial conditions to the 12 km run at 00:00 UTC, as well as lateral boundary condition fields.

2.4.1.2 12 km Data Assimilation

This configuration field has the same domain and resolution as the 12 km no data assimilation run, but incorporates continuous 3DVAR data assimilation. The assimilation process is repeated every six hours, forecasting six hours ahead, i.e. four times a day, but at the 00:00 UTC assimilation update, the model continues to forecast 48 hours ahead.

2.4.1.3 15 km no Data Assimilation

The 15 km horisontal resolution run has a much smaller domain (Figure 2.6), than the two before-mentioned UM runs. It is set up to cover only the South African domain, covering 22ºS to 35ºS and 15ºE to 34ºE, making it less computer intensive. This configuration has no data assimilation, and also uses the 18:00 UTC forecast from the UM Global Model to provide 00:00 UTC initial conditions.

The three UM model runs were nearly 100% available for all three half-years (Table 2.1). Exceptions are the 2006/07 season, due to missing days in October 2006 as a result of computational errors within the archiving system of the two 12 km configuration runs.

TABLE 2.1: Data availability of the UM per season

Model run	2006/2007	2007/2008	2008/2009
12km no-DA	97%	100%	100%
12km DA	86%	100%	100%
15km no-DA	100%	100%	100%

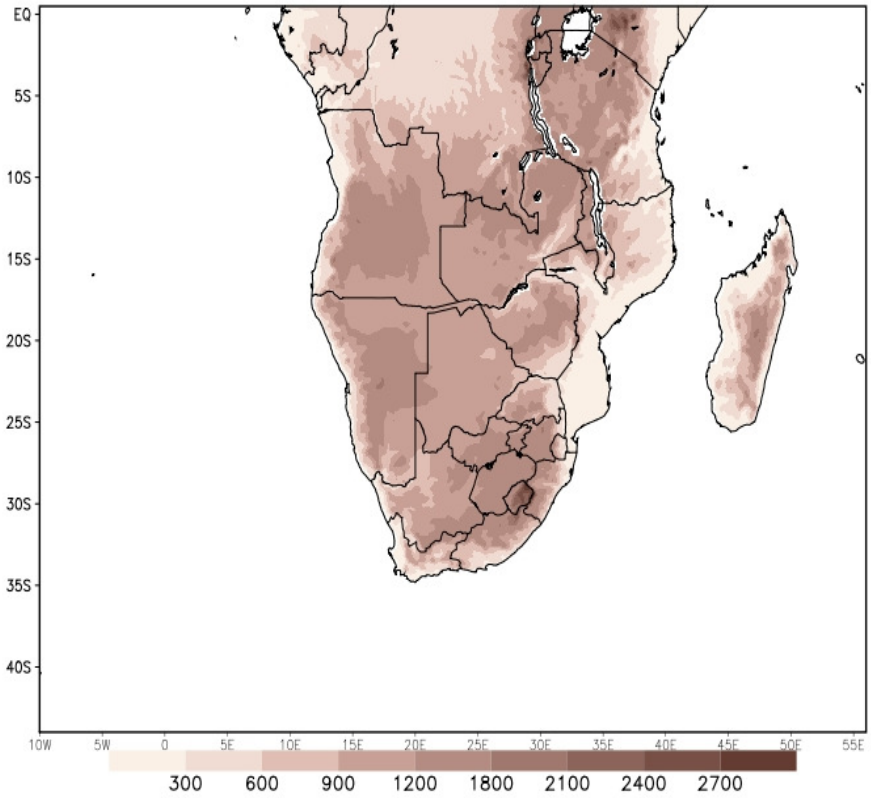


FIGURE 2.5: The 12km horisontal resolution domain of the UM operational at the SAWS. The shaded values are the model topography (m).

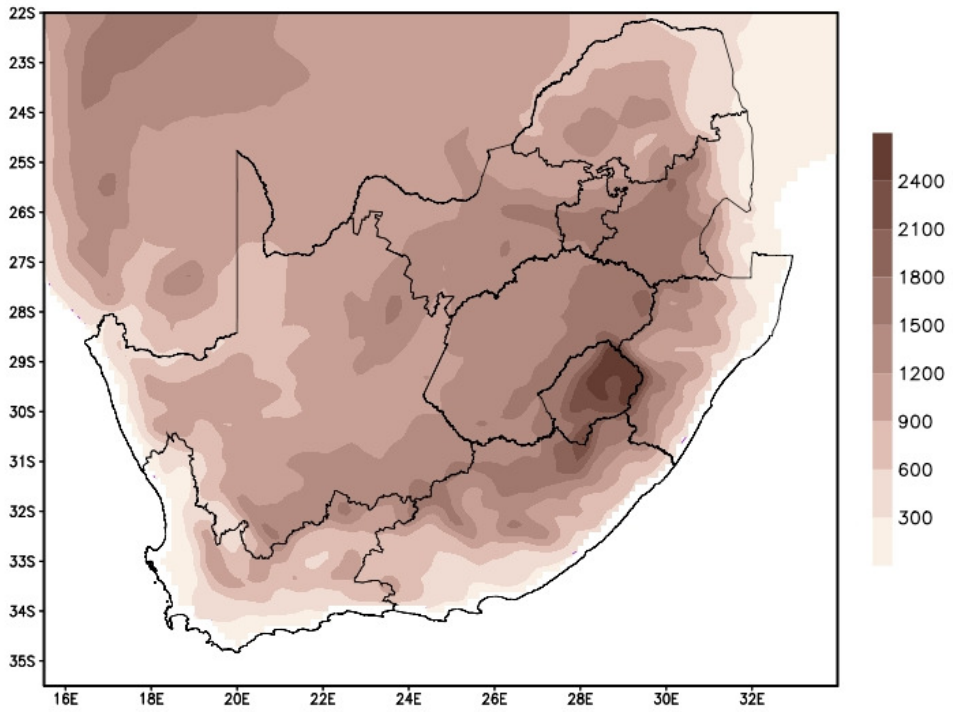


FIGURE 2.6: The domain size of the 15 km UM member

2.4.2 CCAM MODEL

The Cubic-Conformal Atmospheric Model (CCAM) was developed at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Australia (McGregor, 2005). CCAM is a variable-resolution global model, that may be applied either in quasi-uniform mode to function as a global circulation model, or alternatively in stretched-grid mode to provide high resolution over an area of interest. The model solves the hydrostatic primitive equations using a semi-implicit semi-Lagrangian method. It includes a comprehensive set physical parameterizations (see McGregor, 2005), including a cumulus convection scheme that uses a mass-flux closure and which includes downdrafts, entrainment and detrainment (McGregor, 2003). Interactive cloud distributions are determined by the liquid and ice-water scheme of Rotstayn (1997). It has been illustrated by Engelbrecht *et al.* (2009) and Engelbrecht *et al.* (2011) that the CCAM is capable of satisfactorily simulating many attributes of the present-day climatological conditions over southern and tropical Africa. The model has also been shown to produce skillful short-range and seasonal forecasts over the southern African region (Potgieter, 2007; Ghile and Schulze, 2010; Landman *et al.*, 2010).

The CCAM became operational at the Council for Scientific and Industrial Research (CSIR) in 2010, so that hindcast data were created in order to perform verification studies for the three summer half-years relevant to this study. In operational mode,

the CCAM is initialized at 00:00 UTC, using initial condition fields obtained from the Global Forecast System (GFS). Two different 7-day forecasts are produced daily using the 00:00 UTC initial state. A forecast that has a resolution of about 60 km over southern and tropical Africa is performed first. In order to obtain this forecast the model is applied in stretched-grid mode over southern and tropical Africa (Figure 2.7), with the resolution decreasing to about 400 km in the far-field. A high-resolution forecast is subsequently performed using a strongly-stretched grid that provides resolution of 15 km over southern Africa, with this run nudged within the 60 km forecast. Hindcasts for the three half-years under consideration were performed using a set-up that mirrors the operational forecasting system. For both the 60 km and 15 km hindcasts, model output is available at six-hourly time-steps over a domain that covers southern and tropical Africa.

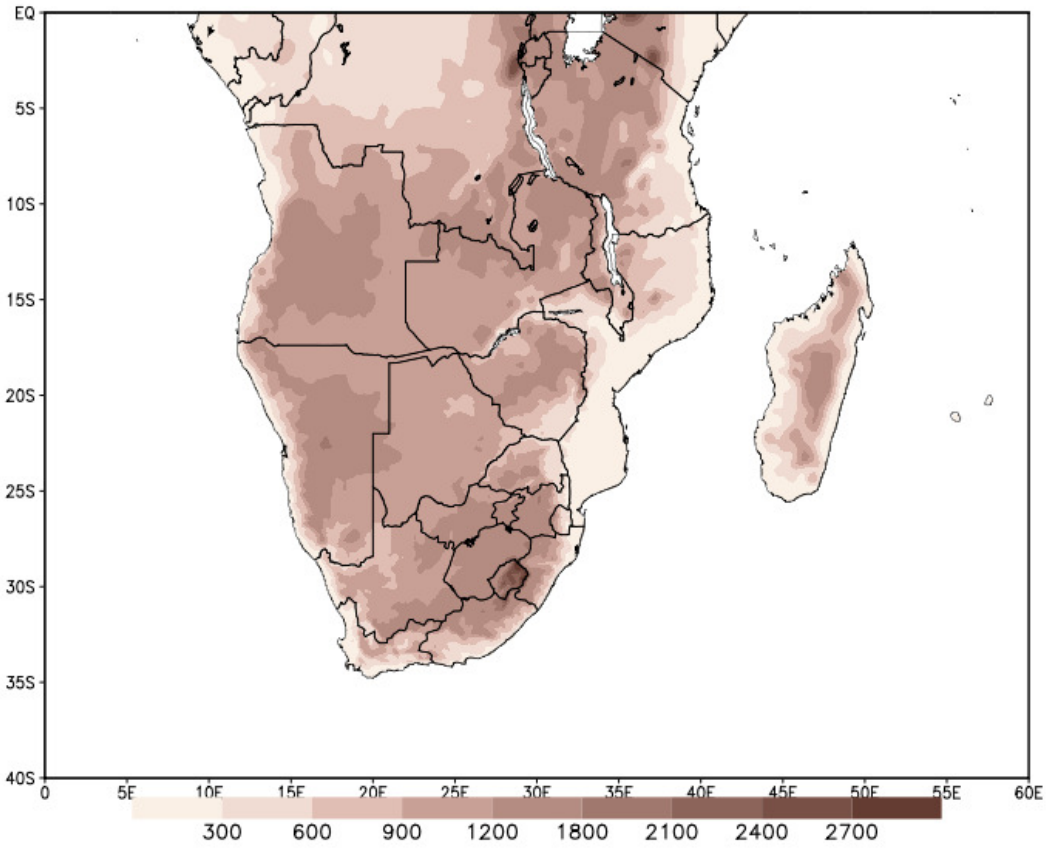


FIGURE 2.7: Output domain of the CCAM forecasts

2.4.3 NCEP GLOBAL ENSEMBLE FORECAST SYSTEM

The National Centres for Environmental Prediction (NCEP) runs the Global Ensemble Forecast System (GEFS) operationally. The output is downloaded at the SAWS and primarily used for medium-range forecasting (4 to 14 days ahead) by the forecasters at SAWS's National Forecast Centre (Tennant, 2007). The GEFS consists of runs configured at 00:00 UTC, 06:00 UTC, 12:00 UTC and 18:00 UTC. For each of these time intervals, there is one control run plus 14 additional members, forecasting 16 days ahead. The horisontal resolution is 1º (~100 km) with 28 levels in the vertical. The ensemble members are created using the ensemble transform perturbation method (Kalnay, 2003). The model used to generate the ensemble is the Global Spectral Model.

Due to data connectivity limitations, the time that it takes for these files to be downloaded to SAWS effectively implies that the full suite of forecasts issued by the GEFS on a given day, is only available to the SAWS forecasters on the next day. The SAWS therefore always use the GEFS forecast with a one day lead time. Note that each forecast also consists of 60 ensemble members (14 members + control run x 4 initialization times).

2.5 DESIGN OF THE SHORT-RANGE MULTI-MODEL ENSEMBLE SYSTEM

The high-resolution multi-model ensemble (MMENS) is only compiled for the South African domain (Figure 2.1). The system is formulated with the purpose of predicting the probability of precipitation exceeding a certain rainfall threshold, over a 24-hour period from 06:00 UTC to 06:00 UTC. Although each of the five ensemble members (UM 12 km, UM 12 km data assimilation, UM 15 km, CCAM 15 km and CCAM 60 km) described in the above-mentioned sections covers a bigger domain than the South African domain (22º to 35ºS and 16ºE to 33ºE) the spatial extent of the observational network limits the verification analysis to this smaller domain. Model output is regridded to a horisontal resolution of 0.25° over the SA domain, by using the same box averaging technique applied to the observational station data. In this study threshold values of daily rainfall totals are considered in order to calculate the probability of these threshold being exceeding within a 24-hour period. These values are 1 mm, 10 mm, 25 mm and 50 mm, with the latter representing extreme, and rare rainfall events. For the extreme rainfall threshold of 50 mm, the number of events is

significantly less than those of the other thresholds. Figure 2.8 shows the number of times the threshold of 50 mm/day was exceeded per month over the three seasons. The maximum number of times the extreme event did occur was 7 times during December over the Eastern Cape Province. Ebert (2001) noted that 1 mm/day threshold is useful in the construction of gridded rainfall field, in order to eliminate dew and insignificant rain. The forecast skill in predicting rainfall occurring at or above each of the various thresholds was subsequently investigated.

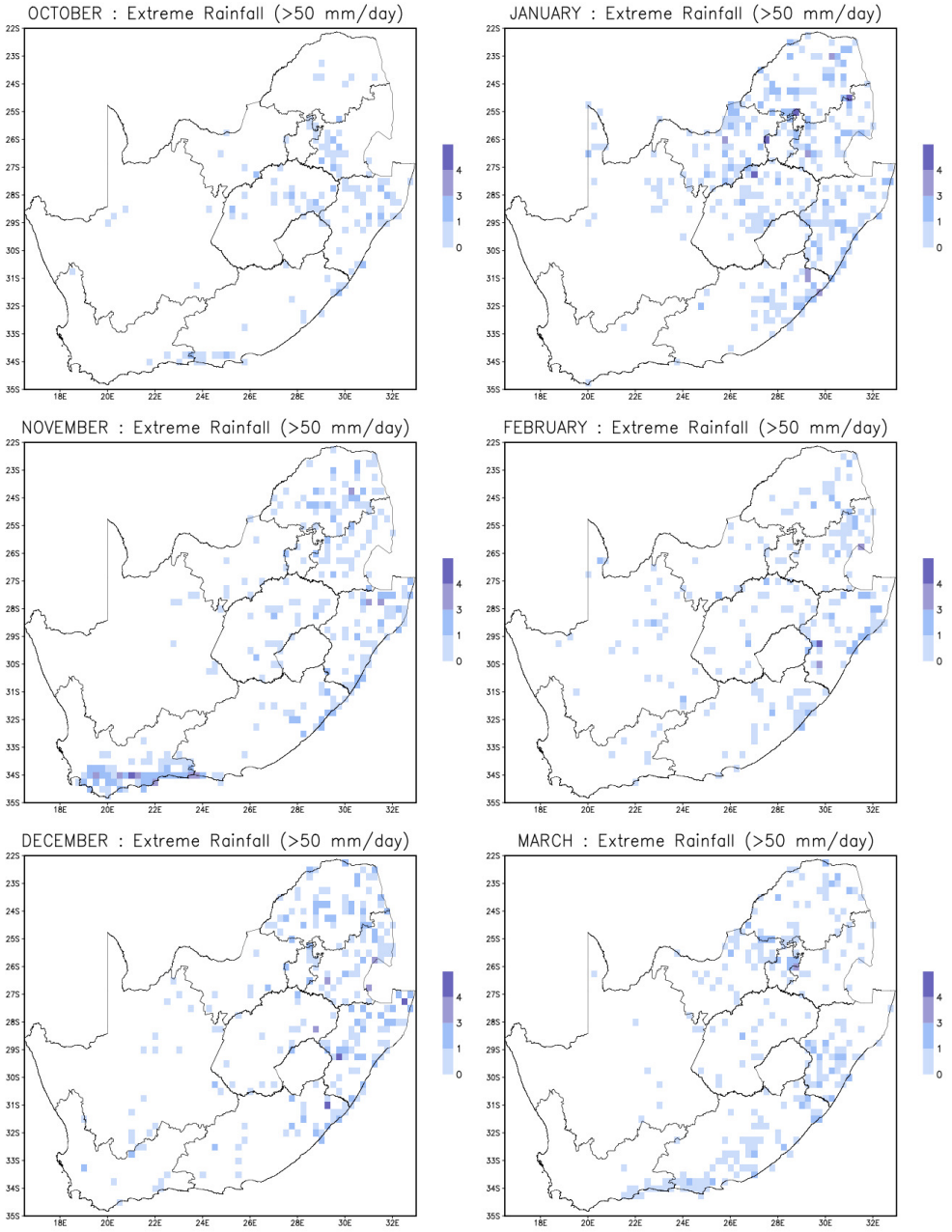


FIGURE 2.8: The total number of extreme rainfall events exceeding 50 mm/day for the six months under investigation

Apart from the MMENS, an ensemble for the individual models (the UM and CCAM) was also created, in order to establish the skill of each of the single-model ensemble forecasts, and to determine the influence that each of the single-model ensemble systems has on the multi-model system.

The four different 24-hour rainfall total thresholds are considered in order to formulate five different dichotomous forecasts for each threshold value. That is, for a given threshold a value of zero is assigned to the forecast if the threshold is not exceeded and a value of one if the threshold is exceeded.

The individual members of each of the ensemble members contribute equal weights to the respective single-model ensemble systems. The UM ensemble (UMENS) probability of precipitation (above a certain threshold) is created by adding the dichotomous forecast (1 or 0) obtained from the deterministic rainfall forecast of each forecast at a specific grid point for each of the members, and to then divide by three (the number of model configurations). Symbolically,

$$UMENS = \frac{UM1 + UM2 + UM3}{3} = \frac{1 + 0 + 1}{3} = 0.66\%$$

The same process is repeated for the CCAM ensemble (CCAMENS), by adding the two dichotomous forecasts and to then divide by two in order to obtain the probability forecast from the CCAMENS:

$$CCAMENS = \frac{CCAM1 + CCAM3}{2} = \frac{1 + 0}{2} = 0.50\%$$

The MMENS is also created by applying equal weights to the two single-model ensemble systems described above.

$$MMENS = \frac{UMENS + CCAMENS}{2} = \frac{0.66 + 0.5}{2} = 0.58\%$$

The probability of precipitation from the UMENS is added to that of the CCAMENS and the total is then averaged. It was decided that with this method, the UM with the most members will not carry more weight in the MMENS system.

For the deterministic forecasts of each of the ensemble systems, the average rainfall amount was calculated in the same manner as described above for the dichotomous forecasts, but by adding the deterministic rainfall amount and averaging the value.

2.6 THE NCEP ENSEMBLE DESIGN

An ensemble forecast based on the 00:00 UTC initialized members of the NCEP GEFS data was created for each day of the study period. The 00:00 forecasts were selected for this purpose in order to keep the forecast similar to the set-up of the multi-model system. This single-model ensemble system is useful to establish a reference value of skill for the MMENS to beat – that is, the MMENS should at least provide forecasts more skillful than that of this single-model, low resolution 15-members ensemble (constructed from 1 control member and 14 perturbed members). The 1° resolution NCEP forecasts were regridded to the 0.25° resolution grid for comparison purposes by using a bi-linear interpolation method. The probability forecast is created in a similar manner as the single-model high resolution ensembles where the individual forecasts is converted to a dichotomous forecast and then averaged in order to obtain a probability of precipitation. Similarly, for the deterministic forecasts, the average rainfall amount from the 15 member ensemble was calculated for each grid point.

2.7 VERIFICATION METHODS

For each of the three summer half-years, verification is performed separately for the six months that make up the summer half-year. That is, for each of the months the daily forecasts issued for that month over the three half-years are considered when calculating the statistical verification scores. In example: for October (assuming you have 100% data availability) the daily forecasts of October 2006, October 2007 and October 2008 will equal to a large sample of 93 days, which in turn is then verified as one forecast sample. Each forecast of daily rainfall totals was converted into four sets of dichotomous forecasts, based on the four different threshold values under consideration (section 2.5). The verification statistics were calculated for the forecasts regridded to the 0.25° resolution grid.

2.7.1 VERIFICATION OF FORECAST ACCURACY

The forecast bias (as defined by Eq. 2.1) explores whether a variable under consideration is systematically over-forecast or under-forecast, and is a measure of forecast accuracy. The perfect forecast would have a bias of 0. In this study, the average bias of the deterministic forecast of all the ensembles and individual members are calculated on a monthly basis (for each grid box on the 0.25 degree resolution grid).

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i) \qquad 2.1$$

Here N represents the total number of forecasts issued for a specific month for each grid box over the three summer half-years under consideration, f_i is the forecast and o_i is the corresponding observed value.

Together with the bias, the forecast frequency is calculated to show how many times during the period of interest a certain threshold was exceeded. For a perfect forecast, the forecast frequency of events occurring in a given threshold category will exactly match the observed frequency of occurrence of events above the threshold under consideration.

2.7.2 VERIFICATION OF FORECAST SKILL

To determine the skill and quantify the quality of a forecast, the forecast has to be compared to that of a reference forecast (Joliffe and Stephenson, 2003). A skill score compares the value of the forecast issued to that of another forecast, for example a forecast obtained from the climatology of the variable. In this study, persistence was used as a reference forecast due to the short-range period forecast (Wilks, 2006). Persistence is the observation of the event during the previous time period, therefore for this study it is the previous days' 24-hour total rainfall observation.

2.7.2.1 Brier Skill Score Calculations

The Brier skill score (BSS; Stanski *et al*. 1990) is derived from the Brier score (BS; Wilks, 2006; Fawcett, 2008). The BS consists of the mean squared error in the probability forecast (Eq. 2.2). Here f_i assumes a value of 1 if the event was forecast and 0 if the event was forecast *not* to occur. Similarly, a value of 1 is assigned to o_i if the event did occur and 0 if the event did not occur. The same method applies in calculating the BS for the reference forecast where o_{n-1} will be equal to 1 or 0 if the

event on the previous day exceeded the threshold value (Eq. 2.3). The BSS (Eq. 2.4) answers the question of the relative skill the forecast (predicting whether the event occurred or not) has over that of the persistence (reference) forecast (Mason, 2004). The BSS is generally known as a verification tool for probability forecasts, but can also be applied to deterministic forecasts when a threshold value is applied. The deterministic forecast is then transformed into a categorical forecast of a dichotomous predictand (Theis *et al*, 2005; Wilks, 2006). The general range of the skill score is from minus infinity to 1, where the perfect score will be equal to 1 and no skill improvement over that of the reference forecast equals to 0, and no forecast skill for all negative values.

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \qquad 2.2$$

$$BS_{ref} = \frac{1}{N}\sum_{i=1}^{N}(o_{n-1} - o_i)^2 \qquad 2.3$$

$$BSS = 1 - \frac{BS}{BS_{ref}} \qquad 2.4$$

For comparing the MMENS and the NCEP ensemble systems, the three independent terms of the Brier score is calculated. In order to calculate the three terms, the following equations needs to be solved where n is equal to the total number of forecasts issued over the evaluation period and grid points and N is the number of forecasts for each forecast probability bin.

$$\bar{o}_i = \frac{1}{N_i}\sum_{k \in N_i} o_k$$

This equations determines the number of times the events was observed o_k for each of the forecasts made for each probability, N_i. The equation below determines the relative frequency ($\bar{o}$)of observations.

$$\bar{o} = \frac{1}{n}\sum_{i=1}^{I} N_i \bar{o}_i$$

The reliability term needs to be as small as possible, which will indicate a well calibrated forecast because it summarizes the conditional bias of the forecast. Here y_i is the probability forecast value.

$$Reliability = \frac{1}{n}\sum_{i=1}^{I} N_i(y_i - \bar{o}_i)^2$$

The resolution term needs to be as large as possible, which will indicate that the forecast resolves the event strongly because it summarizes the ability of the forecasts to discern between events and non-events.

$$Resolution = \frac{1}{n}\sum_{i=1}^{I} N_i\,(\bar{o}_i - \bar{o})^2$$

The uncertainty term is only dependent on the climatological frequency of an event occurring and therefore is not influenced by the forecast.

$$Uncertainty = \bar{o}(1 - \bar{o})$$

2.7.2.2 Mean Squared Error Skill Score Calculations

The mean squared error skill score (MSESS) is another measure of the accuracy of a forecast. It is sensitive to the presence of large individual forecast errors. The MSESS is formulated in a way similar to the BSS, but with f_i and o_i now representing the actual forecast and observed values. That is, the difference with determining the skill of the forecast using the MSESS is that the skill is determined based on the deterministic forecast and not a probability forecast (Wilks, 2006).

2.7.3 ADDITIONAL VERIFICATION OF DICHOTOMOUS FORECASTS

A number of additional statistical scores exist to further explore the accuracy of the dichotomous forecasts based on different thresholds. For each forecast or observation where the threshold is exceeded, the non-probabilistic forecast becomes “yes (or 1)” and “no (or 0)” if the threshold is not exceeded. This forecast is then analysed using a contingency table which shows the frequency of “yes” and “no” forecasts relative to the observed occurrences (Joliffe and Stephenson, 2003; Wilks, 2006, Fawcett, 2008; Table 2.2). The series of verification statistics obtained in this way, for various threshold values, give an indication of the forecast to correctly predict the occurrence as well as the amount of rainfall (Ebert, 2001). This process is applied separately to all the individual members, and to the different ensembles that were formulated. Usually, the contingency table are set-up to explore the average

forecast performance over a model domain. This has the disadvantage that the verification scores represents an area average (Ebert, 2001) and cannot distinguish between different geographical locations of the domain or different weather regimes. For this reason, a contingency table is calculated for each grid box in the domain, and the scores calculated to present forecast performance at each grid box. In this manner the spatial patterns of the forecast performance can be evaluated.

Table 2.2 outlines a number of forecast attribute measures that can be obtained from a contingency table based on dichotomous forecasts.

TABLE 2.2: The contingency table for the analysis of dichotomous forecasts

		OBSERVED		
		YES	**NO**	
FORECAST	**YES**	***a***	***b***	*forecast yes*
	NO	***c***	***d***	*forecast no*
		observed yes	*observed no*	**TOTAL**

a : HITS – the event was forecast and observed
b : FALSE ALARMS – the event was forecast but not observed
c : MISSES – the event was not forecast but observed
d : CORRECT NEGATIVES – the event was neither forecast not observed

Based on the forecast attribute measures obtained from the contingency table, a number of statistical scores measuring forecast performance can be formulated.

2.7.3.1 Frequency Bias Index (Bias Score)

The frequency bias index (FBI) measures the ratio of the forecast frequency of "yes" events to the observed frequency of "yes" events (Eq 2.5). For the FBI the score is perfect when equal to 1. The FBI indicates whether or not the system has a tendency to over- or under-forecast rainfall frequencies (Joliffe and Stephenson, 2003). Because the FBI does not distinguish between the timing of the forecast and the observations, it is not a measure of accuracy.

$$FBI = \frac{a+b}{a+c} \qquad 2.5$$

2.7.3.2 Probability of Detection (Hit Rate)

The probability of detection (POD) or the hit rate is defined by Joliffe and Stephenson (2003) as the proportion of occurrences of the forecast of the "yes" event being correct, relative to the total number of actual occurrences of the event (Eq. 2.6). A perfect score for POD is 1. In equation 2.6 it is shown that the POD is dependent on the "hits" and "misses" of the forecast.

$$POD = \frac{a}{(a+c)} \qquad 2.6$$

2.7.3.3 False Alarm Rate

In contrast to the POD, the false alarm rate (F) is defined as the proportion of events not occurring being forecast incorrectly to occur (Joliffe and Stephenson, 2003). The F is used in conjunction with the POD.

$$F = \frac{b}{b+d} \qquad 2.7$$

2.7.3.4 Critical Success Index

The threat score or critical success index (CSI) addresses how well the "yes" forecast events agrees to the "yes" observed events, and only measures the events that occurred or were predicted and therefore does not take into account correct negatives (Eq. 2.8). The CSI is especially useful during the occurrences when "no" events are forecast more often than "yes" events; i.e. extreme rainfall totals (50 mm/day). The value of the CSI ranges between 0 and 1, where 1 is the best CSI score.

$$CSI = \frac{a}{a+b+c} \qquad 2.8$$

2.7.3.5 Equitable Threat Score

The equitable threat score (ETS) is a modification on the CSI that takes into account the correct negatives (Eq. 2.9 and 2.10), included in the total number of events (n). Therefore its equitability is less sensitive to different atmospheric regimes (Joliffe and Stephenson, 2003). The ETS calculates the forecast events that were correctly predicted, while correcting for hits due to random chance. The ETS ranges from -1/3 to 1; negative scores have no skill and 1 is a perfect forecast (Clark *et al*, 2008).

$$ETS = \frac{a-a_r}{a-a_r+b+c} \quad 2.9$$

$$a_r = \frac{(a+b)(a+c)}{n} \quad 2.10$$

2.7.4 VERIFICATION OF PROBABILITY FORECASTS

2.7.4.1 ROC Curve

The Relative Operating Characteristics (ROC) curve is constructed by calculating the hit rate and the false alarm rate for each of the probability intervals, and is then plotted on an x-y graph (Clark *et al*, 2008). The ROC curve is a visual method of determining the discrimination of a forecast between events and non-events (Wilks, 2006; Fawcett, 2008). Forecasts with good discrimination show ROC curves approaching the upper-left corner of the ROC diagram (the bold curved line in Figure 2.9). A diagonal line (dotted in Figure 2.9) represents a forecast with no discrimination (Joliffe and Stephenson, 2003). The ROC is conditioned on the observations, determining the performance of the forecast if a given event occurred. For each of the ROC curves, the area under the ROC curve is calculated with the trapezoid method. This method is applied by adding the areas of the trapezoids formed by connecting the points on the curve (Wilks, 2006; Clark *et al*, 2008).

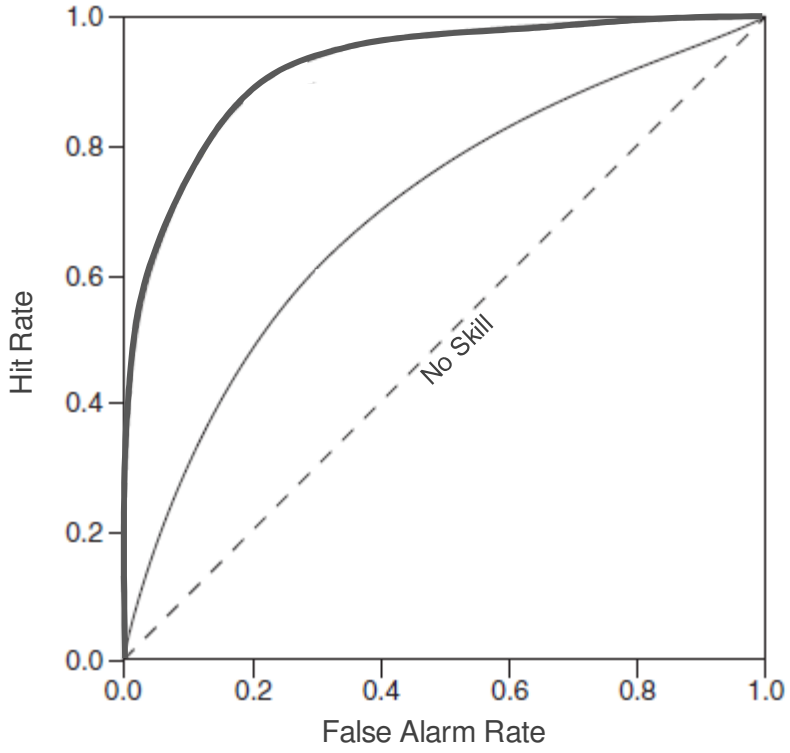


FIGURE 2.9: ROC graph (adapted from Wilks, 2006)

2.7.4.2 Reliability Diagram

The reliability diagram represents the relationship between the observed frequency and the forecast probability of an event (Joliffe and Stephenson, 2003, Wilks, 2006). The diagonal line on the graph in Figure 2.10 represents a forecast with good resolution whereas a horisontal line will indicate a forecast with minimum resolution (Wilks, 2006). The reliability diagram is a good companion to the ROC curve, where the reliability diagram is conditioned on the forecast. The reliability diagram shows what the observed frequency is given the forecast probability for that event to occur.

Together with the reliability diagram a sharpness or frequency diagram is constructed where the forecast probability bins are plotted against the frequency of the event forecast within each probability bin (over the verification period and at all the gridpoints). The sharpness diagram is an indication of the confidence of the forecast system under investigation.

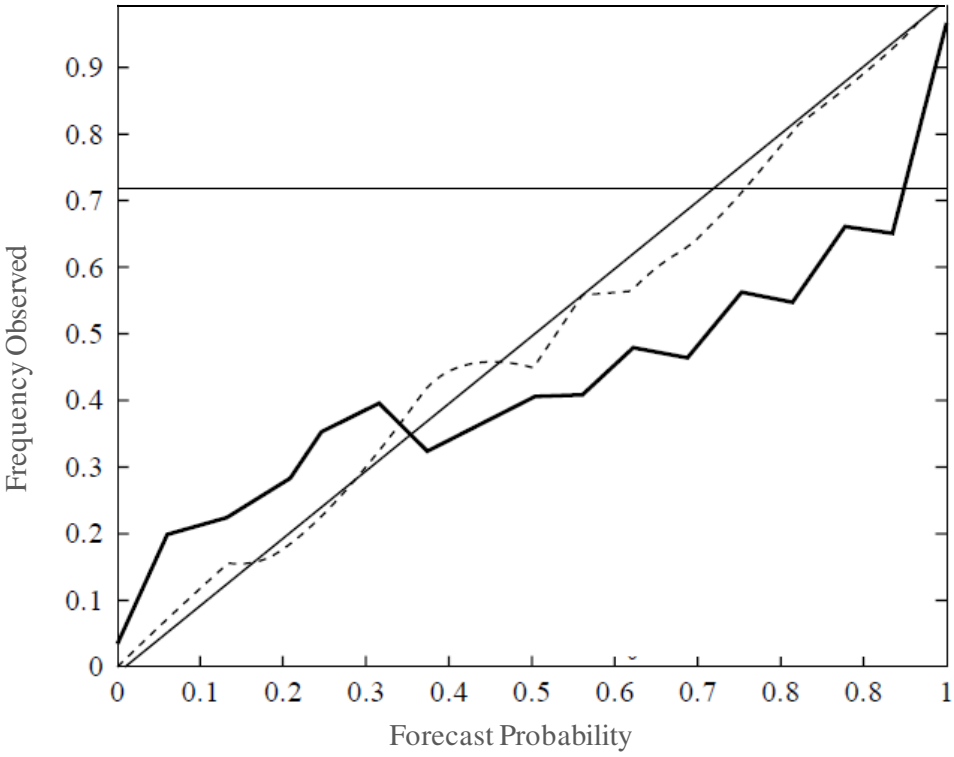


FIGURE 2.10: A reliability Diagram (adapted from Joliffe and Stephenson, 2003)

2.8 SUMMARY

The data, models and methods used to construct the different ensemble systems and multi-model ensemble system have been described. The properties of the high-

resolution regional models and course resolution global ensemble system have been detailed. Finally the data and methods used to verify the members and the ensemble forecast system have been discussed. In the next chapter the verification of the precipitation forecasts of the various systems are discussed.

CHAPTER 3

VERIFICATION OF THE SHORT-RANGE MULTI-MODEL ENSEMBLE PREDICTION SYSTEM

This chapter describes the verification results obtained using the short-range MMENS prediction system based on the Unified and CCAM models applied over the South African domain (Figure 2.1). The skill and accuracy of each of the individual ensemble members, as well as the three respective ensembles are verified. The three ensembles are the UMENS (average of the three UM members) the CCAMENS (average of the two CCAM members) and the MMENS (average of the UMENS and CCAMENS). Verification is performed by comparing the forecast output from each of the ensemble members to observed rainfall data. Firstly, each of the forecasts is verified deterministically over the entire domain for each of the four rainfall thresholds (1, 10, 25 and 50 mm/day), followed by the verification of the probabilistic forecasts as obtained from each of the ensembles. The verification is done on the 0.25° horisontal resolution as described in Chapter 2.

3.1 VERIFICATION OF FORECAST ACCURACY

3.1.1 BIAS CALCULATIONS

In order to better interpret the bias fields of the forecasts, Figure 3.1 shows the monthly observed daily mean rainfall. The average monthly bias in the forecast of 24-hour rainfall totals, for each of the months in the summer half-year, is shown for each of the three ensembles in Figure 3.2 a to c. These spatial maps represent the monthly bias for each grid box obtained from averaging each month in the summer half-year over the three seasons of 2006/07, 2007/08 and 2008/09. The spatial map of the bias provides insight into the distribution of areas with relatively high and low as well as positive and negative biases. For each of the fields of monthly biases, the spatial average bias calculated for the domain is also given on each map in Figure 3.2 (a) to (c). From these values it can be deduced which month displays the lowest area average bias for a given ensemble, as well as which ensemble provides the smallest bias for a given month. A summary of these values are given in Figure 3.3. This figure also contains the area average bias of the five individual members of the MMENS; UM 12 km (UM12), UM 12 km data assimilation (UM12DA), UM 15 km (UM15), CCAM 15 km (CCAM15) and CCAM 60 km (CCAM60).

In Figure 3.2a it is seen that the bias of the UMENS has a general wet bias, especially over the eastern parts of the country. Regions with a negative bias include the southwestern Cape and western interior, during the months of October to December. The Limpopo Province also displays a predominately dry bias during October. The month of March has the lowest overall bias for the domain (about 1.3 mm/day). December displays the highest bias of more than 2 mm per day

The CCAMENS (Figure 3.2b) shows a very similar wet bias pattern over the central interior as that of the UMENS. The CCAMENS under-forecast the rainfall amounts over the east-coast and the south-coast, especially during November. The months of February to March, the CCAMENS tend to over-forecast the rainfall amounts over the east-coast, with the lowest overall bias for November. The CCAMENS has the highest bias in February when the system in general has a wet bias.

The MMENS (Figure 3.2c) has a similar spatial pattern as the UMENS system, with a wet bias over the east-coast; contrary to the dry bias of the CCAMENS. The MMENS as with the UMENS tend to over-forecast the rainfall amounts over the Northern Cape Province for all of the six months considered.

In Figure 3.3 the area average bias for each of the individual members of the MMENS and the UM- and CCAM ensembles is given. Figure 3.3 reveals that the UM12DA member has significantly smaller biases compared to the MMENS forecasts for all of the six months. The remaining members of the UM model has a significant high bias, contributing to the high bias of the UMENS, hence the CCAMENS has the lowest area average bias of the three ensembles, except for March when the UMENS has the lowest bias.

Figure 3.4 shows the frequency of the forecast for each of the ensemble members as well as the frequency of observed rainfall amounts (black line). In Figure 3.4 graphs a to f, the monthly frequency of occurrence of observed and forecast rainfall events for each of the ensemble members and the different ensemble systems, for different rainfall categories are shown for October to March. The categories are calculated from zero rainfall per day observed and forecast, increasing incrementally to greater than 100 mm/day events in order to capture the rainfall frequency distribution of daily rainfall totals. This figure gives a more detailed perspective of the over- and under-forecast of each of the ensemble members.

When looking at graphs (a) to (f) in Figure 3.4, it is seen that for all the months considered, the members and ensemble systems over-forecast the lower threshold frequencies (< 10 mm/day).

The variability of forecast frequencies is the greatest during December, but the MMENS follows the frequency observed line from 10 mm/day to 30 mm/day threshold. The UM12 member over-forecast the frequency for all thresholds up to 50 mm/day after which the occurrence is zero. When only considering thresholds greater than 10 mm/day, November is considered to be the month when all of the members and ensemble systems have the best forecast.

The MMENS tends to under-forecast the frequency of occurrence for thresholds greater than 10 mm/day for four of the six months. During November through to March, the UM12DA member follows the observed frequency the best for thresholds greater than 10 mm/day. The UM12 member is the best for October, but for the months remaining the UM12 member over-forecast the frequency of threshold events. In January the UM12 member under-forecast the frequency of events greater than 10 mm/day amounts.

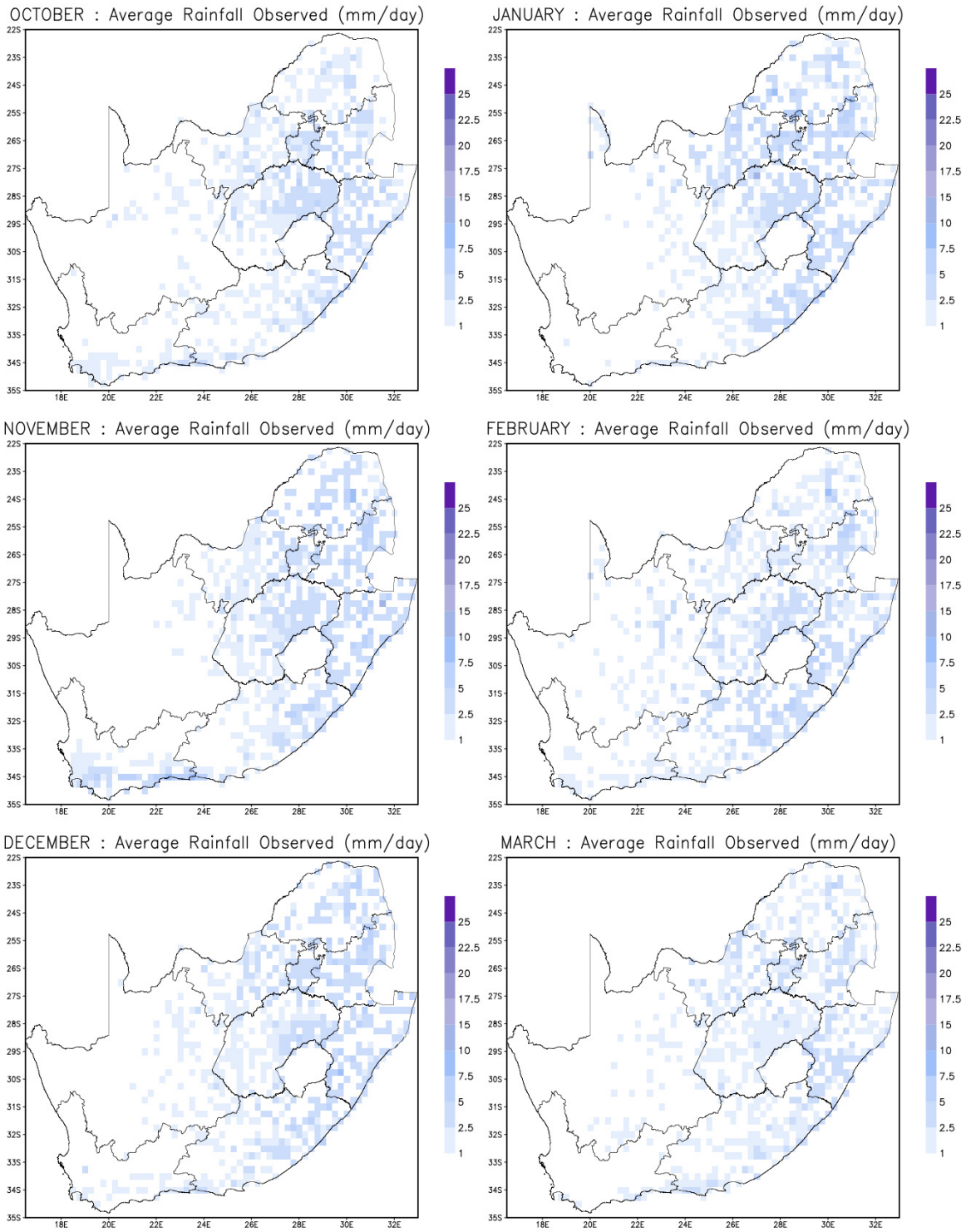


FIGURE 3.1: The average daily rainfall observed for each of the six months under investigation during the three summer half-years.

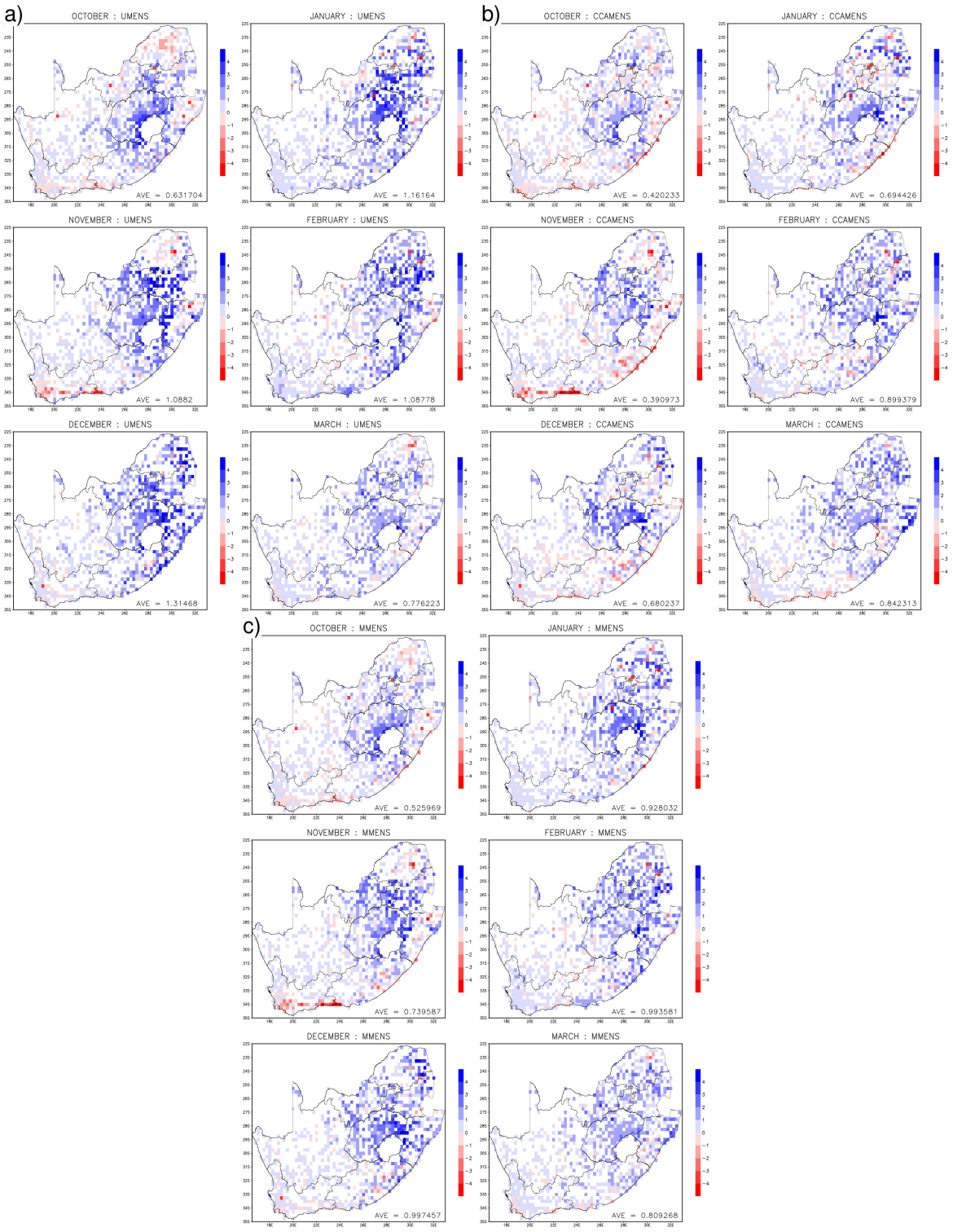


FIGURE 3.2: The spatial maps for the bias for the three ensembles for the six months under investigation. (a) UMENS, (b) CCAMENS and (c) MMENS

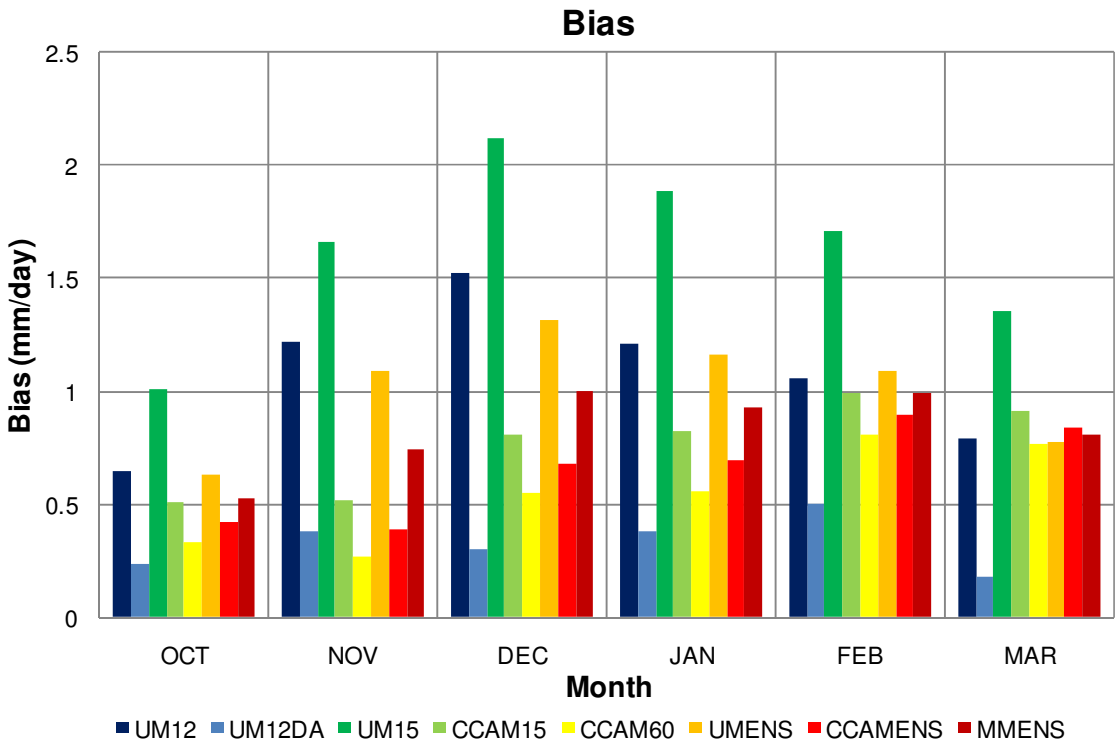
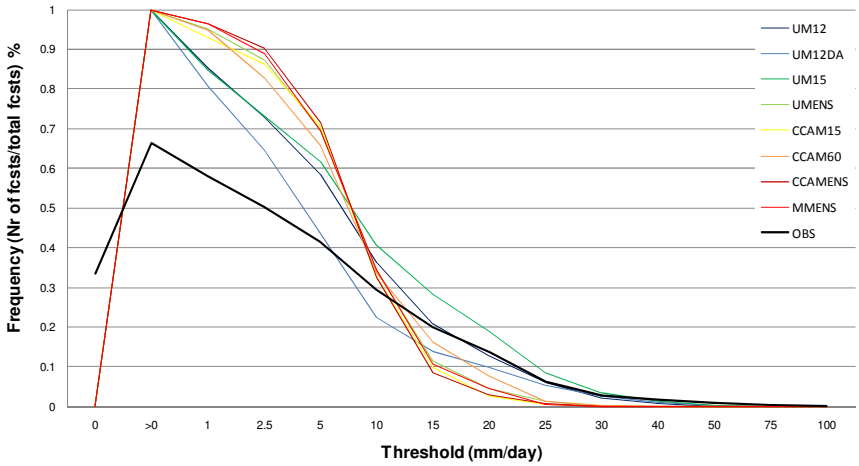
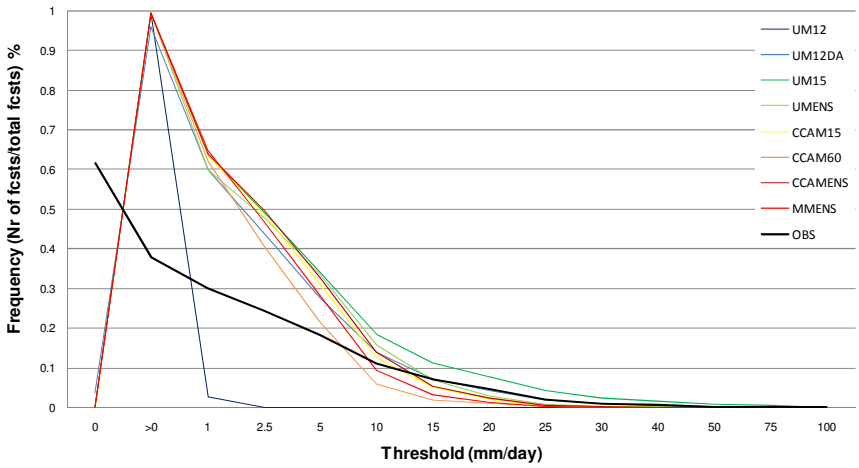


FIGURE 3.3: A summary of the bias for the members and the three ensemble systems for the six months under investigation

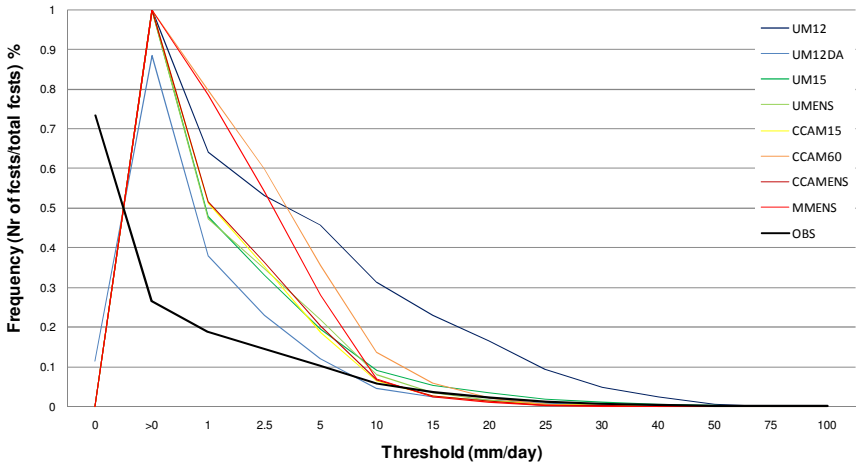
a). October



b). November



c). December



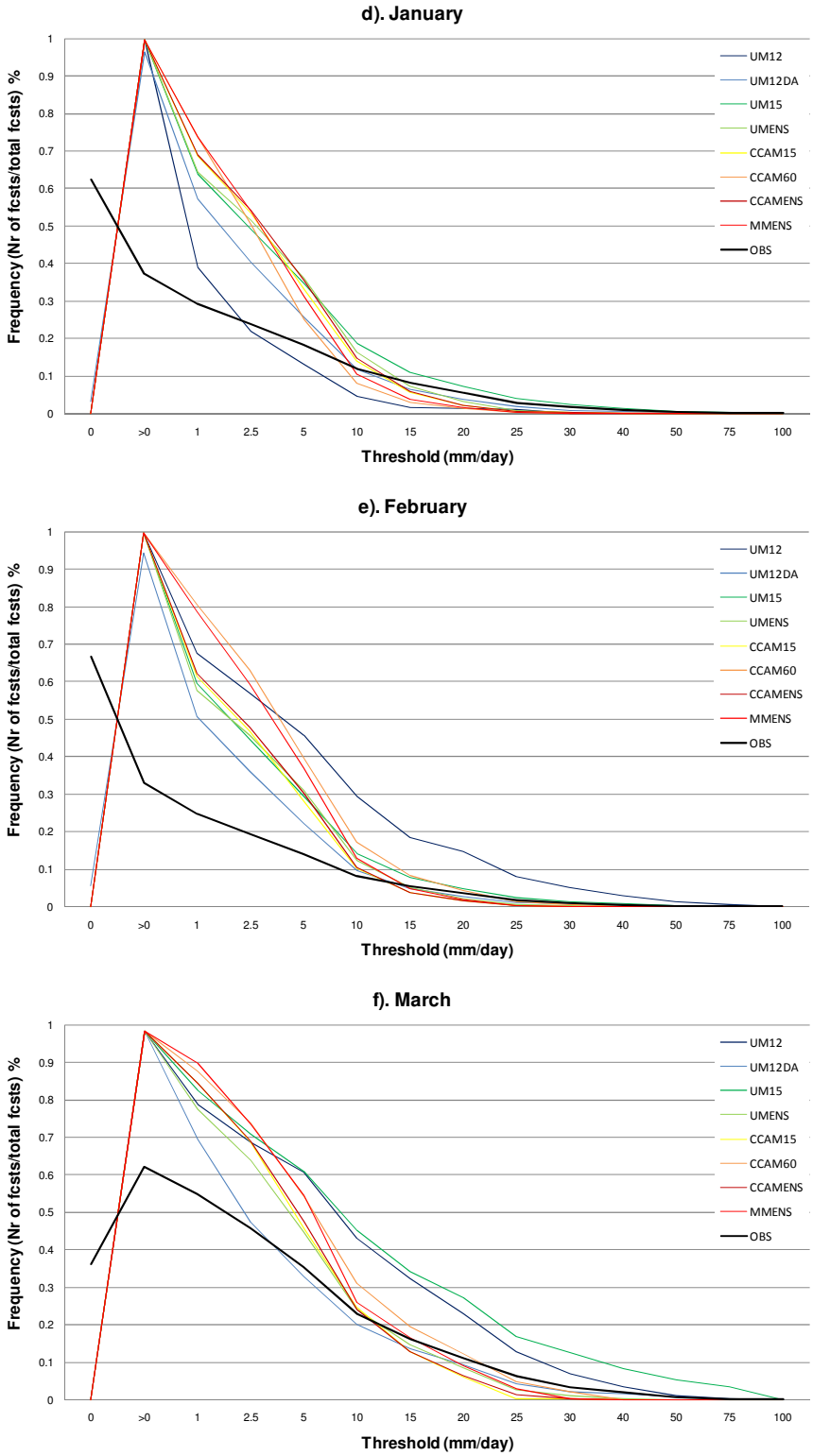


FIGURE 3.4: The frequency of occurrence for the six months considered. (a) October, (b) November, (c) December, (d) January, (e) February and (f) March.

3.2 VERIFICATION OF FORECAST SKILL

3.2.1 BRIER SKILL SCORE

The Brier skill score for the four threshold values are represented spatially by maps in Figure 3.5 to Figure 3.7 for the UMENS, CCAMENS and MMENS, respectively. The 1 mm/day threshold value is represented by (a) in the figures, (b) represents 10 mm/day, (c) and (d) shows the 25 mm/day and 50 mm/day thresholds. The results are shown on a monthly basis for the summer half-year, similar to the maps of the bias. On each of the maps, there is an indication of the percentage of positive grid points of the total number of grid box forecasts for the whole domain. This number gives an indication of the percentage of the forecast grid points over the domain that has skill over that of the persistence (reference) forecast. Therefore, the greater this number, the greater the skill of the forecast is for the period over the domain. In Figure 3.8 these numbers are visualized on a histogram for all the members of the ensemble. This graph is indicative of the relative performance of the various ensembles and ensemble members.

3.2.1.1 1 mm/day Threshold

In Figure 3.5a it can be seen that in general there is no skill for this threshold for the UMENS. This result may be interpreted as the UMENS not being skillful in predicting the occurrence, or non-occurrence of rainfall – at least over the summer rainfall region during summer. Skill is limited to the south-west coast and parts of the south coast, extending into the Eastern Cape Province. This result may indicate that the UMENS is skillful in predicting rainfall occurring in association with mid-latitude weather systems that move along the southern coastal areas during summer – but that is not skillful in predicting the occurrence of convective rainfall over the summer rainfall area. The reason for the forecasts not being skillful over the summer rainfall area may be linked to the general tendency of numerical models predicting too many rain days.

The CCAMENS Brier skill score verification is shown in Figure 3.6, where (a) again shows the 1 mm/day threshold. The same pattern as with the UMENS is evident in the CCAMENS. Areas of skill are visible over the south-west coast, Cape south coast and coastal areas to the north during certain months. With regard to the number of grid points with skill, the CCAMENS generally has lower skill than that of the UMENS.

Figure 3.7 displays the MMENS. For rainfall events of 1 mm/day and greater (map a), it is seen that the MMENS has no skill over most of the domain for all six months under consideration. This spatial distribution of areas with and without skill is similar to the skill distribution of the individual members of the ensemble. The south-west coast is again, as with the CCAMENS and UMENS a region with skill, whereas the Limpopo Province and the west coast have less skill during November and February respectively than with the two individual model ensembles. This is also the case over Mpumalanga during March. Considering Figure 3.8 it is seen that the MMENS has the least amount of positive (skillful) grid boxes for all six months considered, concluding that individual members outperforms the MMENS with this evaluation, because averaging in MMENS increases the likelihood of at least some rain. The UM 12 km data assimilation (UM12DA) member has the highest number of skillful grid boxes for the threshold value.

3.2.1.2 10 mm/day Threshold

The 10 mm/day threshold result is displayed in Figure 3.5b and it is seen that for this threshold the ensemble systems has overall greater skill than for 1 mm/day events. The months with the lowest skill are December through to February. During February however, the Free State Province is dominated by areas of skill for the 10 mm/day threshold. The same is true for the south and south-west coast during March. The skill is generally lower for the Limpopo Province as well as for parts of the central interior for several of the six months under investigation.

The spatial distribution of skill for the CCAMENS is similar to that of the UMENS, but considering Figure 3.8b it is seen that the CCAMENS generally has greater skill in predicting 10 mm/day rainfall events over that of the UMENS. The CCAMENS has skill over the Northern Cape Province, especially during December, as well as the Free State Province when the CCAMENS has skill over the region during January, as well as the north eastern parts of the country.

The influence of the combination of the UM and the CCAM ensembles into the MMENS is most evident over the south and south-west coast during March (Figure 3.7b). The region of skill over the south coast for the UMENS, in combination with the region with skill over the south-west coast of the CCAMENS produced a greater region with skill with the MMENS. The overall skill during all six months is improved with the MMENS for the 10 mm/day threshold over that of the two single-model

ensemble systems (Figure 3.8). However, during March and December the UM12DA member has greater skill than the MMENS. It is also evident that the members generally forecast with more skill the 10 mm/day threshold than any other of the thresholds.

3.2.1.3 25 mm/day Threshold

The Brier skill score for the 25 mm/day threshold is displayed in Figure 3.5c to 3.7c. The UMENS in Figure 3.5c in general has skill in predicting daily rainfall totals exceeding and equal to 25 mm, with the important exception that the forecasts are not skillful for the larger part of the domain during December and January. During these months, the absence of skill is particularly notable over the eastern Free State and western Kwa-Zulu Natal (KZN) area. The months with the best overall skill are October and March.

Figure 3.6c shows that the CCAMENS is outperforming the UMENS at the 25 mm/day threshold value. The greater part of the domain has skill for all the six months, with only the eastern parts of the country having regions with no skill during December, February and March.

The influence of the UMENS is evident in the MMENS. The MMENS skill is decreased especially over the Eastern Cape and KZN provinces during December. During January, the skill over the Free Sate is also decreased from that of the CCAMENS. This is evidence that an ensemble will not have skill over a region or for a period if one or more of the members is unskillful, especially with a small ensemble size.

For the 25 mm/day the MMENS is less skillful than the CCAM model members (Figure 3.8). Both the CCAM15 and CCAM60 forecasts have more positive grid points than the MMENS. Considering only the UM members, it is seen that the data assimilation model outperforms the other two UM members.

3.2.1.4 50 mm/day Threshold

This threshold represents the extreme rainfall events. It is clear from map (d) in Figure 3.5 to 3.7 that this threshold was not exceeded that often during the study period (in both observations and forecasts – Figure 2.8). However, it is also clear that when this threshold was exceeded, it mostly occurred over the northern and eastern parts of the country, with some events occurring over the south-coast. The Brier skill

score has the highest values during the months of October, February and March for the UMENS. The forecasts are not skillful over Mpumalanga during November.

The CCAMENS, as with the 25 mm/day threshold is more skillful than the UMENS (Figure 3.6d). The CCAMENS forecasts are skillful over Mpumalanga during November, with only the Free State Province having some areas in January where the forecasts are not skillful.

The extreme Brier skill score values are reduced in the MMENS from that of the UMENS. This is evident in Figure 3.7d over the eastern parts of the country. It is noted however that when looking at the area averages of the values, the values do not change much from one ensemble to the other. This might be ascribed that there are not as many extreme values in the individual model ensembles and therefore the influence of one model ensemble over the other is not great in the MMENS.

In Figure 3.8 it is noted that for the 50 mm/day threshold the MMENS and CCAMENS skill is comparable. It is also noted however that the UM 15 km member had the lowest percentage of positive grid boxes for all six months.

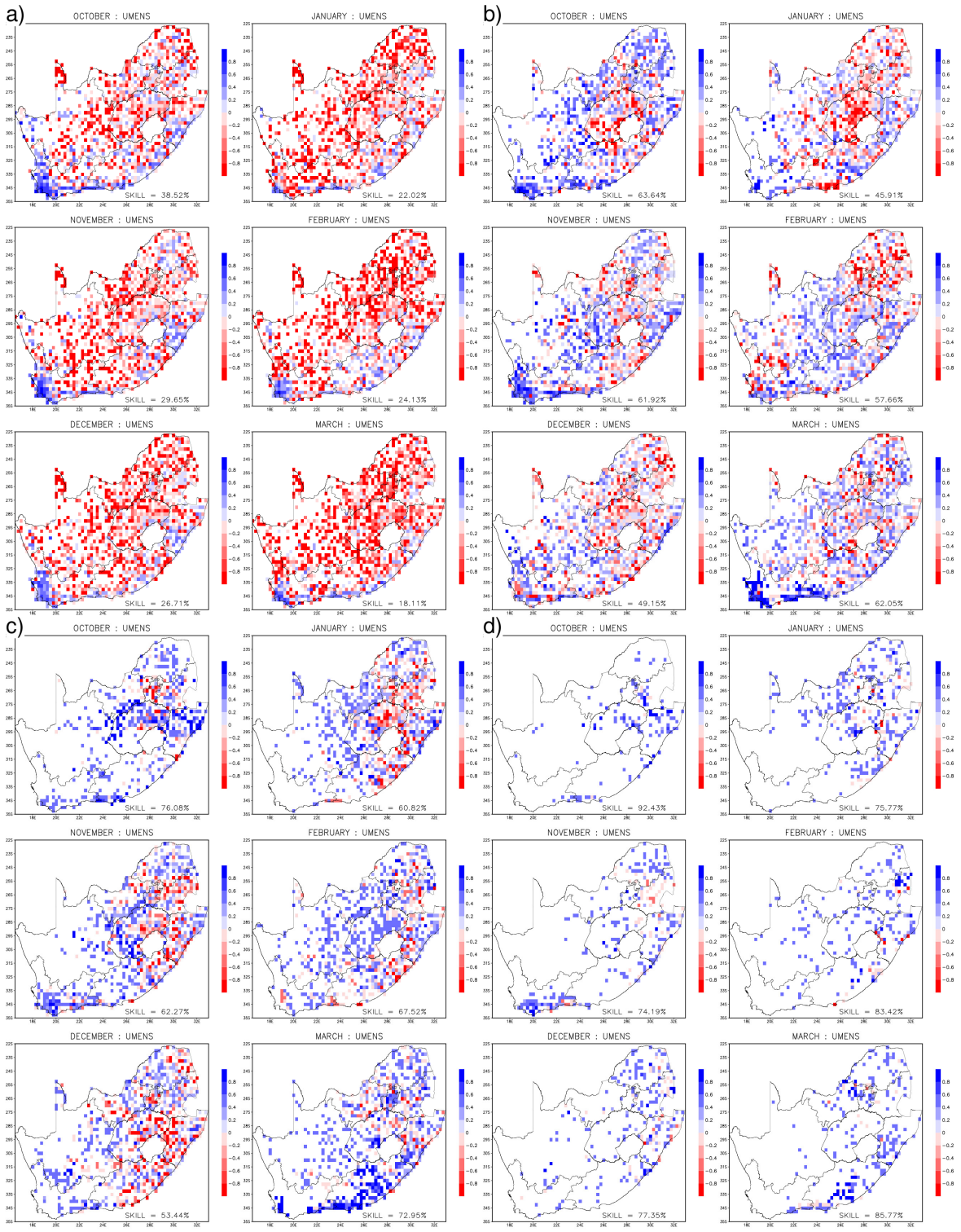


FIGURE 3.5: Spatial maps for the BSS for the UMENS system for the four thresholds for all six months. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

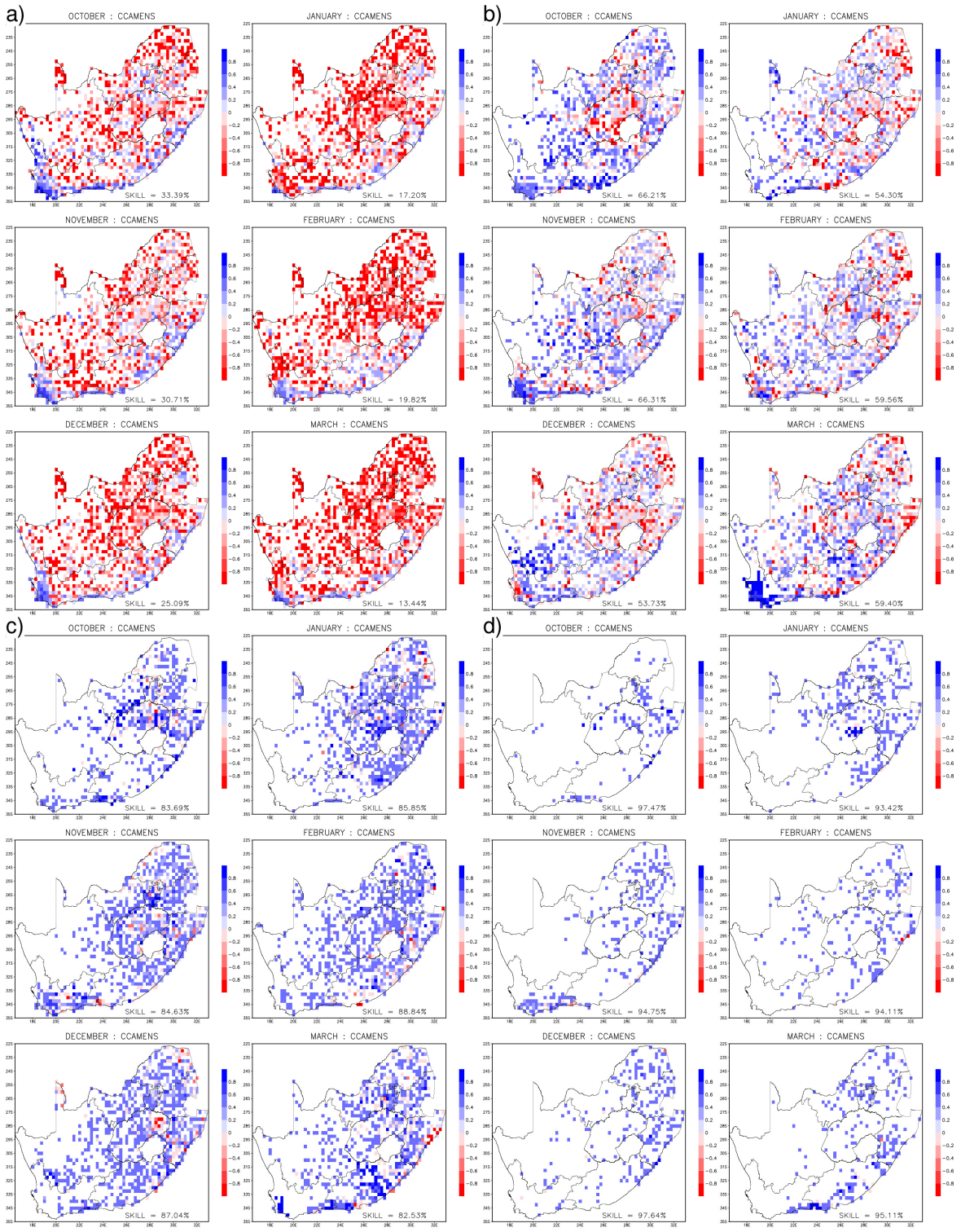


FIGURE 3.6: Spatial maps for the BSS for the CCAMENS for the four thresholds for all six months (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

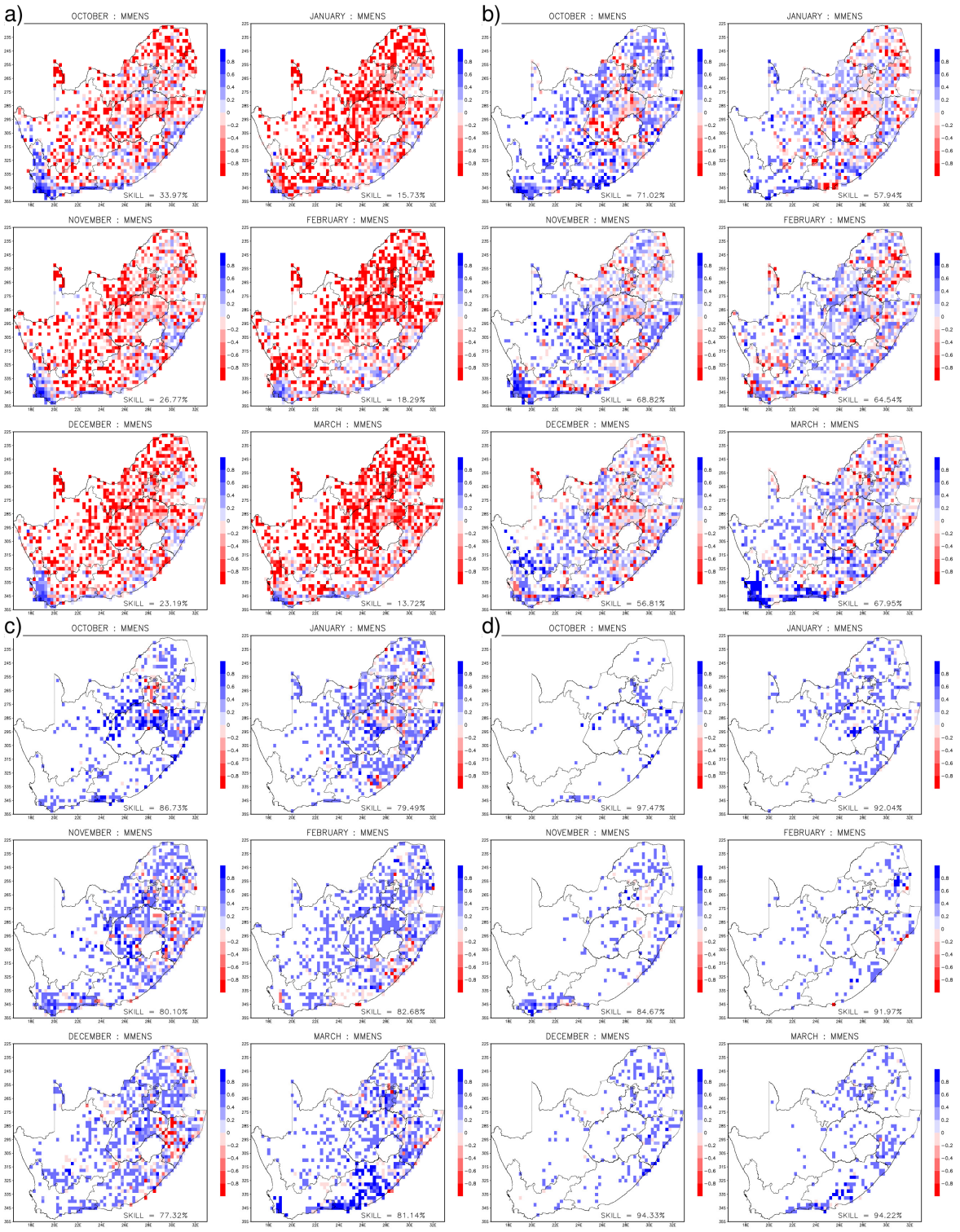


FIGURE 3.7: Spatial maps for the BSS for the MMENS for the four thresholds for all six months (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

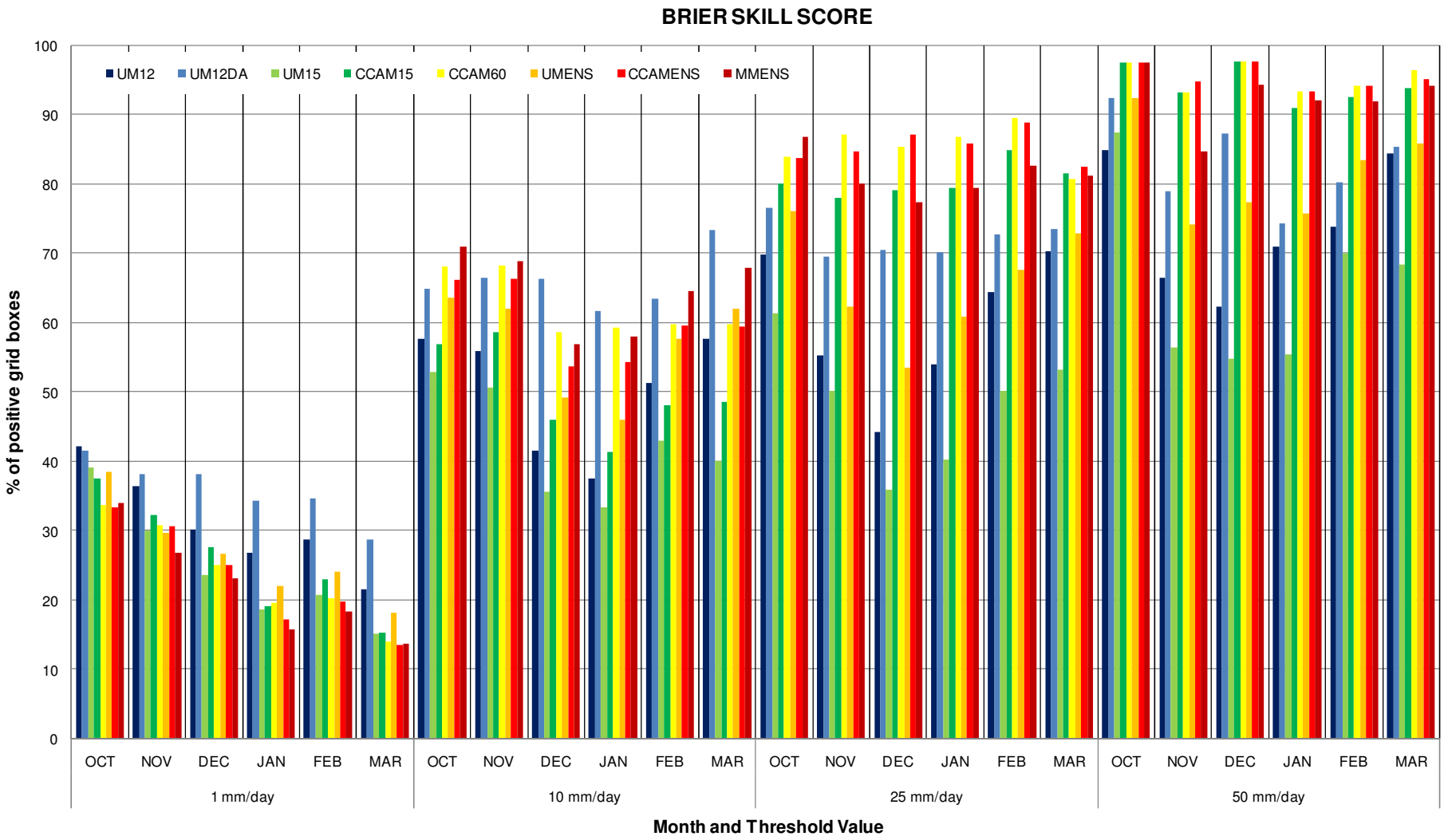


FIGURE 3.8 A summary of the percentage of positive BSS grid boxes for the four threshold values per member for the six months under investigation

3.2.2 MEAN SQUARED ERROR SKILL SCORE

In Figure 3.9 the mean squared skill score is represented in the same spatial format as discussed above for the Brier skill score, and was calculated after rescaling all forecasts to a common grid of resolution 0.25°. A positive skill score is a sign that the forecast is outperforming the reference skill. In each of these figures, map (a) represents the UMENS, map (b) the CCAMENS and map (c) the MMENS.

Figure 3.9a reveals that the UMENS in general produces skillful forecasts of rainfall totals, with the exception being an area over central South Africa (including Gauteng). The south-west coast has skill over persistence over the entire period. The south-coast has no skill during the months of January and February but is skillful the rest of the period. For the north-eastern parts of the country the skill varies from one month to the other, no pattern is distinguishable over this region.

Considering Figure 3.9b it can be seen that the CCAMENS provides generally skillful forecasts, with areas along the west-coast and the Limpopo Province being the exceptions. The CCAMENS outscores the UMENS for all months (considering the total number of positive grid boxes).

The overall skill of the domain is improved over that of the UMENS and also some regions of the CCAMENS, such as the Limpopo Province during November and the Western Cape in December and Gauteng during March. In Figure 3.10, a summary of the percentage of positive grid boxes per month per ensemble is given. In this graph it can be seen that MMENS only outperforms the single-model ensemble system during October and March. During October all of the members are the most skillful and January (the month with the highest rainfall in the summer rainfall season) the least. For all six months under investigation it is seen that the UM 15 km (UM15) member is the least skillful member. The UM12DA member is continuously the best member of the UMENS and the CCAM 60 km (CCAM60) member the best for the CCAMENS system.

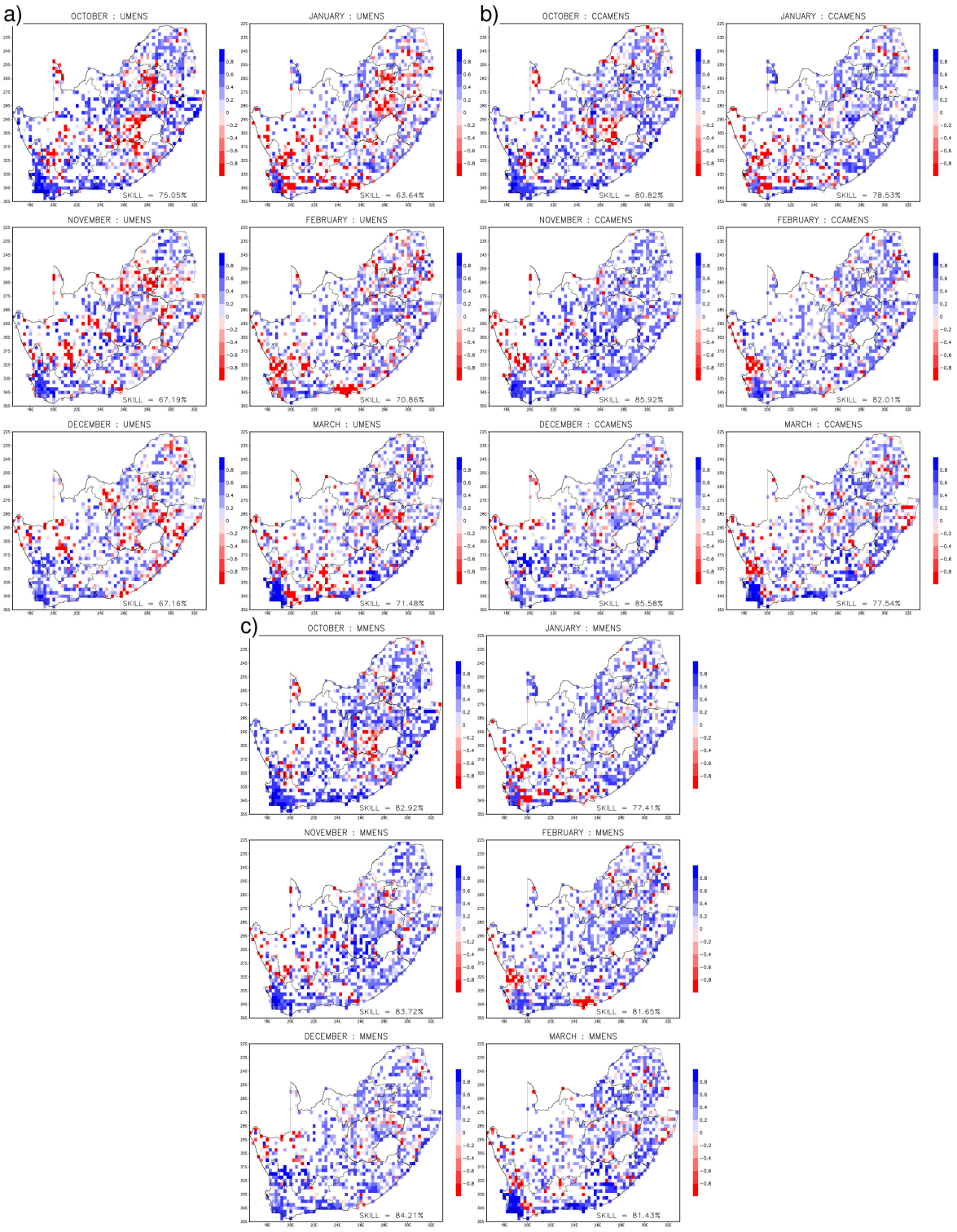


FIGURE 3.9: Spatial maps for the MSESS. (a) UMENS, (b) CCAMENS and (c) MMENS.

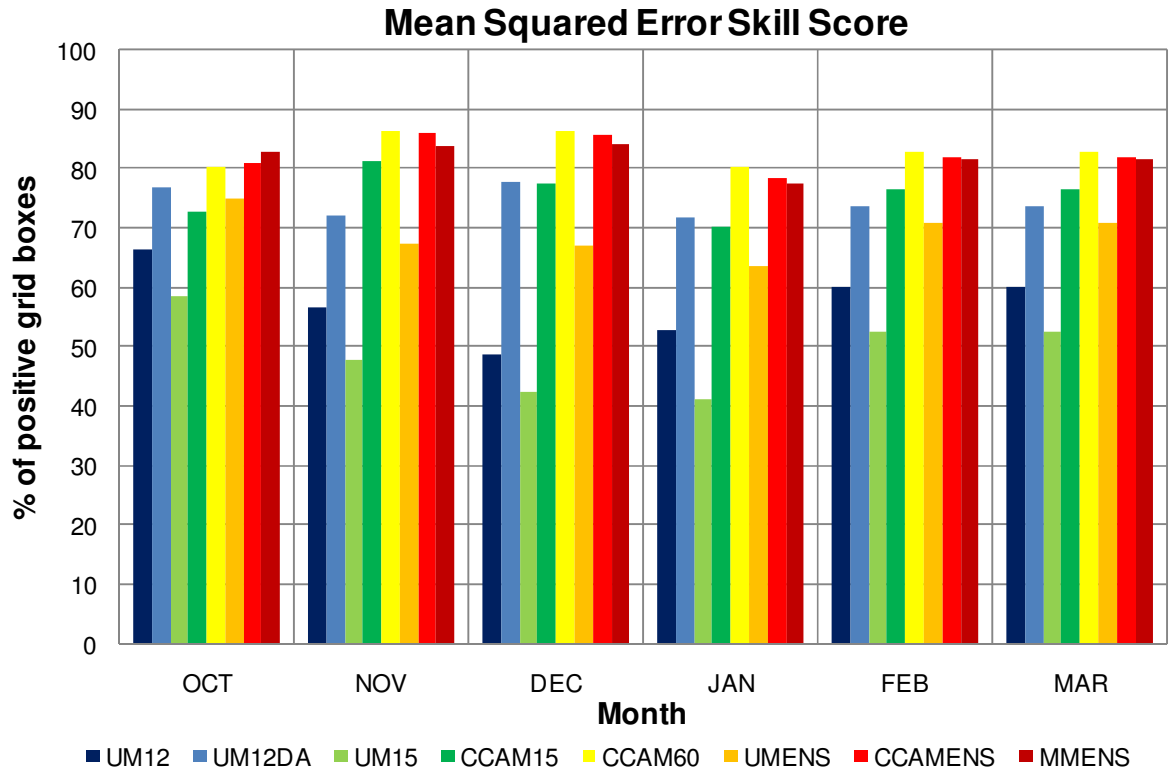


FIGURE 3.10: A summary of the MSESS for each of the members for the six months under investigation.

3.3 ADDITIONAL VERIFICATION OF DICHOTOMOUS FORECASTS

In this section, only the MMENS maps will be shown. The single-model ensembles will be referred to and discussed and summarised in the accompanying histograms. In all the figures map (a) shows verification results for the threshold of 1 mm/day, map (b) for 10 mm/day, (c) for 25 mm/day and map (d) for 50 mm/day.

3.3.1 FREQUENCY BIAS INDEX

3.3.1.1 1 mm/day Threshold

The FBI spatial distribution maps for the MMENS are shown in Figure 3.11. As discussed in Chapter 2, the perfect forecast will have a FBI of one. It is seen in Figure 3.11 that the MMENS system over-forecast the frequency of occurrence of events greater than 1 mm/day threshold. There are however, small regions to the south-west that have under-forecast characteristics. The same spatial pattern and over-forecasting is present for the UMENS and CCAMENS (not shown).

3.3.1.2 10 mm/day Threshold

Considering Figure 3.11b, it is seen that MMENS system has a tendency of over-forecast the frequency of events above the 10 mm/day threshold over the eastern part of the domain. This is true in particular for the eastern coastal areas, eastern escarpment and central interior. It is interesting to note that over the Limpopo Province in northeastern South Africa, there is a pattern of the frequency of these events to be over-forecast for all months, except for October (the start of the rainy season), when the frequency is under-forecast. Over the western interior, the frequency of occurrence of events of 10 mm/day and higher is under-forecast.

3.3.1.3 25 mm/day Threshold

Contrary to the 1 mm/day threshold events, the MMENS in general under-forecasts the frequency of events occurring above the 25 mm/day threshold. The over-forecast of such events is confined to the eastern escarpment and parts.

3.3.1.4 50 mm/day Threshold

These events do not occur often, and the MMENS under-forecasts across the domain. The only exception to this statement is the over-forecast of events above the 50 mm/day threshold along the Cape south-coast in March. This pattern of over-forecast stems from the UMENS, with the CCAMENS in fact under-forecast the frequency of 50 mm/day events over the region.

Figure 3.12 indicates the misses for the 50 mm/day threshold events for October for the MMENS. The MMENS never over this period predicted rainfall amounts greater than 50 mm/day.

Considering Figure 3.13 which shows the area average FBI value on the y-axis and the six months under investigations and the four threshold values on the x-axis. This figure summarizes a result that is evident from the spatial maps of the FBI shown earlier: there is a general tendency for the ensemble members and different ensemble systems to over-forecast the frequency of occurrence of rainfall (events above the 1 mm/day threshold), but extreme rainfall events (25 mm/day threshold and higher) are in generally under-forecast. This is an important result for operational weather forecasts to take note of – the MMENS and individual models too often forecast the occurrence of rain, and may not provide sufficient guidance on the possible occurrence of extreme rainfall events.

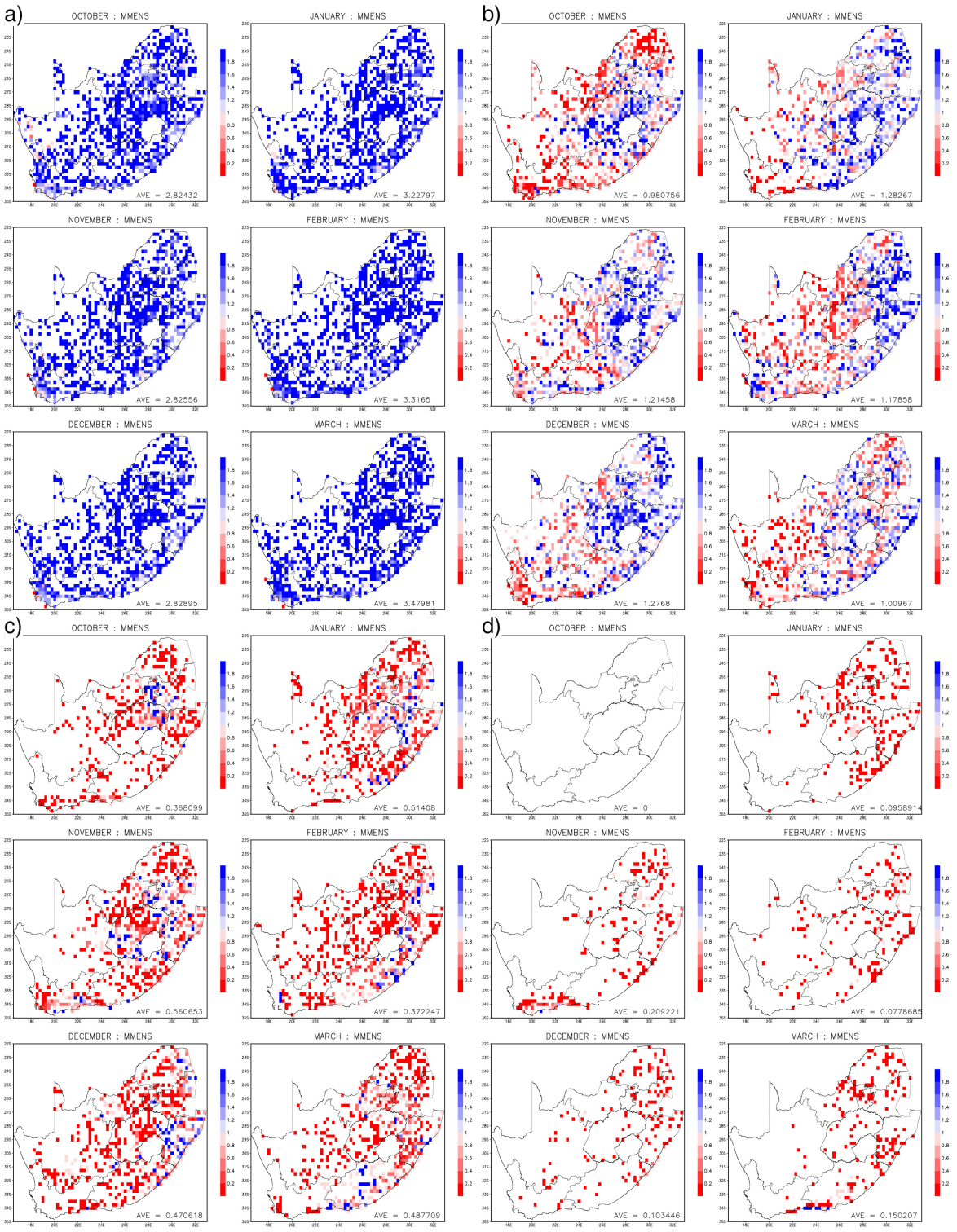


FIGURE 3.11: The FBI for the MMENS system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

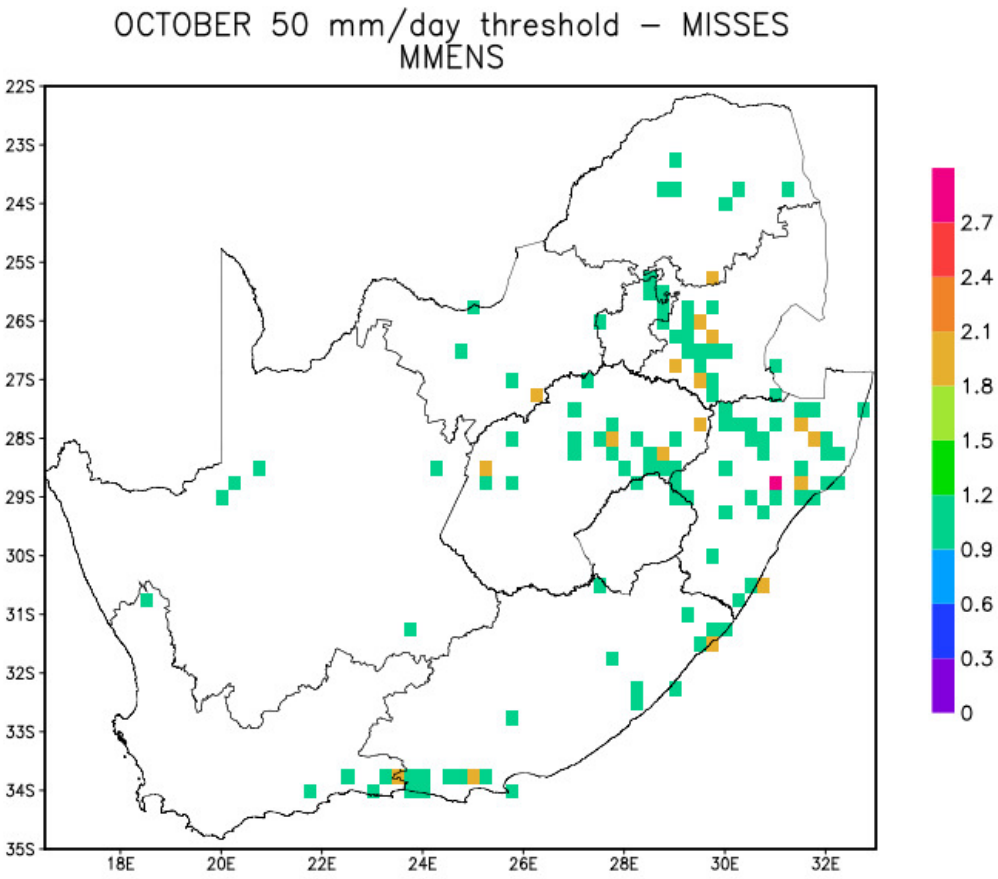


FIGURE 3.12: Misses for 50 mm/day events for October for the MMENS

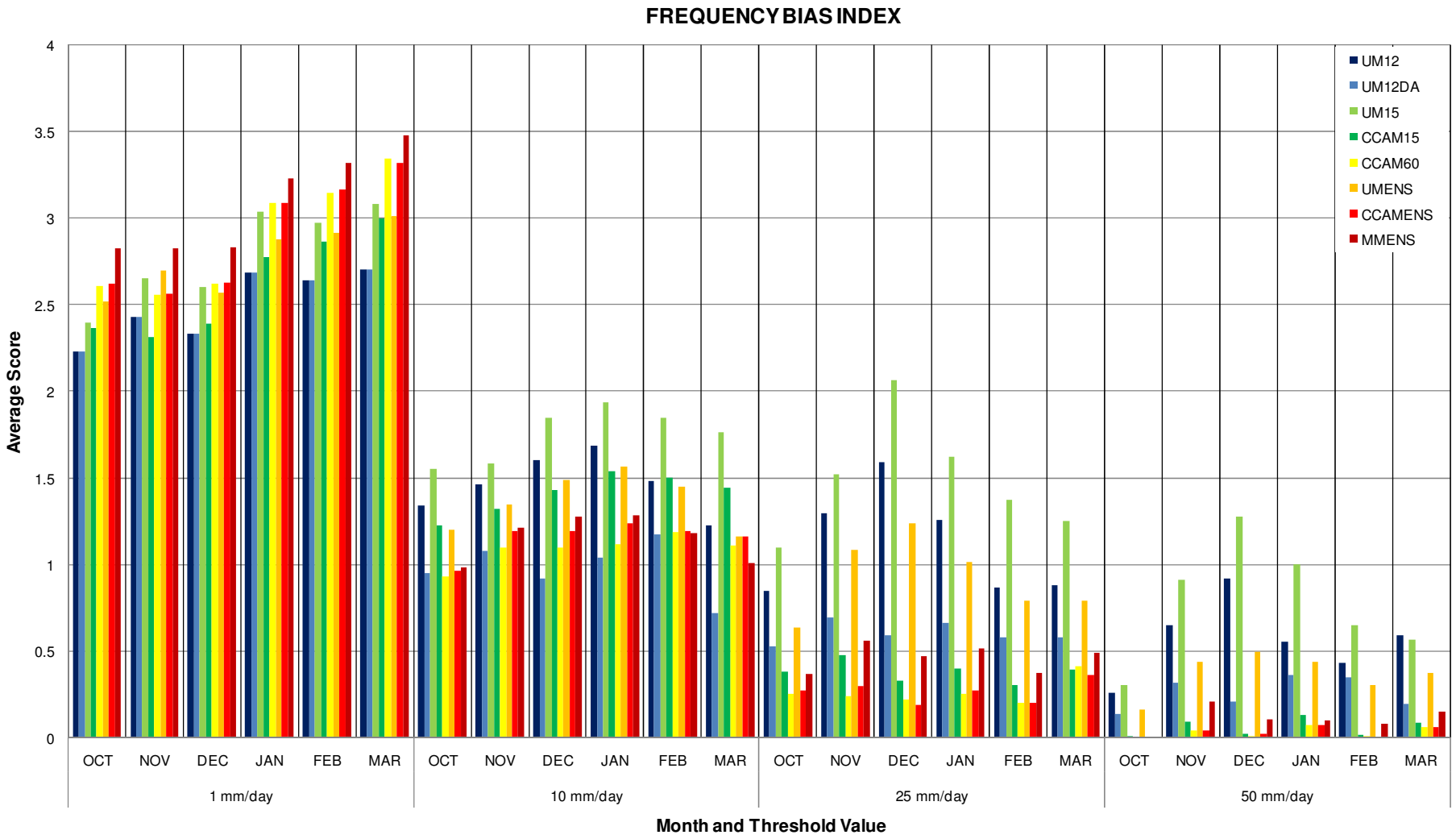


FIGURE 3.13: The FBI summary for the ensemble systems and members for the four threshold values and six months under investigation.

3.3.2 PROBABILITY OF DETECTION

The POD spatial distribution maps for the MMENS are shown in Figure 3.14.

3.3.2.1 1 mm/day Threshold

In Figure 3.14a it can be seen that the POD of exceedance of the 1 mm/day threshold is high over the eastern parts of the country as well as over the southwestern Cape. The lowest values occur over the western interior. The generally high POD values should be seen within the context of the MMENS over-forecasting the frequency of occurrence of events above the 1 mm/day threshold, so that the high POD values are accompanied with high false alarm rates (see the next section).

When comparing the area average of POD values of the UMENS to that of the CCAMENS for October (see Figure 3.15), the UMENS has a higher value. This is true for all the months except for February and March, when the CCAMENS has a greater probability of detecting the occurrence of events above the 1 mm/day threshold. It is also seen in Figure 3.15 that the MMENS has the highest POD value for all members and ensemble systems considered for all six months.

3.3.2.2 10 mm/day Threshold

There is a noticeable decrease in POD values for the case of threshold events of 1 mm/day, compared to threshold events of 10 mm/day (Figure 3.14b). This decrease is due to the fact that there are fewer occurrences of rainfall events exceeding 10 mm/day, and when such events do occur, the models have a greater tendency to miss them than for the case of the 1 mm/day threshold. POD values for the threshold values of 10 mm/day are relatively high for the month of March over the southwestern Cape and the Cape south-coast (Figure 3.14b). This may be due to the models sufficiently simulation the rainfall events occurring over this region in March in association with cut-off lows (e.g. Singleton and Reason, 2007).

3.3.2.3 25 mm/day Threshold

Rainfall events of 25 mm/day or higher are confined to the east-coast and eastern escarpment, and central and Eastern Cape interior and Cape south-coast during the summer half-year (see Figure 3.1). For the MMENS, the POD is highest for these events over the central interior during October and south-coast and Eastern Cape during March. For the rest of the period and domain only fractions of 0.5 or less of the 25 mm/day were detected.

Considering Figure 3.15 it is seen that the CCAMENS POD values are lower than that of the MMENS for all months, with the MMENS values lower than those of the UMENS for all months. The relatively high values of the UMENS in turn result from the high POD obtained by the UM15 and UM12DA ensemble members. Note that high POD values may be obtained through the persistent over-forecasting of the event, and is not an indication of the ensemble member or ensemble to distinguish between events and non-events.

3.3.2.4 50 mm/day Threshold

Rainfall events of 50 mm/day occurring within 24-hours (on the average over a grid box of 0.25º by 0.25º) are rare and largely limited to areas along the Cape south coast during the summer half-year (Figure 2.8). The POD of such events is generally low for the MMENS, with values being relatively high for the UMENS and relatively low for the CCAMENS. One implication of this result, is that the numerical forecasts cannot be used to provide explicit warning of the possible occurrence of extreme rainfall events

During both October and February, the CCAMENS POD is zero (see Figure 3.15). The same can be said for the MMENS during October (Figure 3.14c). However, during the rest of the months there are small areas of fractions greater than 0.8 for the occurrence to have happened and been forecast.

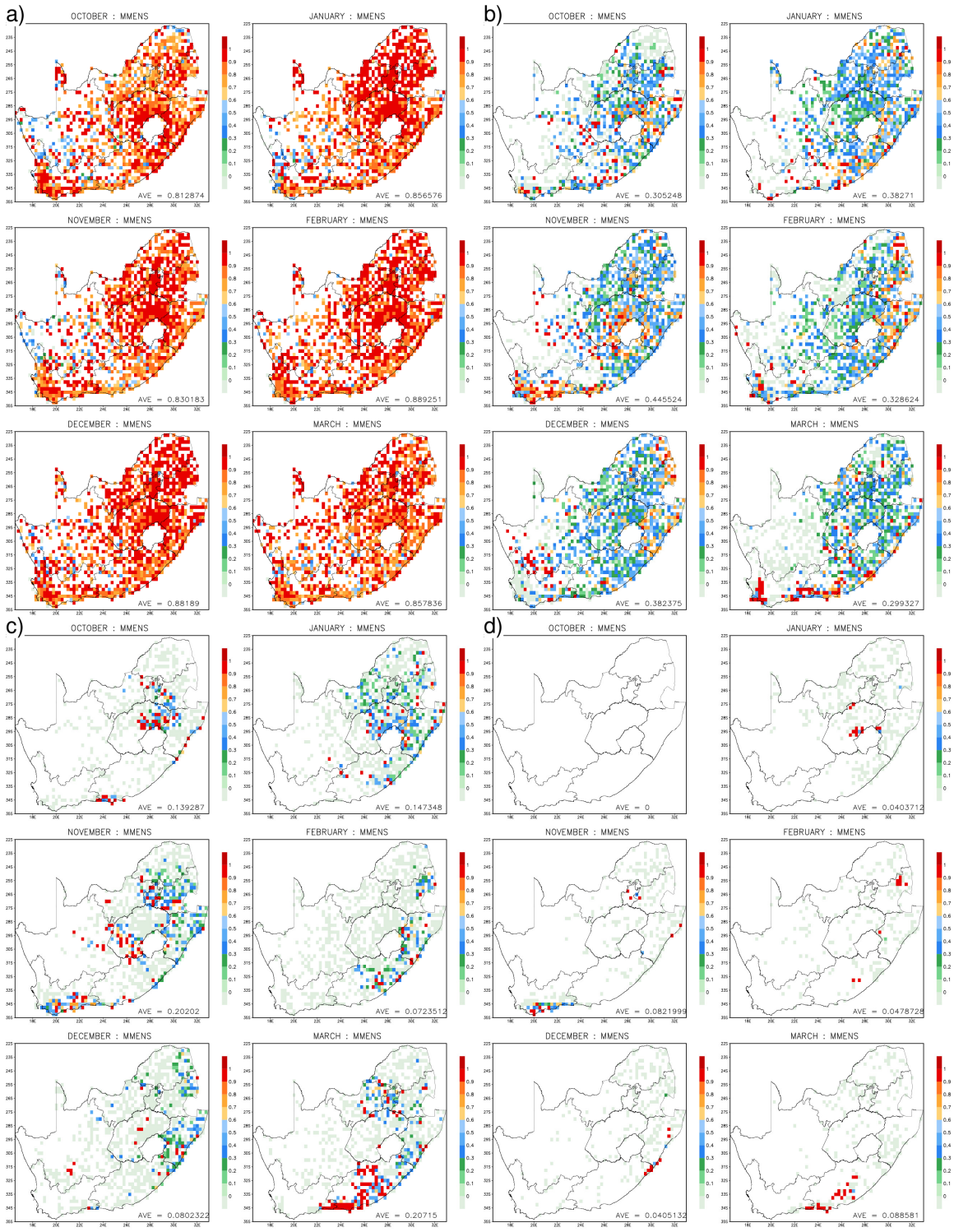


FIGURE 3.14: POD for the MMENS system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

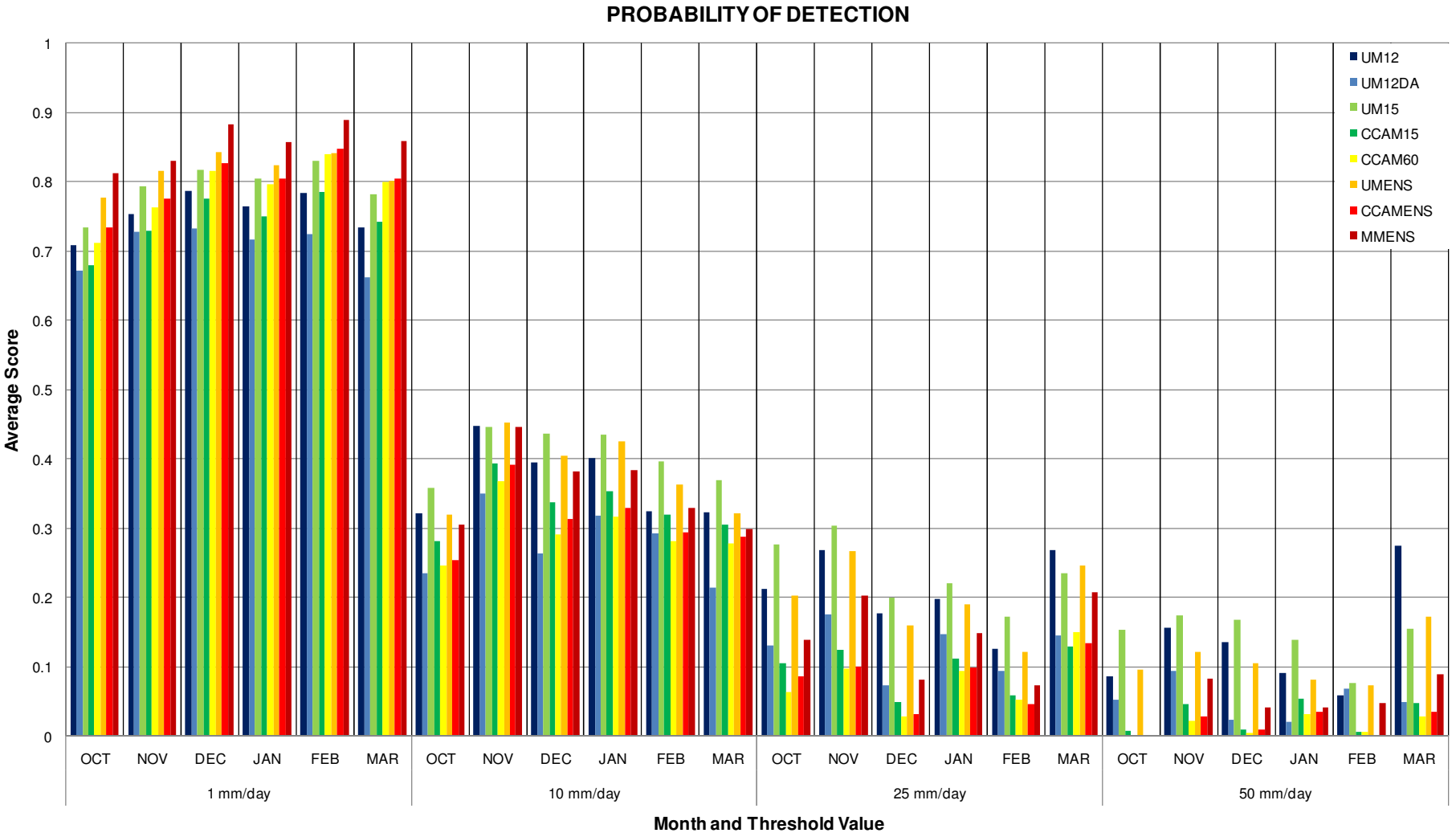


FIGURE 3.15: POD for the ensemble systems and members for the four threshold values and six months under investigation.

3.3.3 FALSE ALARM RATE

The F spatial distribution maps for the MMENS are shown in Figure 3.16. The F spatial distribution maps for the UMENS and CCAMENS are not shown.

3.3.3.1 1 mm/day Threshold

The F of the MMENS in forecasting the mere occurrence of rainfall (events of 1 mm/day) and larger is high (about 0.5 or higher over large portions of eastern South Africa – see Figure 3.16a). F values in excess of 0.8 are recorded for parts of Lesotho and the KZN Province during January and February. High F values are also seen over the Mpumalanga Province during November. These results confirm earlier deductions that the UMENS forecast too many rain days, and cannot be used as a reliable indicator of the possibility of occurrence of rainfall. All three models have the lowest false alarm rate during October (onset of the rainy season). The F is higher than for the UMENS and CCAMES, as well as for the individual ensemble members, for all months (Figure 3.17). This stems from the threshold of 1 mm/day being exceeded in the MMENS when only one or some of the constituting forecasts indicate exceedance of this limit.

3.3.3.2 10 mm/day Threshold

There is a noticeable decrease in the F for the 10 mm/day threshold compared to the 1 mm/day threshold, for all three ensembles (Figure 3.16b). For most months, the F is lower for the CCAMENS than for the UMENS, consistent with the generally higher POD values of the UMENS recorded for this threshold.

3.3.3.3 25 mm/day Threshold

The F is small for the 25 mm/day threshold for all six months for all three ensembles (Figure 3.16c). It may be noted that all the ensembles systematically under-forecasts the occurrence of 25 mm/day events, leading to both the F and POD values being low.

3.3.3.4 50 mm/day Threshold

The F is small for the 50 mm/day threshold for all six months for all three ensembles (Figure 3.16d).

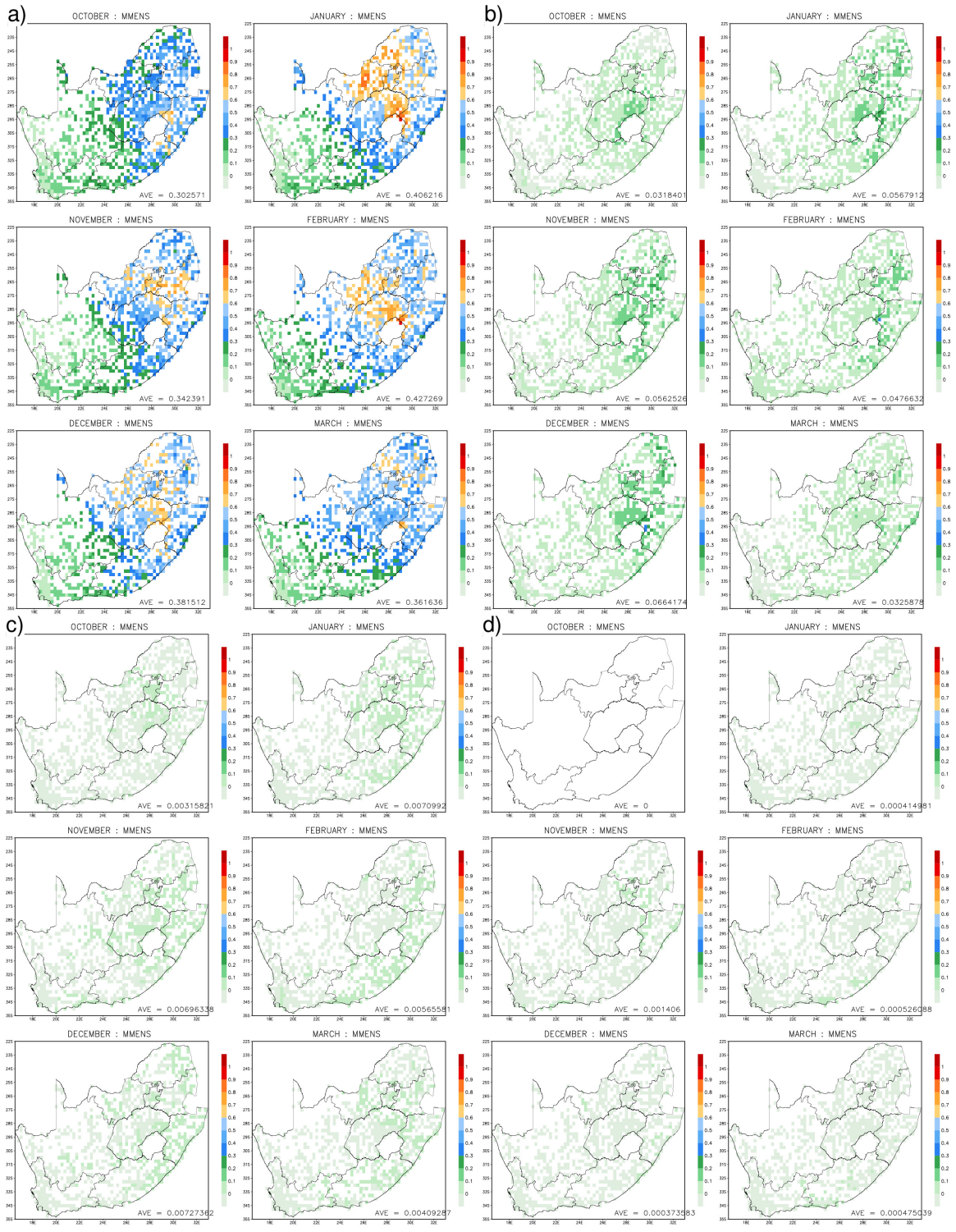


FIGURE 3.16: F for the MMENS system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

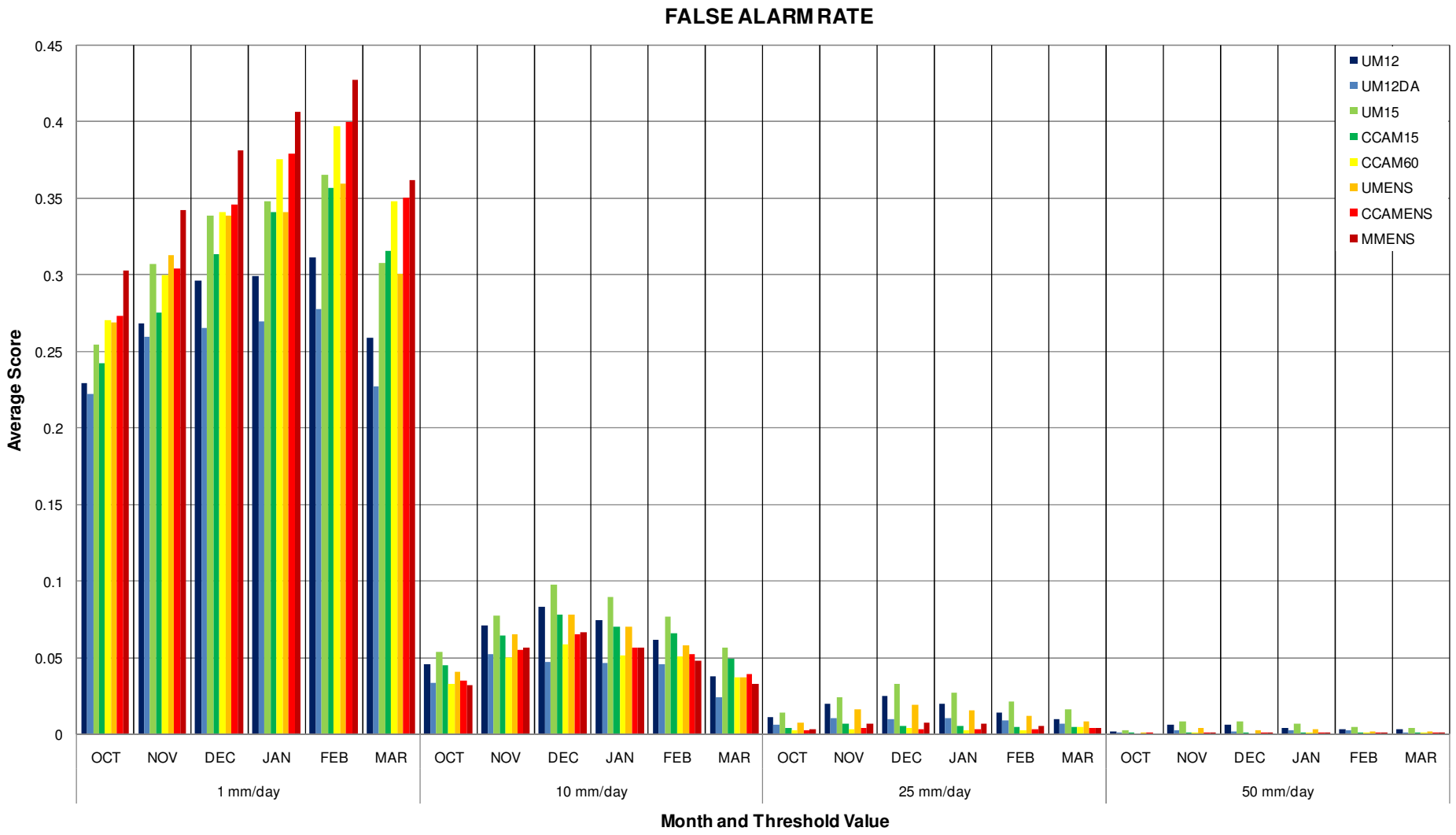


FIGURE 3.17: F for the members of the MMENS System for all six months under investigation.

3.3.4 CRITICAL SUCCESS INDEX

The CSI spatial distribution maps are shown in Figure 3.18.

3.3.4.1 1 mm/day Threshold

The CSI for all three ensembles rarely exceeds 0.7. This result is consistent with the generally high F (incorrect “yes” forecasts). The UMENS consistently obtains the highest CSI values for this threshold, compared to the other two ensembles (Figure 3.19).

3.3.4.2 10 mm/day Threshold

CSI values are similar but in general somewhat lower for the 10 mm/day threshold than for the 1 mm/day threshold. It is noteworthy that for most months, the MMENS obtains higher CSI values than the constituting UMENS and CCAMENS ensembles.

3.3.4.3 25 mm/day Threshold

CSI values are relatively low for the 25 mm/day threshold, compared to the smaller thresholds. This results from the MMENS in general under-forecasting the frequency of occurrence of extreme rainfall events. Noteworthy is the UMENS persistently outscoring the CCAMENS in this respect, with the MMENS values of a similar magnitude than those of the UMENS (Figure 3.19).

3.3.4.4 50 mm/day Threshold

Almost all of the occurrences over the study period where the 50 mm/day event was forecast had a CSI value greater than 0.7, except for the south-coast during November. It is however again noted that for most of the grid box where 50 mm/day was observed, the MMENS missed the event.

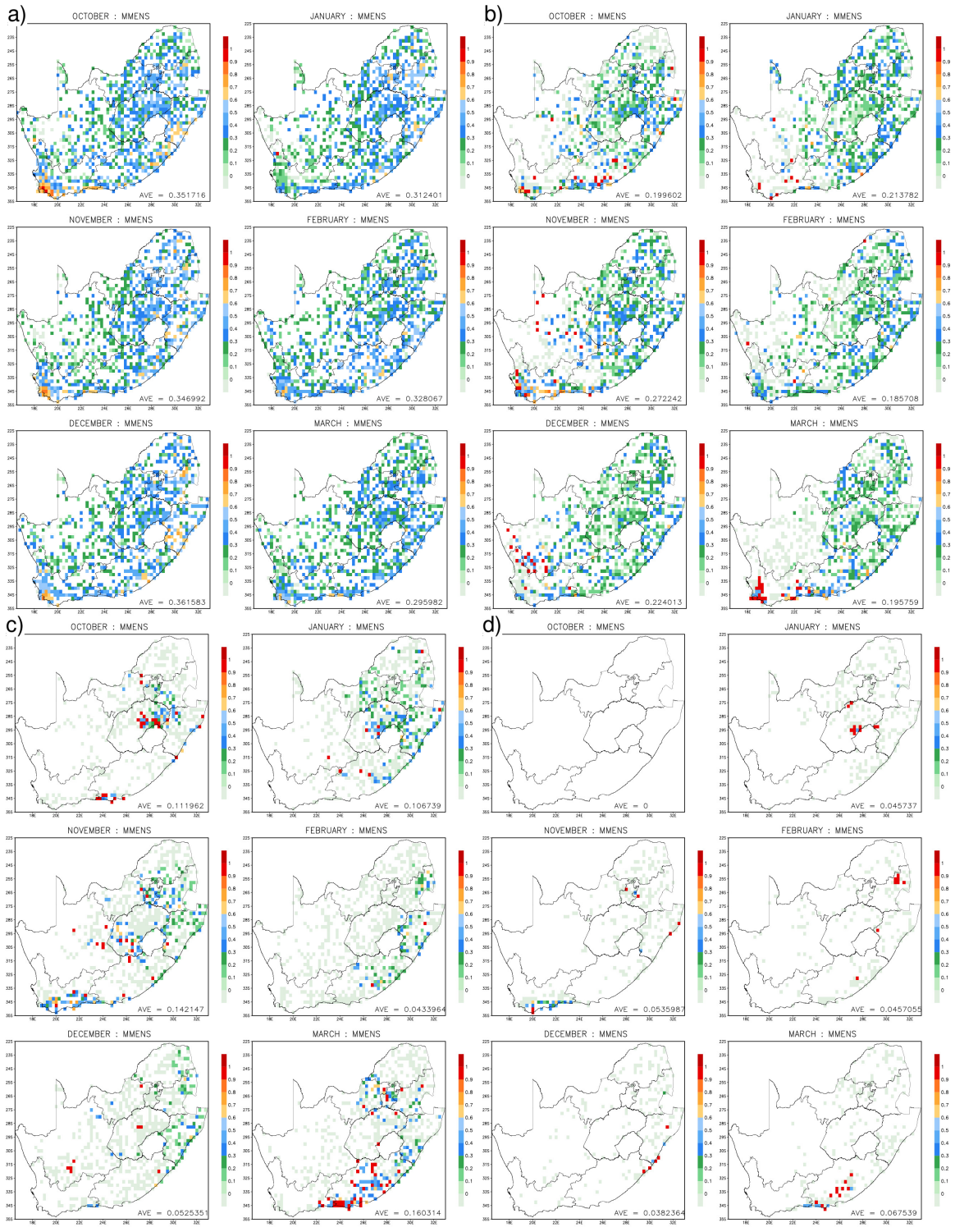


FIGURE 3.18: CSI for the MMENS system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

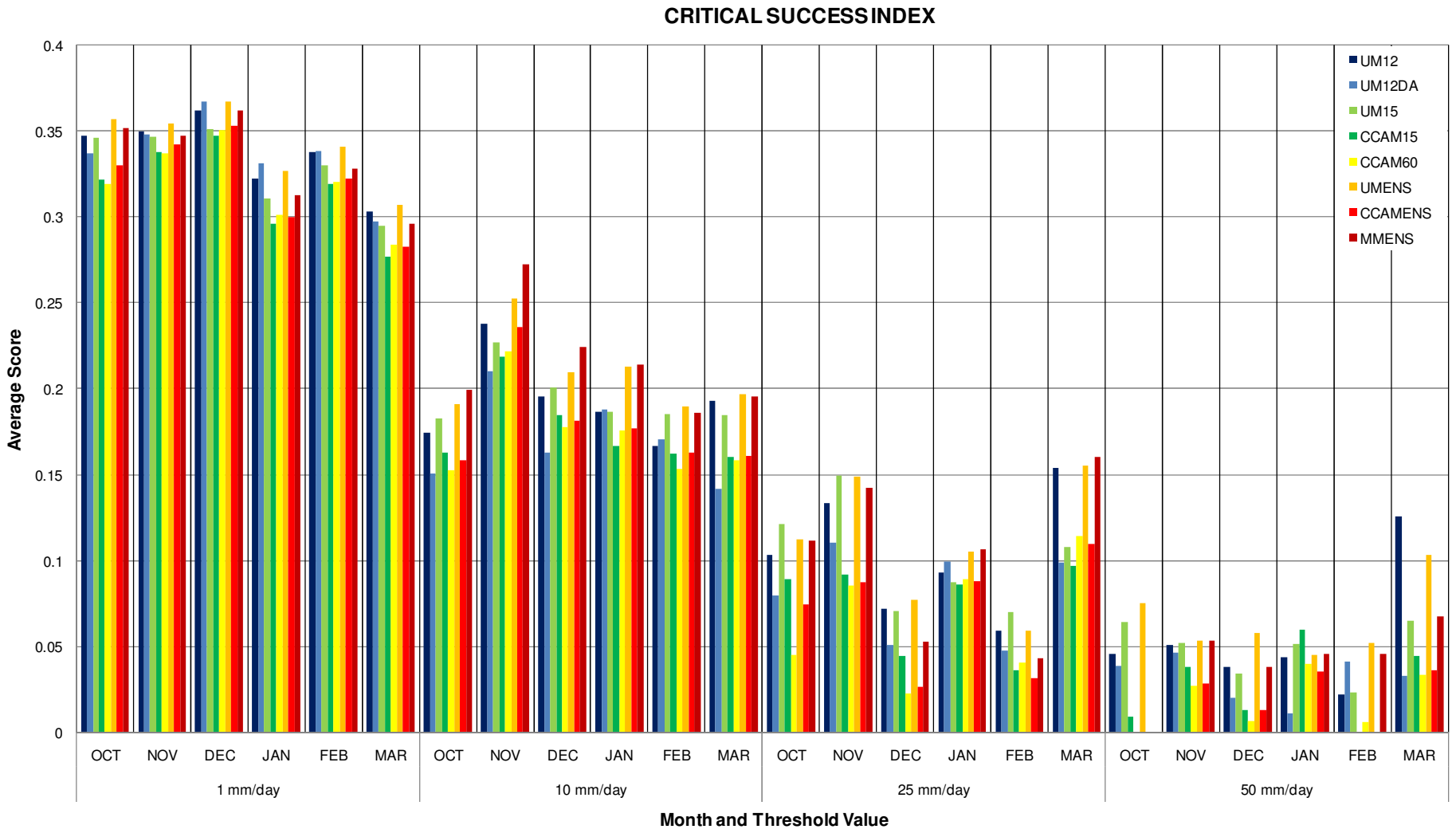


FIGURE 3.19: CSI for the members of the MMENS System for all six months under investigation.

3.3.5 EQUITABLE THREAT SCORE

The ETS spatial distribution maps are shown in Figure 3.20.

3.3.5.1 1 mm/day Threshold

The ETS in general display positive values, for all six months across the study period, implying that the forecast is skillful in predicting rainfall events above this threshold and corresponded well with the observed occurrences. There are only isolated regions where the ETS values are negative, especially over the Northern Cape the first half of the season and the northern parts of the country the second half.

3.3.5.2 10 mm/day Threshold

For this threshold, it is seen that the system is not skillful over the Northern Cape Province, which is generally dry during this period. The ETS is corrected for hits due to random chance and therefore the lower skill over the drier climate regions with very few rainfall stations. The system however performs well over the northern and central regions of the country, especially over the Free State and Gauteng Provinces. The MMENS also has skill over the Mpumalanga and KZN Provinces.

3.3.5.3 25 mm/day Threshold

The skill of the MMENS system is drastically less for this threshold. Regions of skill are limited to areas in the east that do receive a relatively large number of extreme rainfall events. The UMENS did considerably better than the other two systems when looking at Figure 3.21.

3.3.5.4 50 mm/day Threshold

Not one of the three systems had great skill for this threshold. The CCAMENS had no positive skill across the domain for February. The UMENS offers some improvement over the CCAMENS. In Figure 3.21 it is seen that the percentage of positive number of ETS grid boxes (number of boxes having skill) for all the members are very similar. It is however noted that the MMENS only had a higher percentage of skillful grid boxes with the 1 mm/day threshold

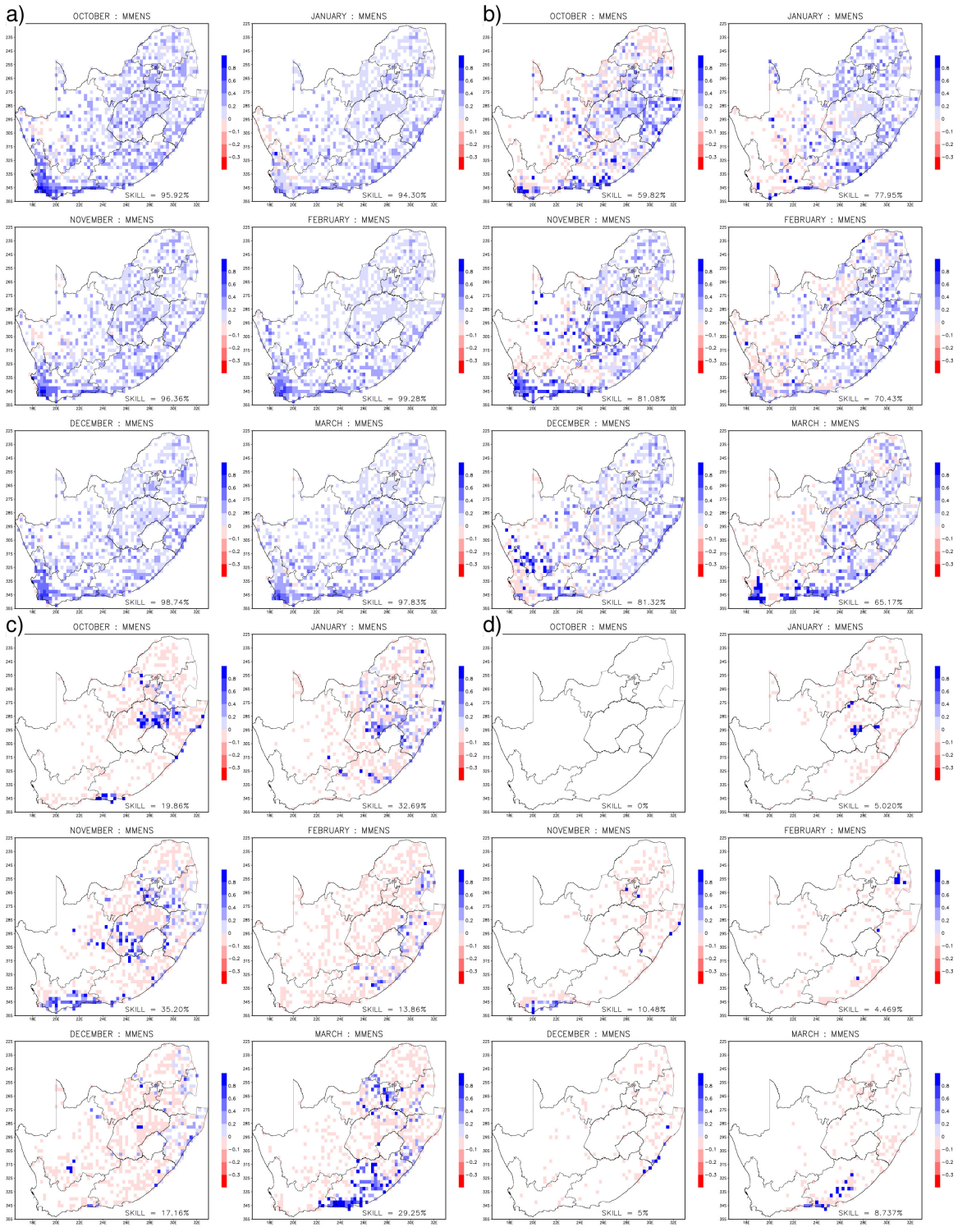


FIGURE 3.20: ETS for the MMENS System. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

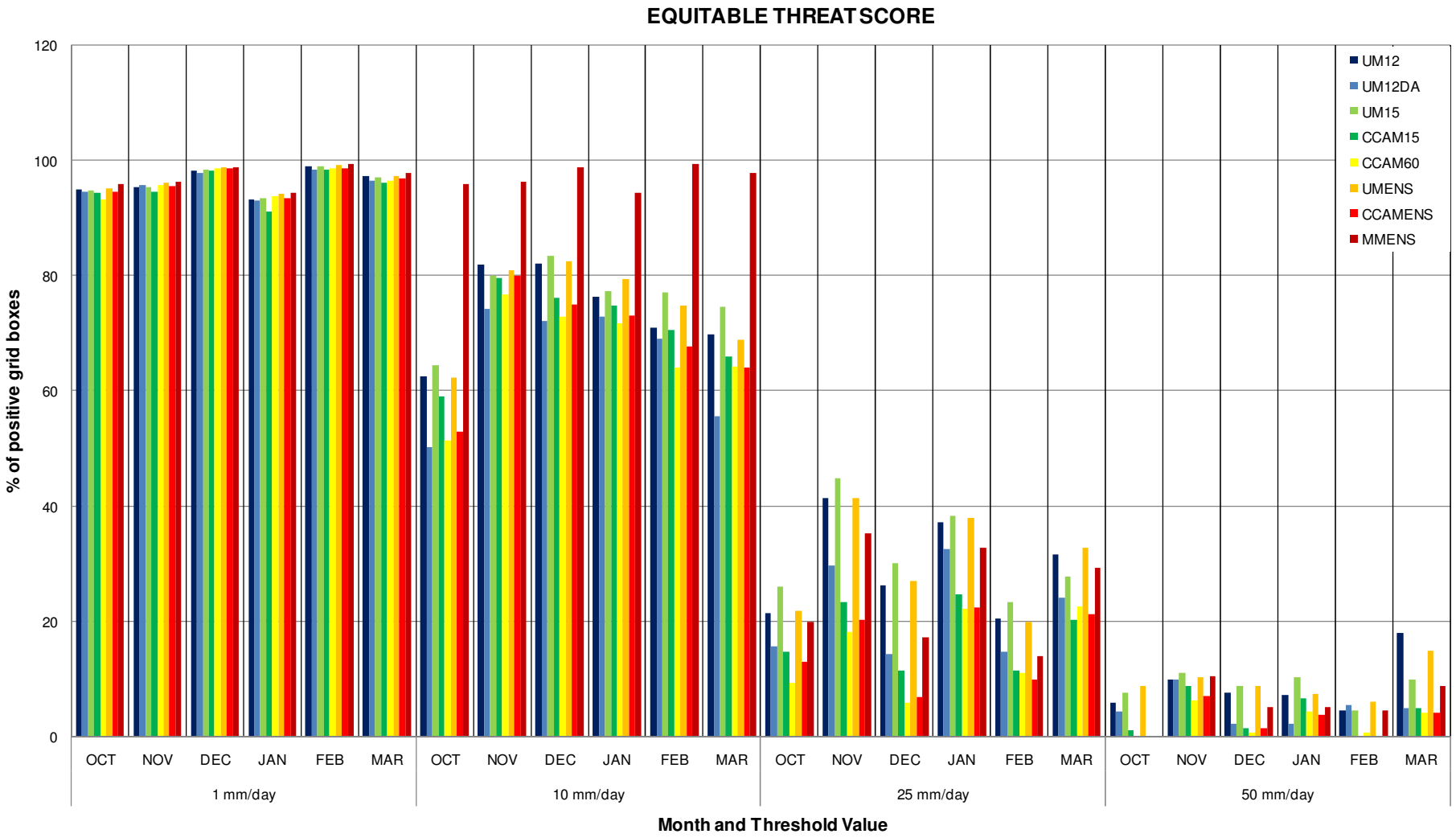


FIGURE 3.21: ETS for the members of the MMENS system for all six months under investigation.

3.4 VERIFICATION OF PROBABILITY FORECASTS

3.4.1 ROC CURVE AND AREA VALUES

The ROC curve for all three ensembles and all four thresholds are shown per monthly graph in Figure 3.22. The UMENS is represented by blue lines, the CCAMENS by green lines and the MMENS by red lines.

For most months (Figure 3.22) the MMENS shows the best discrimination for events exceeding the 1 mm/day threshold (compared to the other thresholds) indicating that the ensemble can distinguish between rainfall and no-rainfall events (ROC areas > 0.7). The scores obtained by the MMENS for the 1 mm/day and 10 mm/day thresholds are in general very similar, with the scores being lower for the 25 mm/day threshold, and decreasing further for the 50 mm/day threshold. However, the MMENS can skillfully distinguish between events and non-events for all the different thresholds.

Looking at Table 3.1, it is shown that the single-model ensemble systems also display the best discrimination abilities for the 1 mm/day and 10 mm/day thresholds, and both systems are skillful for all thresholds. It is interesting to note that although the CCAMENS scores systematically lower than the UMENS for all months and thresholds, the MMENS forecasts are more skillful than those of both the constituting single-model ensembles – for all months and thresholds.

TABLE 3.1: ROC Area values for the UMENS, CCAMENS and MMENS systems

	OCTOBER				JANUARY			
Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
UMENS	0.7509	0.7152	0.7011	0.5936	0.711	0.7331	0.6903	0.5662
CCAMENS	0.7027	0.6662	0.5609	0.5034	0.6265	0.6779	0.581	0.5298
MMENS	0.7961	0.7493	0.7177	0.597	0.7555	0.7735	0.7145	0.5761
	NOVEMBER				**FEBRUARY**			
Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
UMENS	0.7272	0.7386	0.6899	0.6422	0.7165	0.7054	0.5964	0.5741
CCAMENS	0.6764	0.6893	0.5754	0.5257	0.6345	0.6711	0.5419	0.5035
MMENS	0.7732	0.771	0.6974	0.6442	0.7665	0.7504	0.6092	0.5741
	DECEMBER				**MARCH**			
Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
UMENS	0.7149	0.7271	0.6338	0.5877	0.6975	0.7129	0.7205	0.6777
CCAMENS	0.6492	0.6695	0.5277	0.504	0.5989	0.6832	0.6148	0.5288
MMENS	0.7637	0.7652	0.6435	0.5876	0.7387	0.7582	0.7295	0.6815

This might be due to the use of the trapezoid rule and the lower number of forecast probabilities in the CCAMENS forecasts. This is a significant result, which indicates the value of a multi-model ensemble system constructed from single-model members that have independent skill.

All the ensembles obtain the highest scores in October and the lowest values in March.

The 25 mm/day threshold for all three systems shows that there is an insufficient number of forecast categories, resulting in the straight lines for high-probability values.

3.4.2 RELIABILITY DIAGRAM

The ROC analysis has shown that the MMENS system is best able to discriminate between rainfall events exceeding predetermined thresholds, from non-events. Hence, reliability diagrams are only constructed for this system (Figure 3.23). These diagrams however show that the MMENS system exhibit over-confidence for all thresholds and all months considered.

Considering the 1 mm/day threshold for all six months, the system is under-forecasting the events with low probabilities and over-forecasting for higher probabilities (< 70%).

Looking at October, November, January and March, the systems have similar slopes for both the 10 and 25 mm/day threshold, with the 10 mm/day threshold having slightly better reliability.

For December and February, the 10 mm/day threshold has similar reliability as for the afore-mentioned months, but the 25 mm/day curves fluctuate more and have a lower reliability. Considering the same two months, the 50 mm/day threshold has no meaningful resolution for these events. The 50 mm/day threshold events for the remaining four months have some reliability with November having the most.

Looking at the sharpness diagrams in Figure 3.23 (a)-(f), it is seen that for all of the thresholds the MMENS has high confidence. In all four of the threshold events, the most number of forecasts are made in the lower probabilities, with some increase with the 10 mm/day threshold events in the higher probabilities.

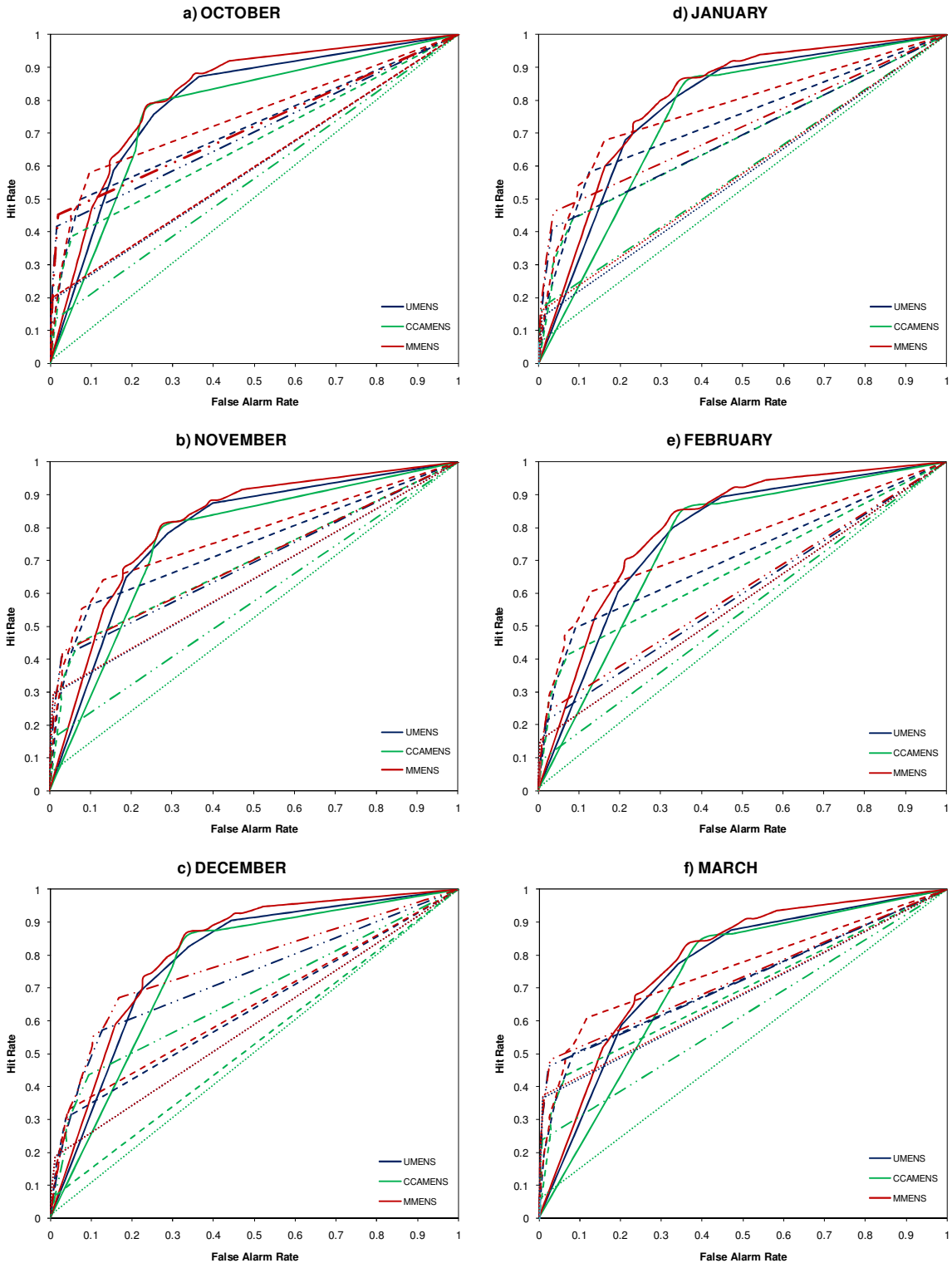
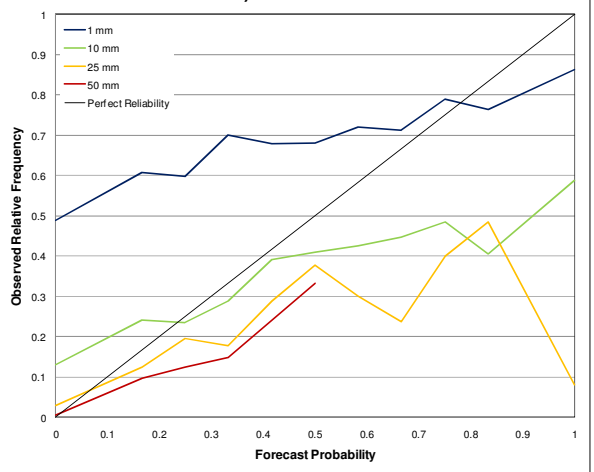
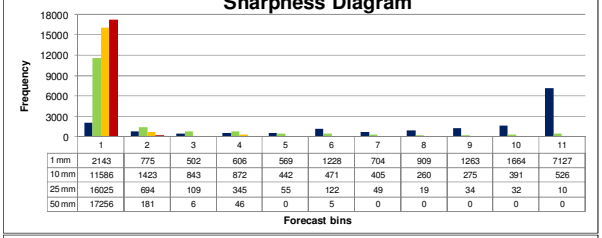


FIGURE 3.22: ROC curves for all three ensembles. (a) October, (b) November, (c) December, (d) January, (e) February and (f) March. Each of the threshold are also represented with the 1 mm/day threshold shown by solid lines, the 10 mm/day threshold represented by the dashed lines, 25 mm/day given by the dashed-dot lines and the 50 mm/day lines represented by dots.

a) OCTOBER

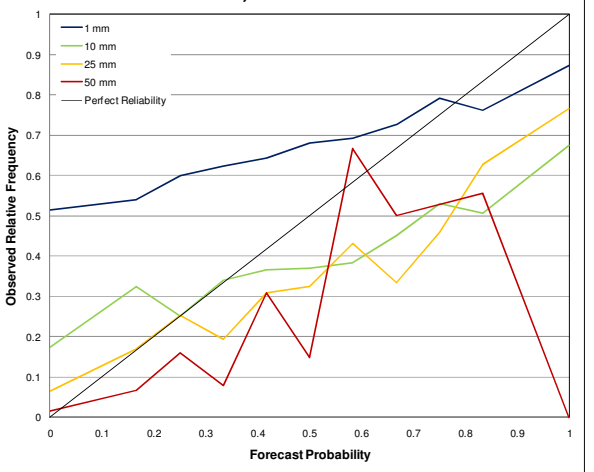


Sharpness Diagram

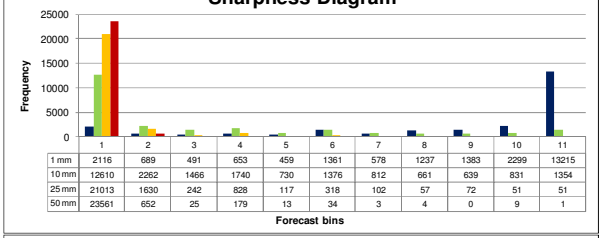


	1	2	3	4	5	6	7	8	9	10	11
1 mm	2143	775	502	606	569	1228	704	909	1263	1664	7127
10 mm	11586	1423	843	872	442	471	405	260	275	391	526
25 mm	16025	694	109	345	55	122	49	19	34	32	10
50 mm	17256	181	6	46	0	5	0	0	0	0	0

d) JANUARY

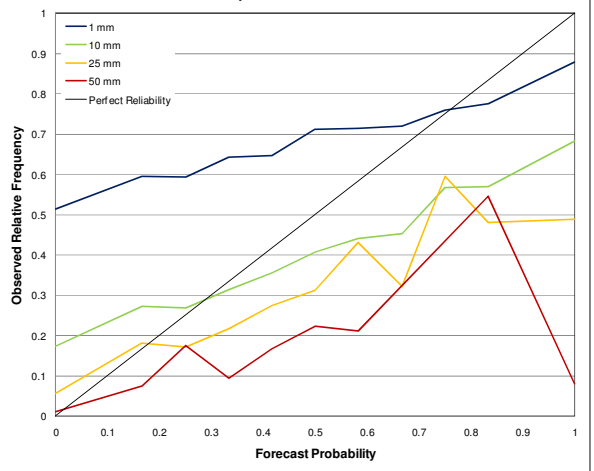


Sharpness Diagram

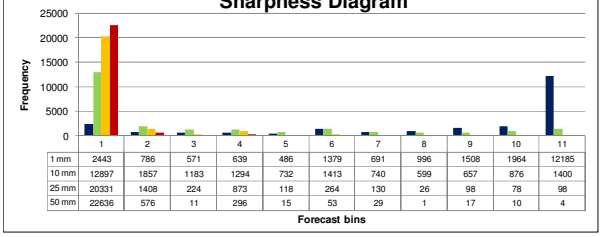


	1	2	3	4	5	6	7	8	9	10	11
1 mm	2116	689	491	653	459	1361	578	1237	1383	2299	13215
10 mm	12610	2262	1466	1740	730	1376	812	661	639	831	1354
25 mm	21013	1630	242	828	117	318	102	57	72	51	51
50 mm	23561	652	25	179	13	34	3	4	0	9	1

b) NOVEMBER

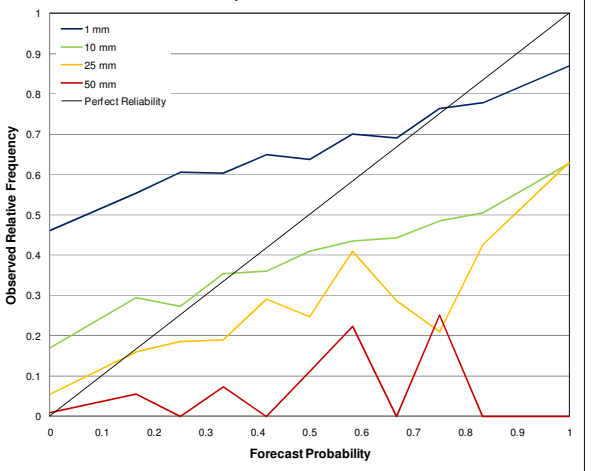


Sharpness Diagram

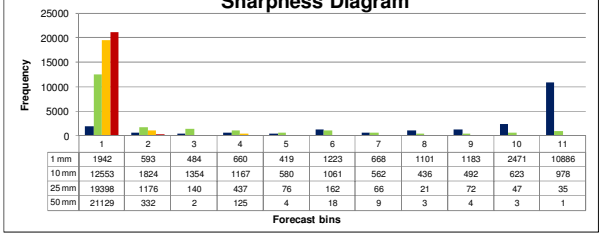


	1	2	3	4	5	6	7	8	9	10	11
1 mm	2443	786	571	639	486	1379	691	996	1508	1964	12185
10 mm	12897	1857	1183	1294	732	1413	740	599	657	876	1400
25 mm	20331	1408	224	873	118	264	130	26	98	78	98
50 mm	22636	576	11	296	15	53	29	1	17	10	4

e) FEBRUARY



Sharpness Diagram



	1	2	3	4	5	6	7	8	9	10	11
1 mm	1942	593	484	660	419	1223	668	1101	1183	2471	10886
10 mm	12553	1824	1354	1167	580	1061	562	436	492	623	978
25 mm	19398	1176	140	437	76	162	66	21	72	47	35
50 mm	21129	332	2	125	4	18	9	3	4	3	1

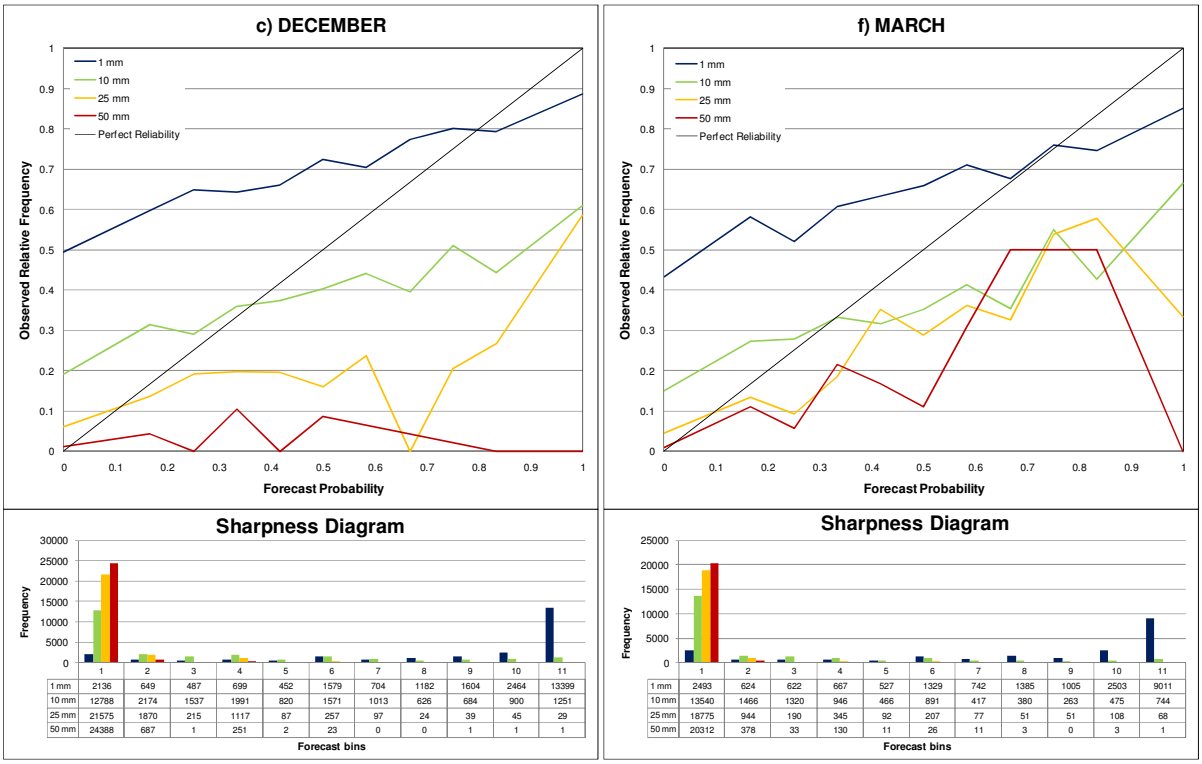


FIGURE 3.23: Reliability and sharpness diagrams for the MMENS system. (a) October, (b) November, (c) December, (d) January, (e) February and (f) March

3.5 SUMMARY

The precipitation forecasts of the individual members, single-model ensemble and multi-model ensemble systems have been presented. The ensemble systems are tested over three seasons from 2006/07 to 2008/09. The ensemble systems are constructed for each of the individual members to have equal weight in the respective single-model ensemble and in turn the equal combination of the single-model ensemble systems forms the multi-model ensemble system.

In terms of the skill for each of the members, the five different members are skillful in predicting rainfall for the South African domain. All of the members are however less skillful in predicting the extreme rainfall occurrences (>= 50 mm/day). It is determined that during the months where one or two of the members had low skill, this affected the skill of the multi-model system. The accuracy (bias) of the members seems to increase to the end of the rainfall season and their ability to detect the predefined threshold events decrease.

Again considering the skill of the members, it is seen that for the UM, the data assimilation member has the best skill whereas the 15 km horisontal resolution

member has the lowest overall skill. The CCAM 15 km member has the best skill with threshold values greater than 25 mm/day. This is an indication that a multi-model system has an advantage over that of a single-model ensemble system. The multi-model ensemble system has the advantage of the skillful UM data assimilation member with lower thresholds after which the CCAM 15 km members contributes greater skill at the higher threshold.

Looking at the multi-models' ability to distinguish between events and non-events, the multi-model has a better discrimination than the two single-model ensemble systems.

The multi-model ensemble system can possibly be improved by removing the model errors within ensemble members as well as through the use of a different combination method.

CHAPTER 4

VERIFICATION OF THE LOW RESOLUTION SINGLE-MODEL ENSEMBLE PREDICTION SYSTEM

This chapter describes the verification results of a low resolution single-model ensemble forecast system. The system in question is the NCEP GEFS, of which the details and set-up are provided in Chapter 2. The purpose of this investigation is to determine a reference level of forecast skill that the high-resolution multi-model system constructed earlier should be able to exceed. This follows from the fact that the NCEP GEFS forecasts are available via the internet, implying that the computational effort required to locally run high-resolution regional models, and their eventual uptake into a multi-model ensemble, can only be justified if the multi-model system is more skillful than the low-resolution ensemble.

Both the accuracy and skill of the single-model ensemble system is quantified. Verification is performed by comparing the forecast output from the ensemble to observed rainfall data. Firstly, the forecast output is verified deterministically over the entire domain for each of the four rainfall thresholds, and secondly the forecasts are verified probabilistically. The verification results are subsequently used to compare the skill of the single-model low resolution system to that of the high resolution multi-model system verified in chapter 3.

4.1 VERIFICATION OF FORECAST ACCURACY

4.1.1 BIAS CALCULATIONS

In Figure 4.1 the spatial bias for the domain of the NCEP forecast is shown. As in Chapter 3, the monthly maps are shown chronologically from top-left to bottom-right. Looking at October, it is seen that the NCEP under-forecast for the majority of the western and northern parts of South Africa. This is the trend for most of the period except during January and February when the system tends to over-forecast the south western parts. The central and eastern parts are mostly over-forecast during December. During January and February the rainfall forecasts for the Gauteng Province are mostly less than the rainfall observed, where during the rest of the period it was mostly over-forecast. Over the Free State Province region, the system

tends to severely over-forecast rainfall during December, as well as the eastern parts of the KZN Province during January and February.

When looking at the area average bias values, it is seen that the month of October had the lowest overall bias. In Figure 4.2 it is seen that the NCEP forecast has a lower overall bias during each of the six months of the study than MMENS. Both systems did however have the lowest bias during October and the greatest during December. This graph however should be interpreted together with the spatial maps. Looking again at Figure 3.2c, it is seen that the MMENS has a wet bias but does not experience the extreme bias values as that of the NCEP system, (like those calculated for November).

Figure 4.3 is similar to Figure 3.4 (explanation in section 3.1.1). The MMENS tends to over-forecast the occurrence of events for the lower thresholds, and under-forecast the occurrences for the higher thresholds (Figure 4.3). The NCEP forecast displays very different characteristics. For October and March, the NCEP forecast under-forecasts the frequency of rainfall events across the different threshold categories, whilst for the remaining months the over-forecast of the frequency of events is less than for the multi-model system. In January and February months the NCEP the frequency of NCEP forecasts is very similar to the observation except for slightly higher values in 2-3 mm/day range.

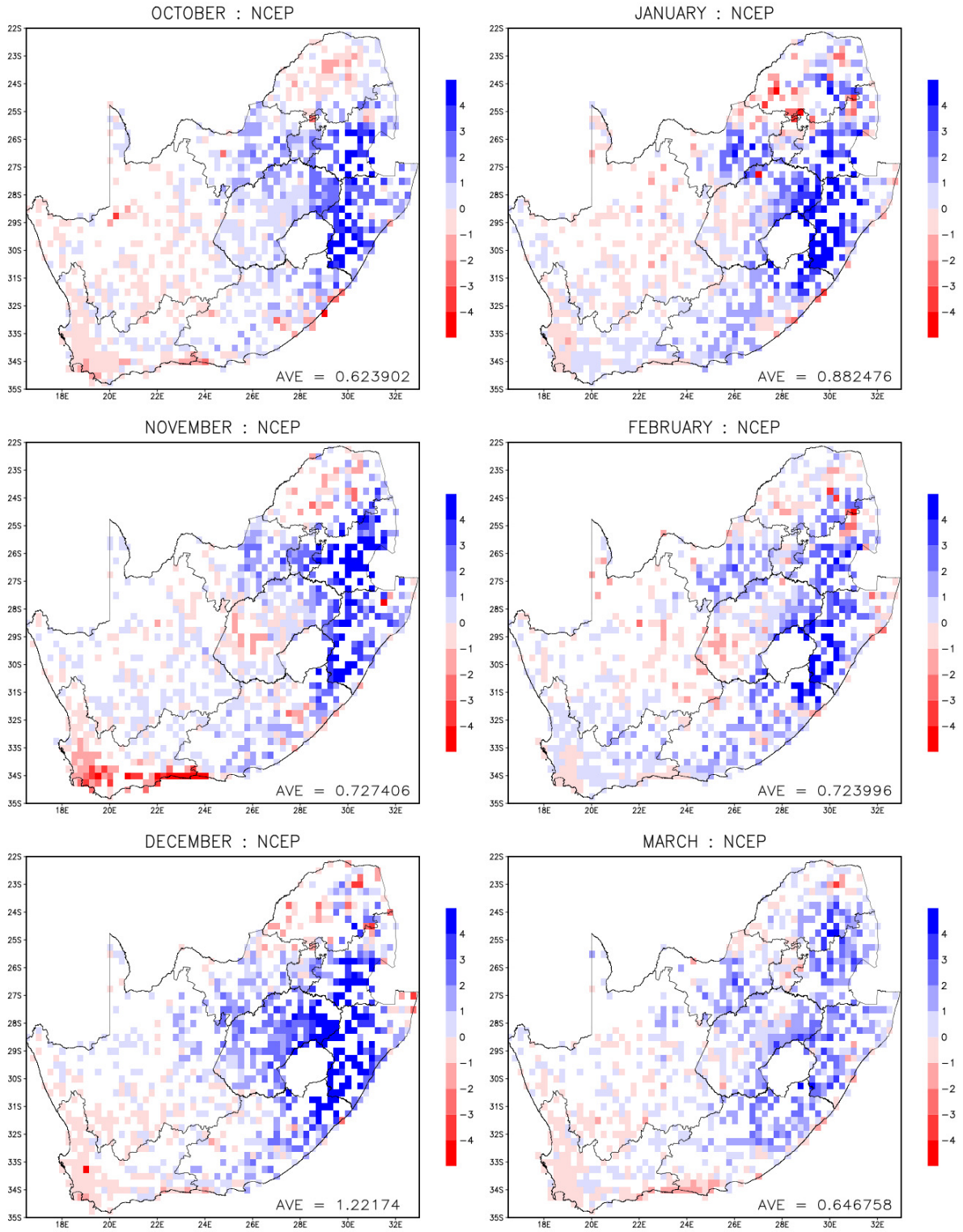


FIGURE 4.1: The spatial maps for the bias for the single-model ensemble for the six months considered.

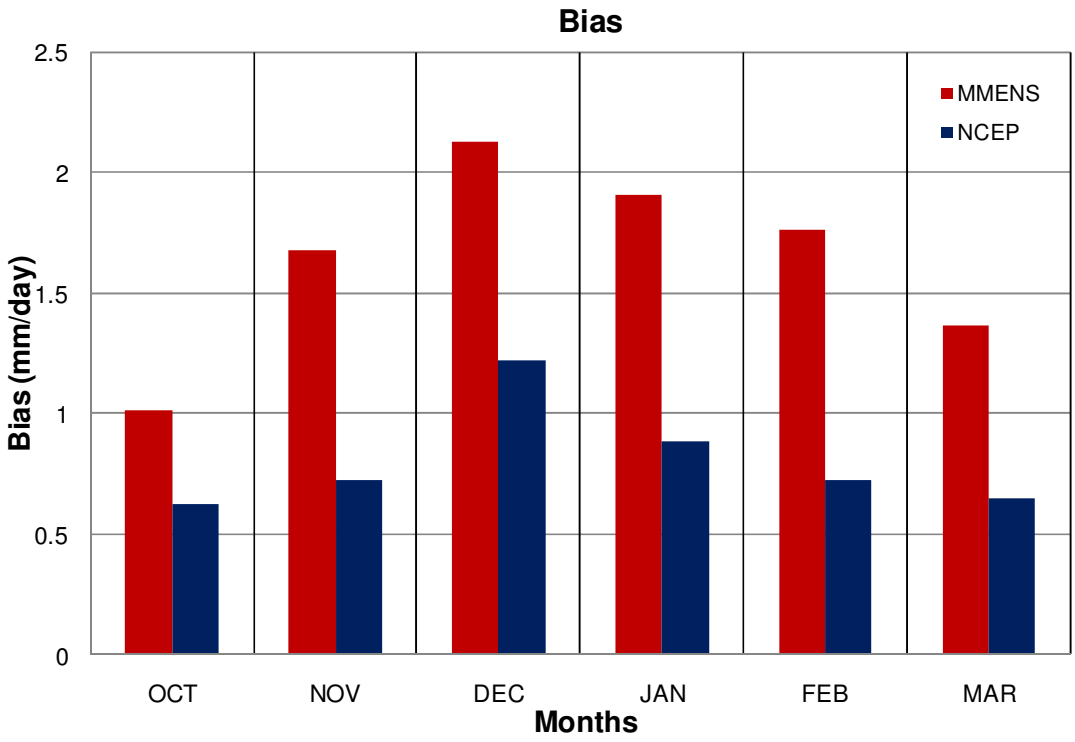
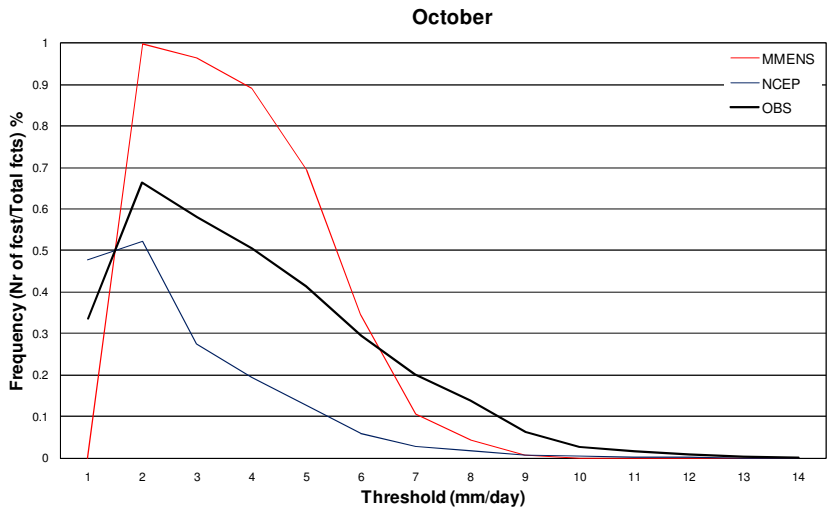
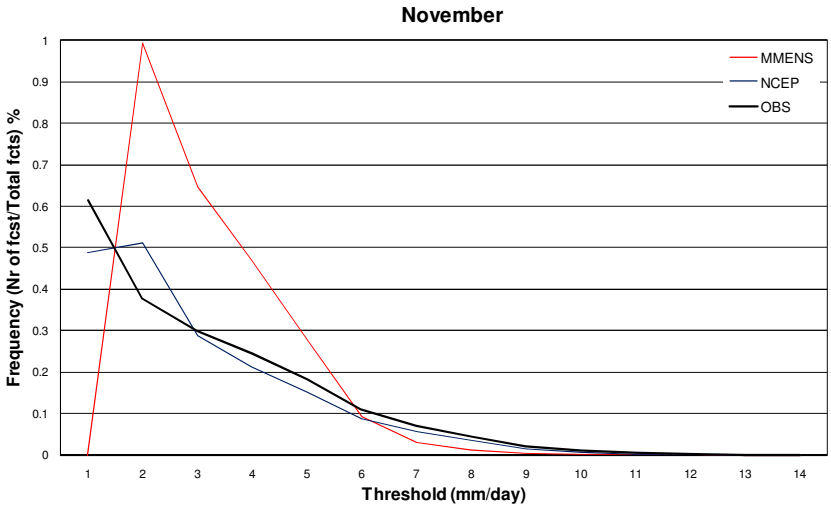
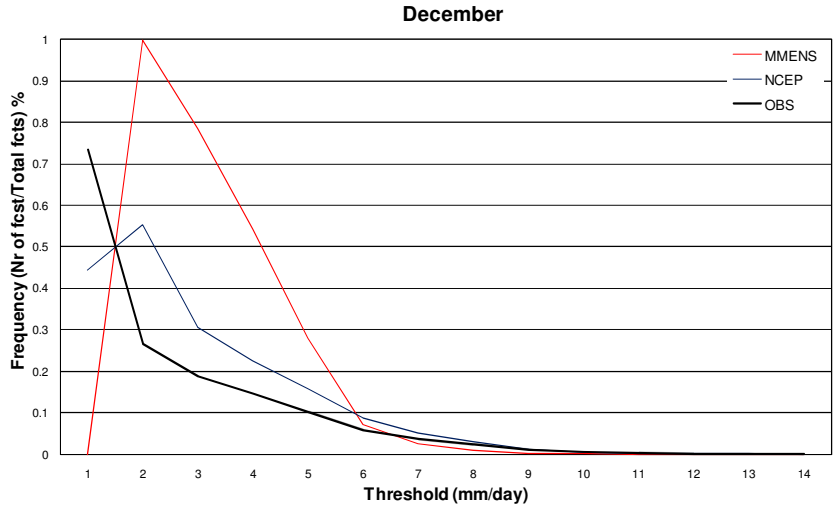


FIGURE 4.2: A summary of the bias for the individual members for the six months under investigation


October
MMENS
NCEP
OBS
Frequency (Nr of fcst/Total fcts) %
Threshold (mm/day)


November
MMENS
NCEP
OBS
Frequency (Nr of fcst/Total fcts) %
Threshold (mm/day)


December
MMENS
NCEP
OBS
Frequency (Nr of fcst/Total fcts) %
Threshold (mm/day)

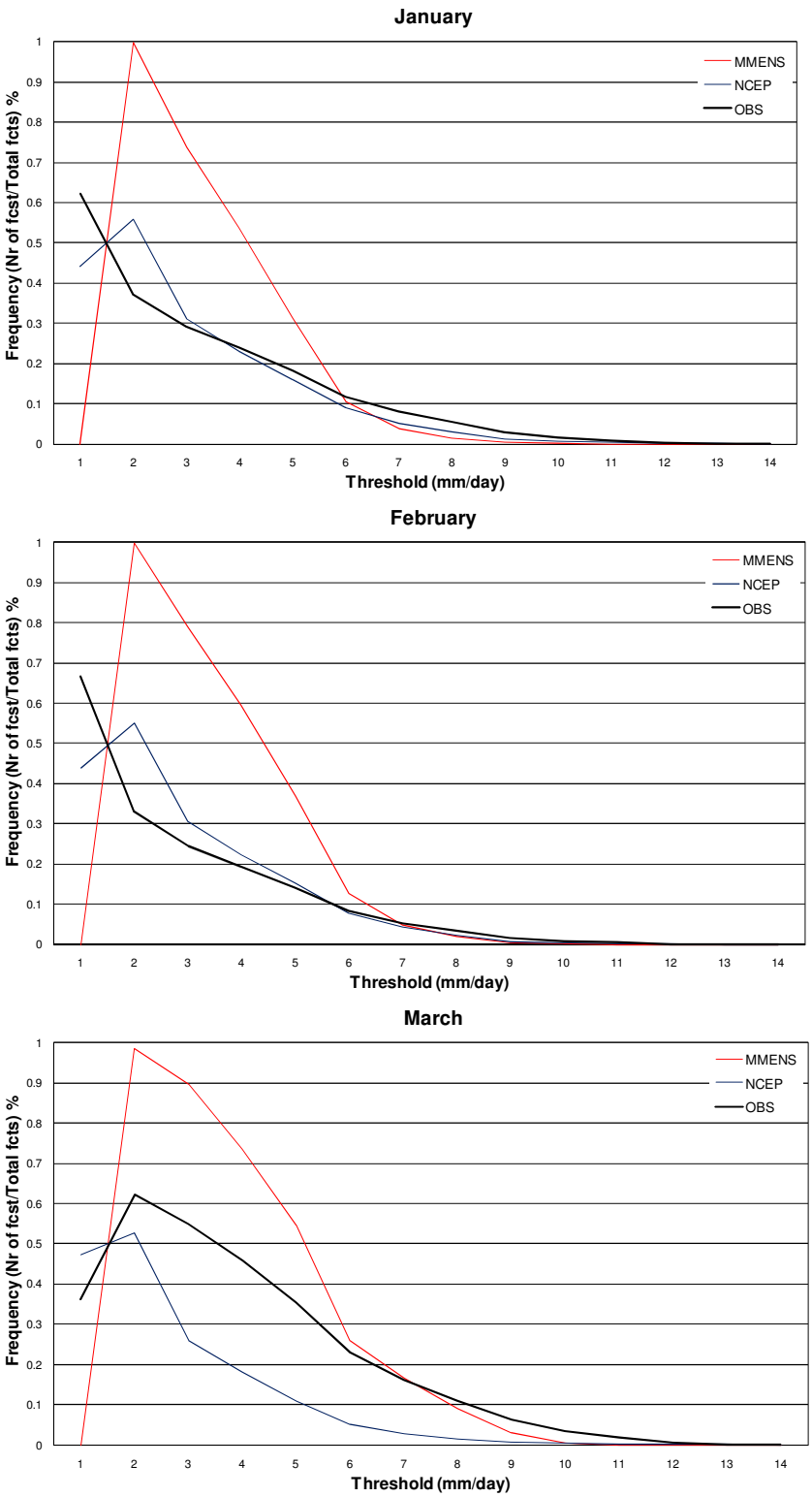


FIGURE 4.3: The observed and frequency of occurrence for the six months under investigation. (a) October, (b) November, (c) December, (d) January, (e) February and (f) March.

4.2 VERIFICATION OF FORECAST SKILL

4.2.1 BRIER SKILL SCORE

As with the MMENS system (Figure 3.7a), the NCEP system has no skill in predicting the occurrence of events above the 1 mm/day threshold, over the largest part of the domain (Figure 4.4a). The NCEP system is more skillful in predicting the occurrence of events above the 10 mm/day threshold (compared to the 1 mm/day threshold forecasts, but still lacks skill over significantly large parts of eastern South Africa.

The NCEP system's skill in predicting events above the 25 mm/day threshold display a similar pattern as for the 10 mm/day forecasts. There is no skill over the eastern Highveld in October and November and over the central interior in December. From January to March NCEP does show skill over these areas. The MMENS system does not show these large discrepancies in skill between early and late summer (Figure 3.7). For the 50mm/day events the NCEP system is generally skillful, with the forecast for 50 mm/day events being generally skillful over eastern South Africa.

Considering Figure 4.5 it is seen that the percentage grid boxes for which the NCEP forecasts are skillful are the highest for the 10 mm/day threshold, and relatively low for the 1 mm/day threshold. The NCEP system is outscored by the MMENS systems for all six months for the 10 mm/day, 25 mm/day and 50 mm/day thresholds. This is a significant result, showing that the locally constructed multi-model system offers level of skill exceeding that of the low resolution NCEP ensemble.

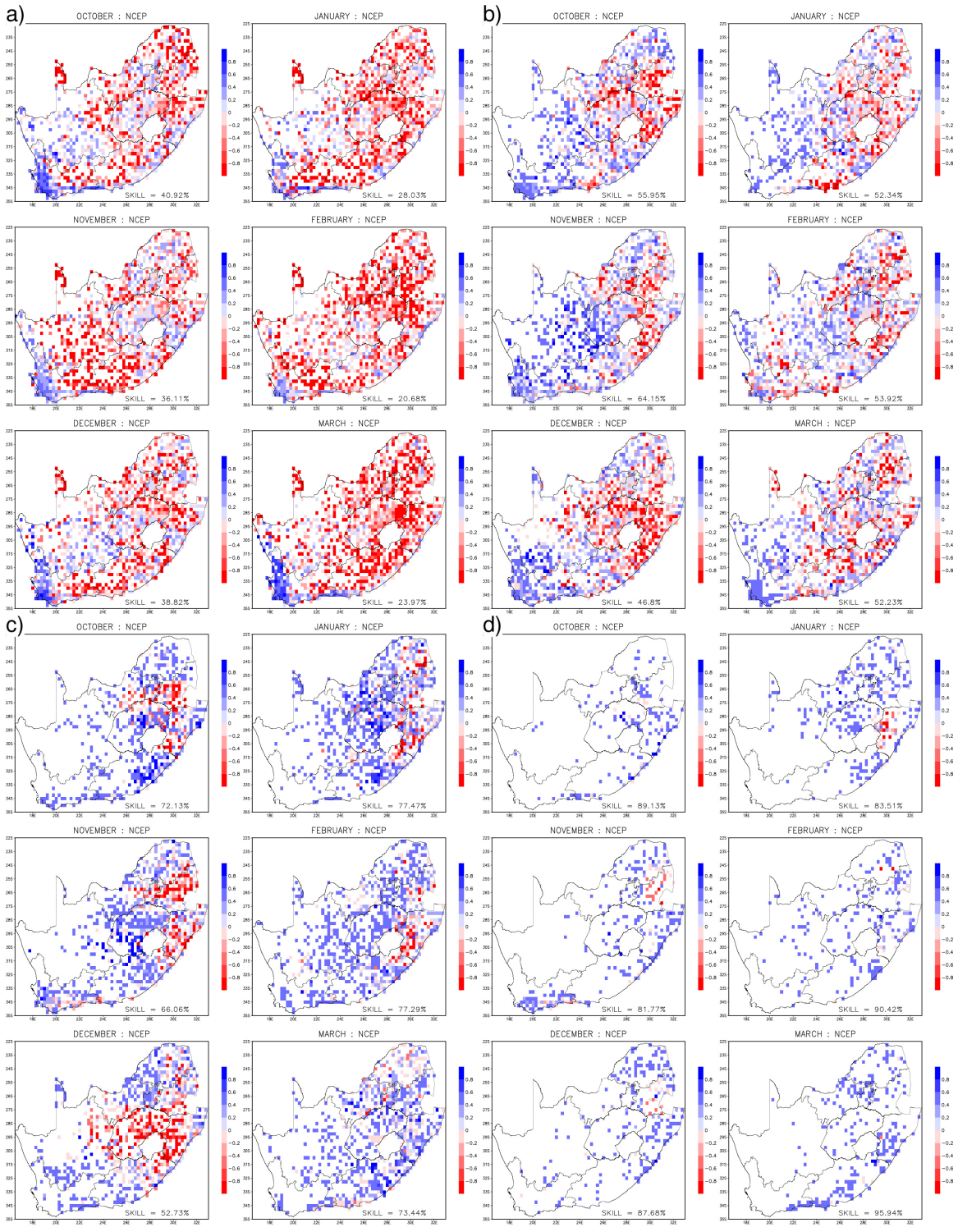


FIGURE 4.4: Spatial maps for the BSS for the NCEP system for the four thresholds for all six months. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

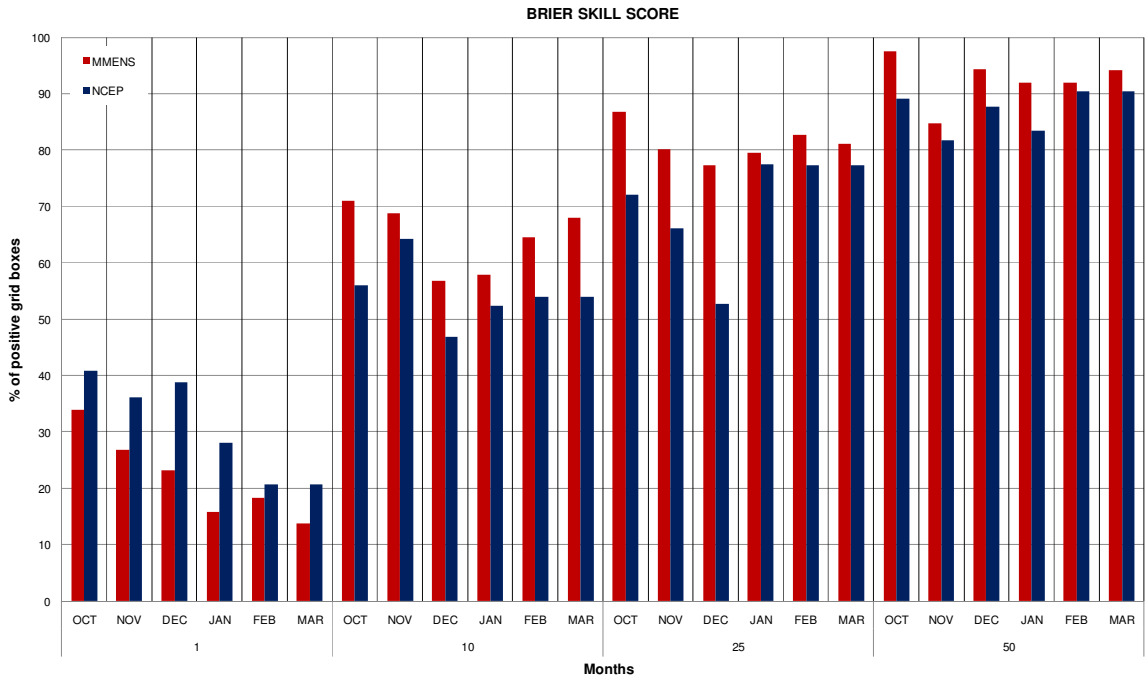


FIGURE 4.5: A summary of the Brier skill score for the four threshold values per system for each of the six months under investigation.

4.2.2 MEAN SQUARED ERROR SKILL SCORE

Figure 4.6 reveals that the largest skill of the NCEP system in predicting daily rainfall totals is for the months of January and February, with large areas where the forecasts are not skillful present for October and December. Looking at Figure 4.7, it is seen that the MMENS system is more skillful in predicting rainfall totals than the NCEP system, for all months except for January.

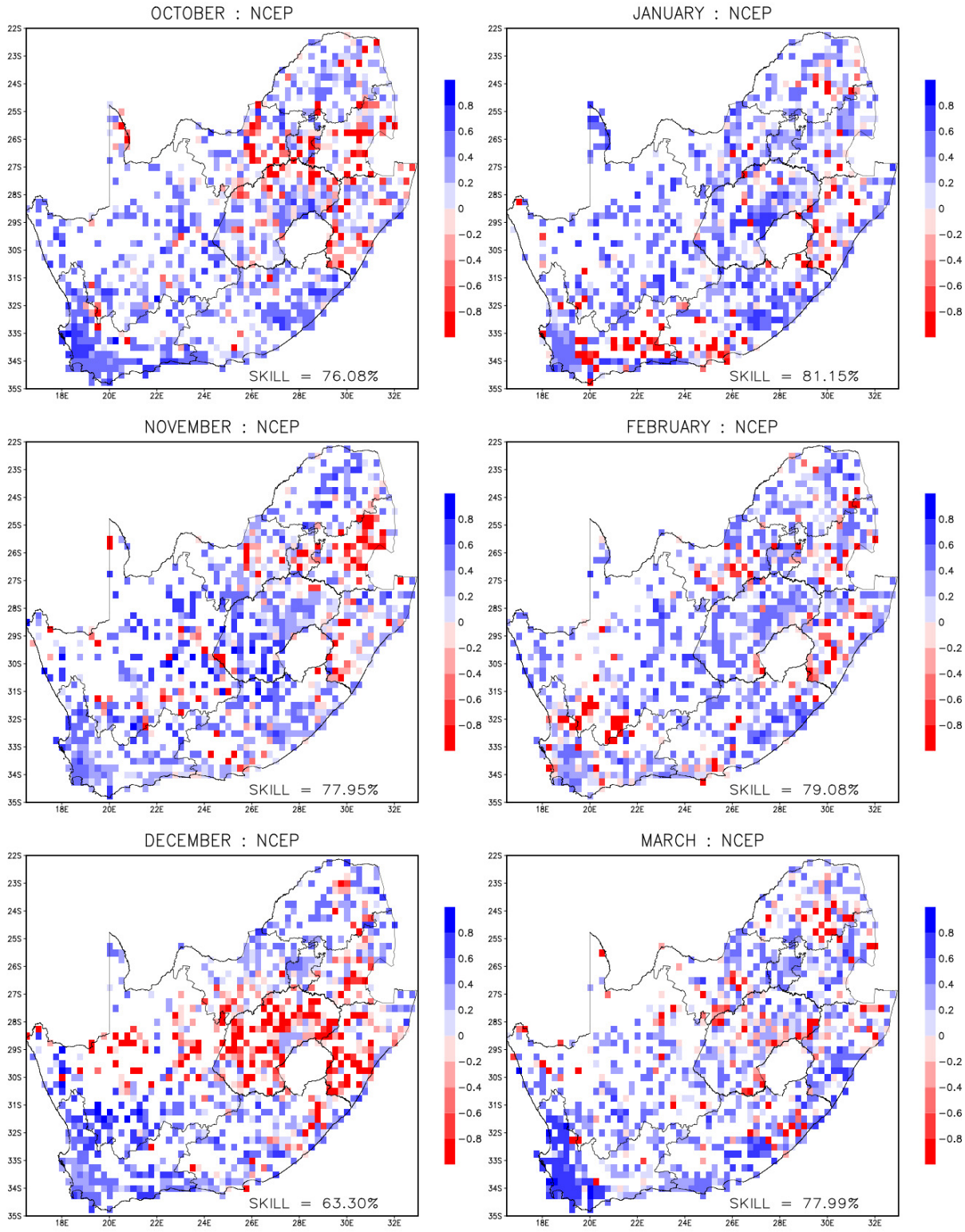


FIGURE 4.6: Spatial maps for the MSESS for all six months under investigation.

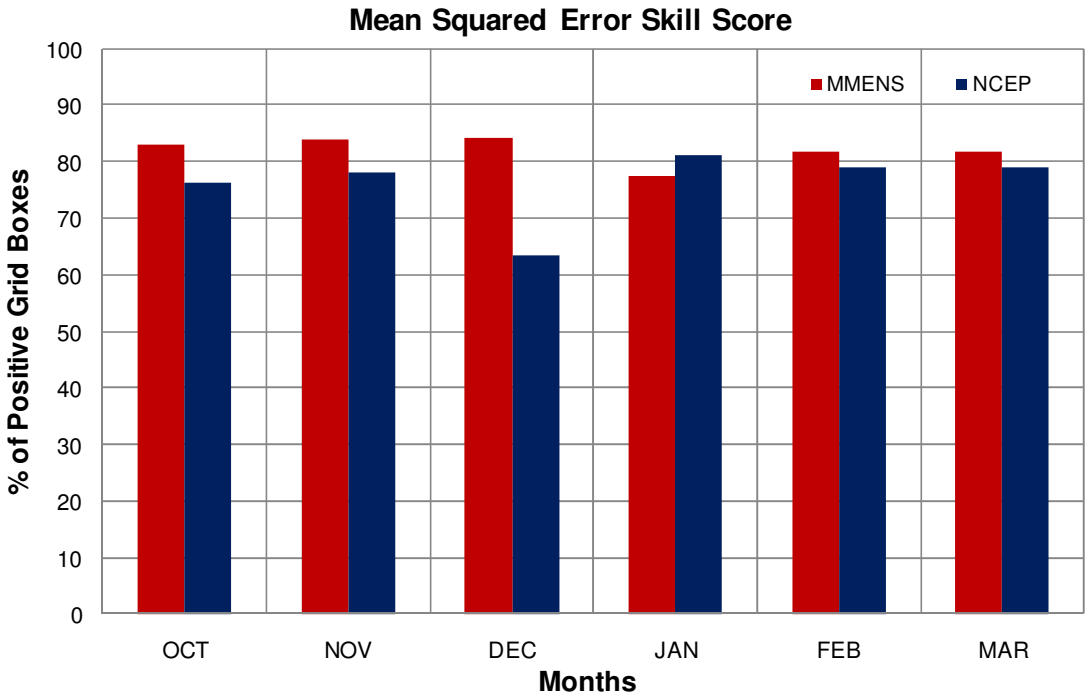


FIGURE 4.7: A summary of the MSESS for each system for the six months under investigation.

4.3 ADDITIONAL VERIFICATION OF DICHOTOMOUS FORECASTS

4.3.1 FREQUENCY BIAS INDEX

The FBI for 1 mm/day thresholds for the NCEP system is shown in Figure 4.8a. Over most of the domain the system has a score greater than 1, indicating an over-forecast for this threshold. The 10mm/day rainfall events are mostly under-forecast over the western half of the domain and over-forecast over the eastern part of the domain (Figure 4.8b). For the 25mm/day threshold, the frequency of events is over-forecast over areas in the eastern Provinces (Mpumalanga and KZN; Figure 4.8 c), with 50 mm/day events being generally under-forecasted.

An analysis for the FBI for both the NCEP and MMENS systems (performance across the domain) are shown in Figure 4.9. Both systems overcast the occurrence of rainfall events (1 mm/day threshold), with the overestimation of the frequency of such events being larger for the MMENS. The MMENS provided a more realistic estimate of the frequency of occurrence of 10 mm/day events than the NCEP forecasts, but is less realistic in the case of the 25 mm/day threshold.

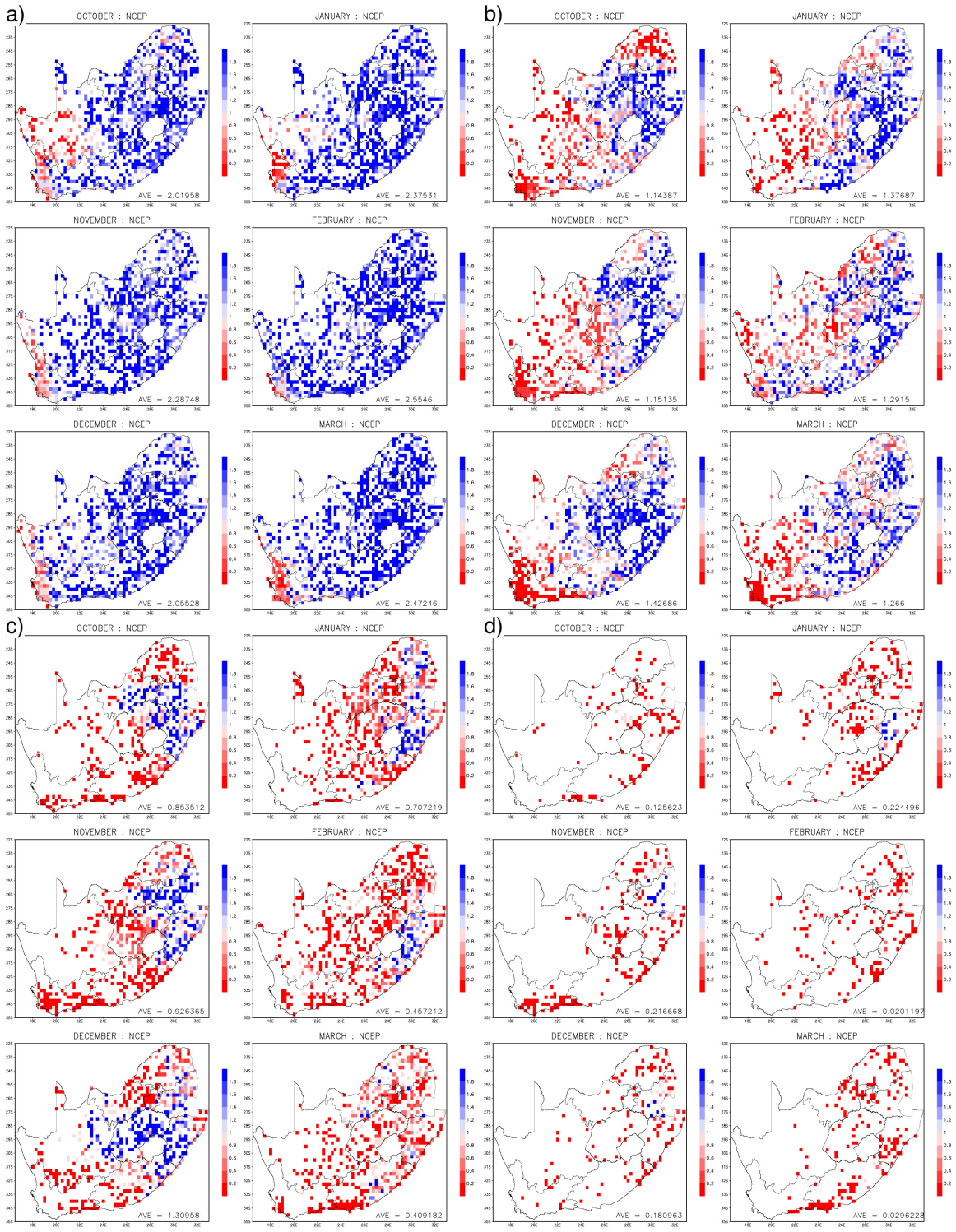


FIGURE 4.8: The FBI for the NCEP system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

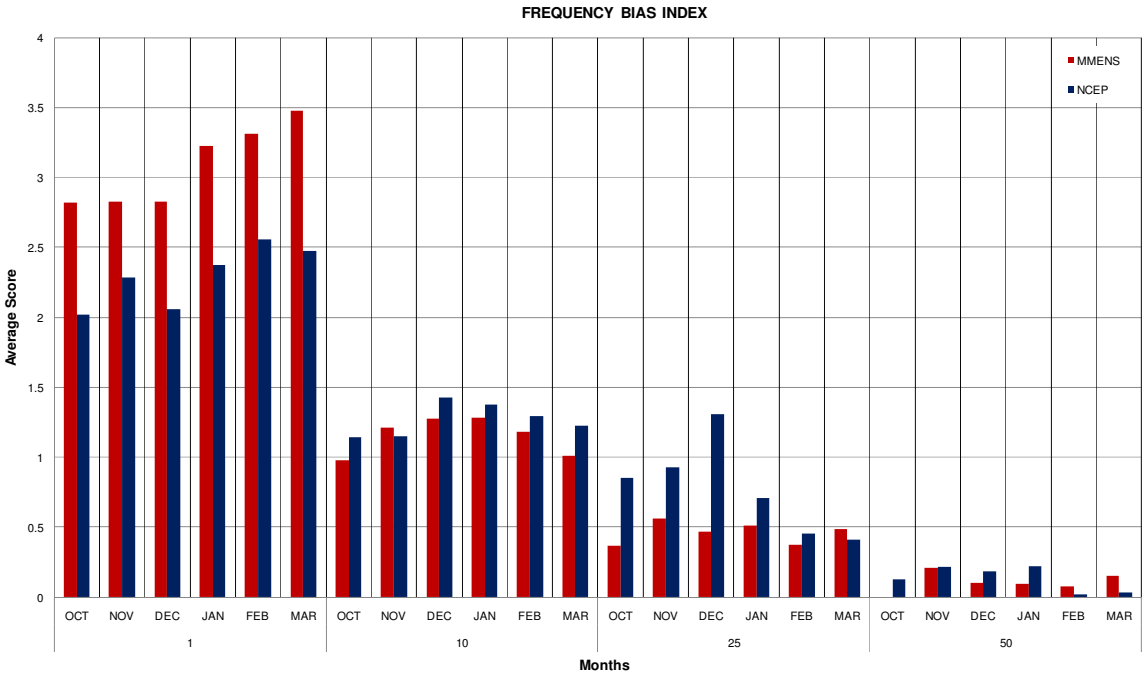


FIGURE 4.9: The FBI summary for the systems for the four threshold values and the six months under investigation.

4.3.2 PROBABILITY OF DETECTION

The POD spatial distribution maps for the NCEP system are shown in Figure 4.10. The NCEP system has a high probability of detection for the 1 mm/day events (Figure 4.10a), especially over the eastern half of the country. This high probability decreases for the 10 mm/day events (Figure 4.10b) and decreases further for the higher threshold events.

Looking at the summary of the POD in Figure 4.11, it is shown that MMENS attains higher POD values than the NCEP system for the 1, 10 and 50 mm/day thresholds. For the 25 mm/day events, the NCEP system has a higher area average POD value for the months of October and December. The higher POD of the 1 mm/day events by the MMENS is likely to be the result of the excessive (compared to the NCEP ensemble) over-forecast of such events by the MMENS.

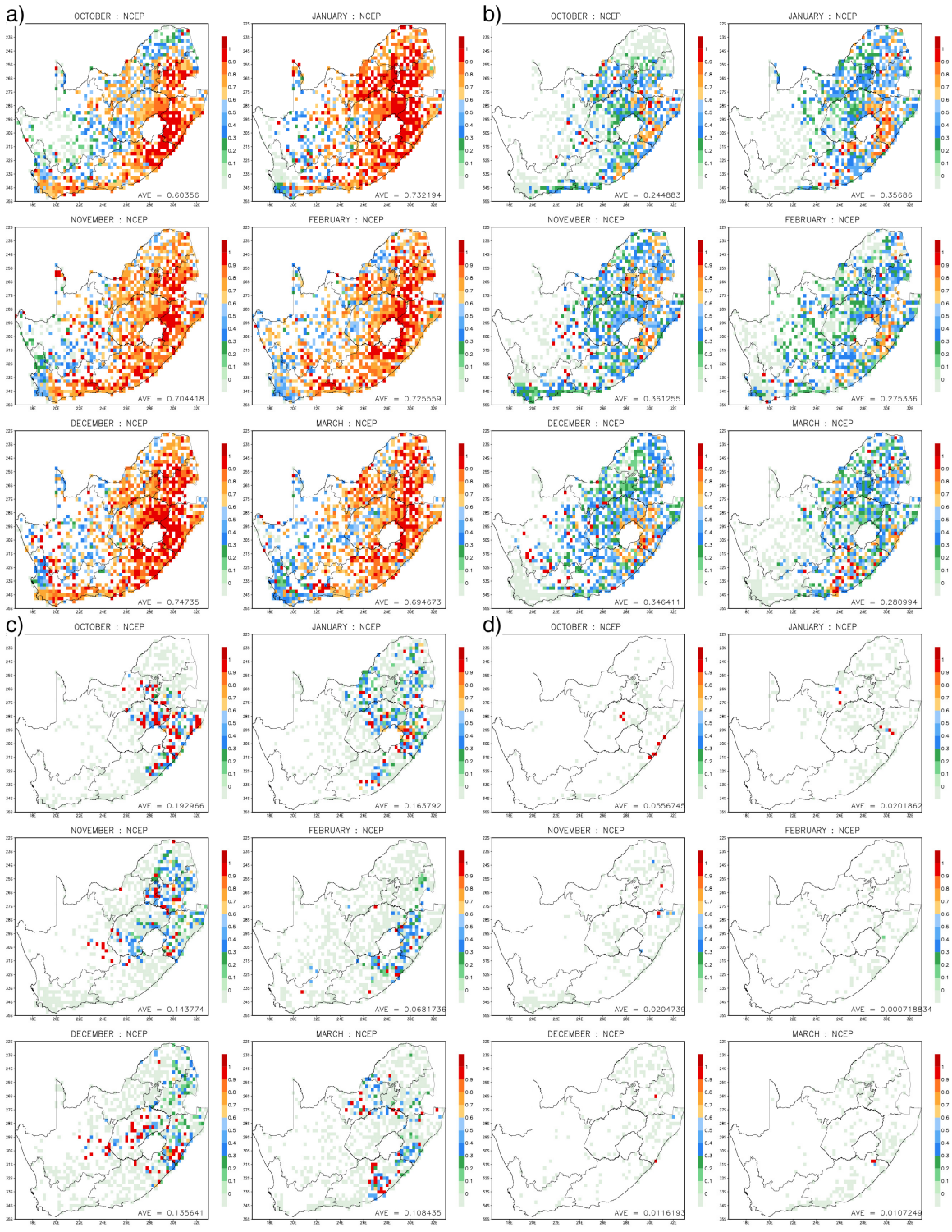


FIGURE 4.10: POD for the NCEP system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

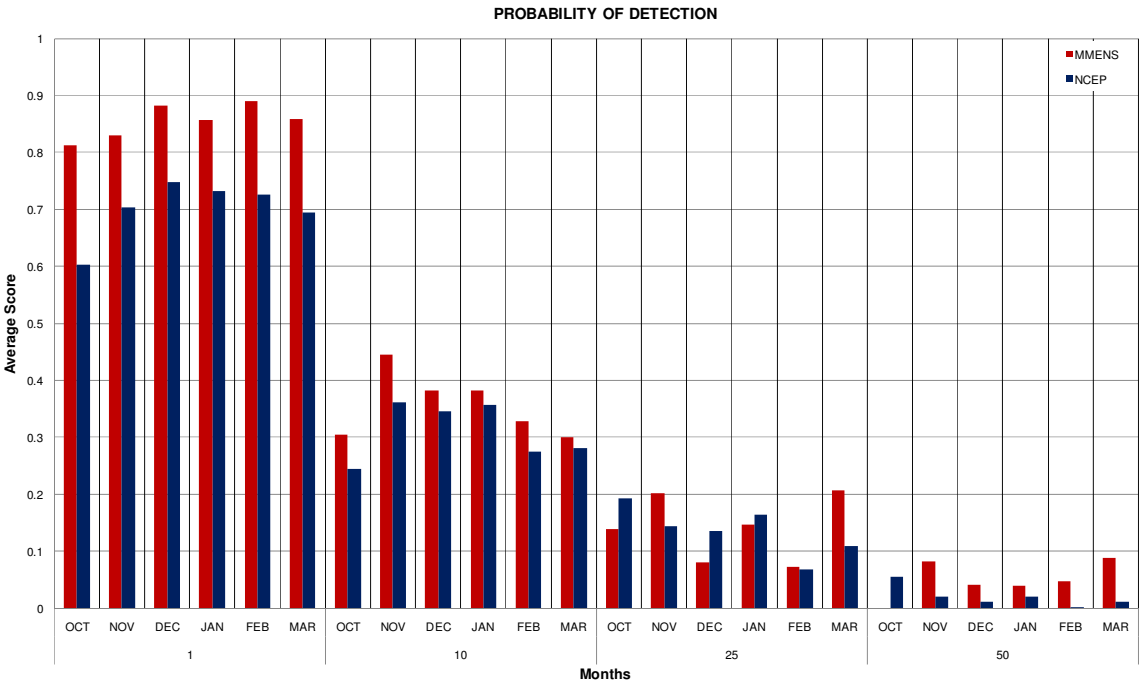


FIGURE 4.11 POD for the NCEP system for the four threshold values and six months under investigation.

4.3.3 FALSE ALARM RATE

In Figure 4.12, the F for the NCEP system is high for the far eastern parts of the domain for the 1 mm/day events (Figure 4.12a), especially over the western parts of the KZN Province during January. For the 25 and 50 mm/day events respectively the NCEP system has negligibly small F values due to the NCEP system not predicting these events that often.

Considering Figure 4.13, the graph shows that the area average F values are the highest for 1 mm/day threshold for the MMENS compared to the NCEP forecasts, but for the other thresholds the MMEMS produces fewer false alarms than the NCEP system

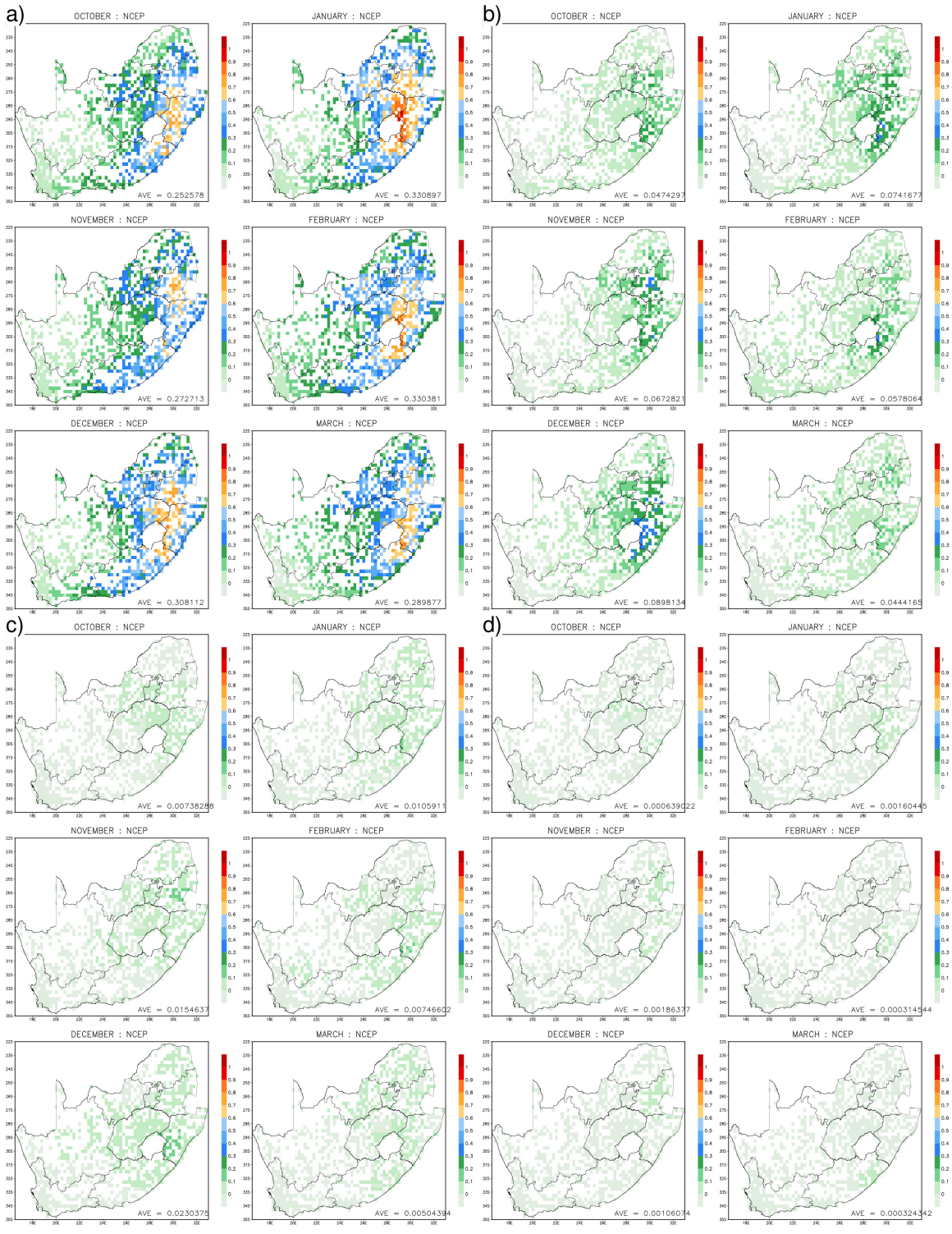


FIGURE 4.12: F for the NCEP system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold

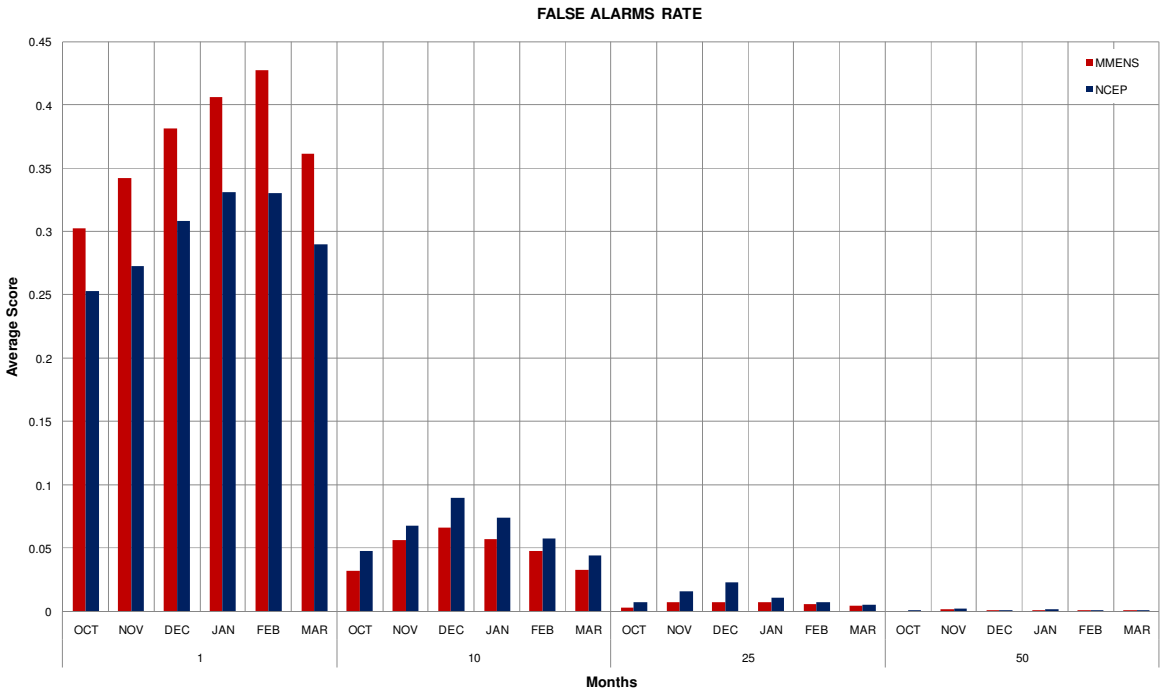


FIGURE 4.13: F for the MMENS and NCEP systems for all six months under investigation.

4.3.4 CRITICAL SUCCESS INDEX

The CSI spatial distribution maps are shown in Figure 4.14. The CSI distribution spatially is greater for the 1 mm/day threshold than any of the remaining three values.

Figure 4.15 reveals that the CSI area average values are higher for MMENS system than for the NCEP system for almost all months and all thresholds, the only exception being October and January for the 25 mm/day threshold.

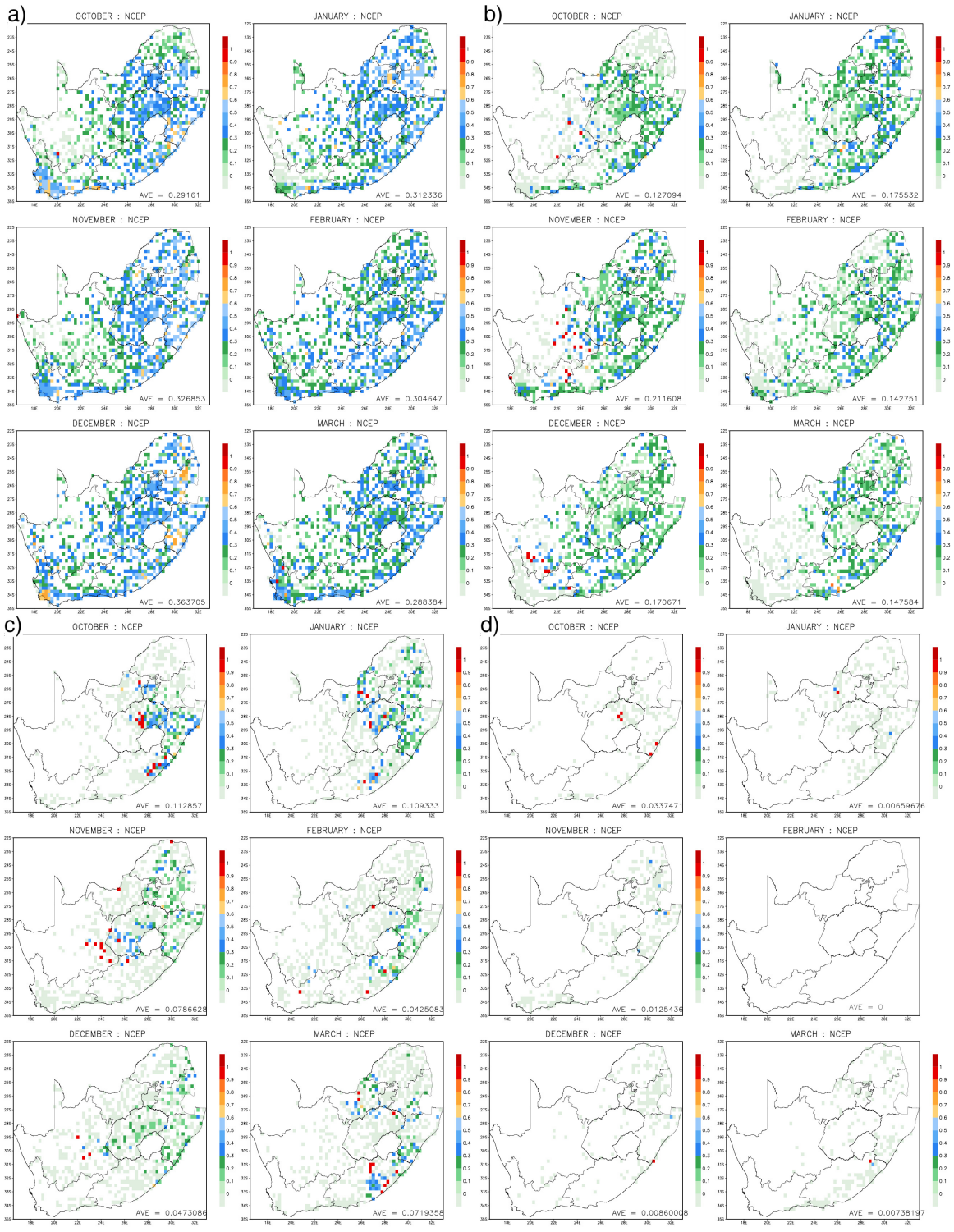


FIGURE 4.14: CSI for the NCEP system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

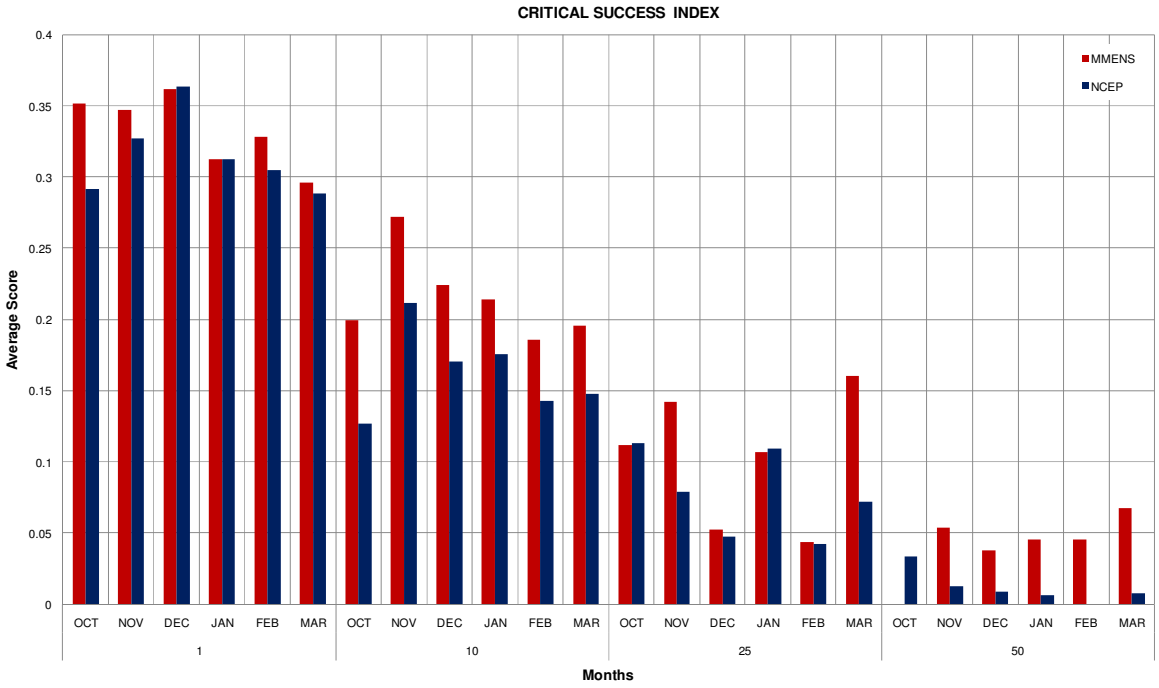


FIGURE 4.15: CSI for the MMENS and NCEP systems for all six months under investigation.

4.3.5 EQUITABLE THREAT SCORE

The ETS spatial distribution maps are shown in Figure 4.16. The NCEP system has no skill over that of random chance in predicting 50 mm/day events, whereas the system is skillfull in the case of 1 mm/day events (Figure 4.16a). For the 10 mm/day events (Figure 4.16c) the NCEP forecast is skillful over the eastern part of the domain, except for the Limpopo Province during October and February. Skill decreases significantly from the 10 mm/day threshold to the 25 mm/day threshold.

Figure 4.17 shows that the number of positive grid boxes (indication of skill) is higher for the MMENS for the 1, 10 and 50 mm/day rainfall events. The MMENS system only performs better than the NCEP system for the 25 mm/day threshold during the months of November and March.

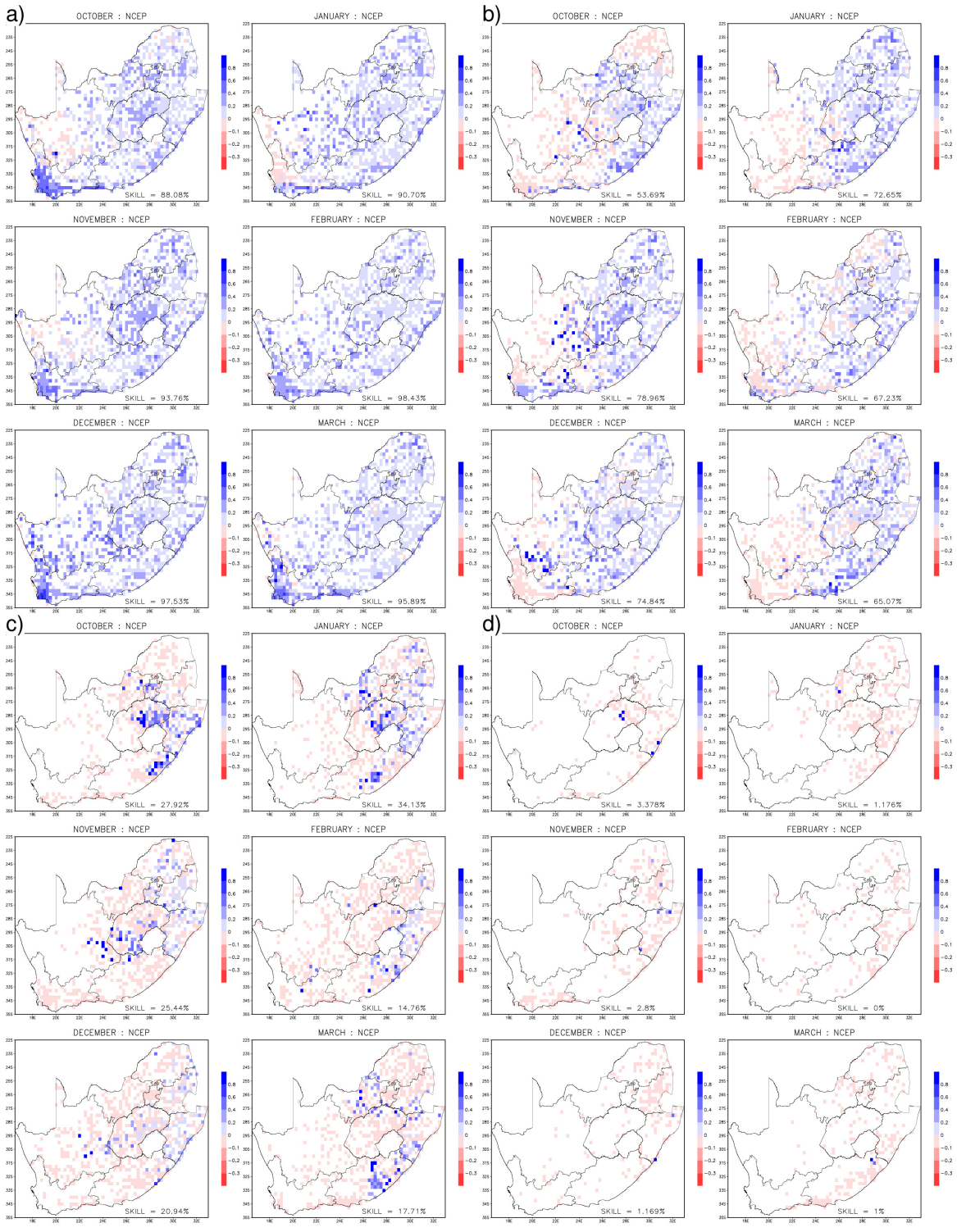


FIGURE 4.16: ETS for the NCEP system. (a) 1 mm/day threshold, (b) 10 mm/day threshold, (c) 25 mm/day threshold and (d) 50 mm/day threshold.

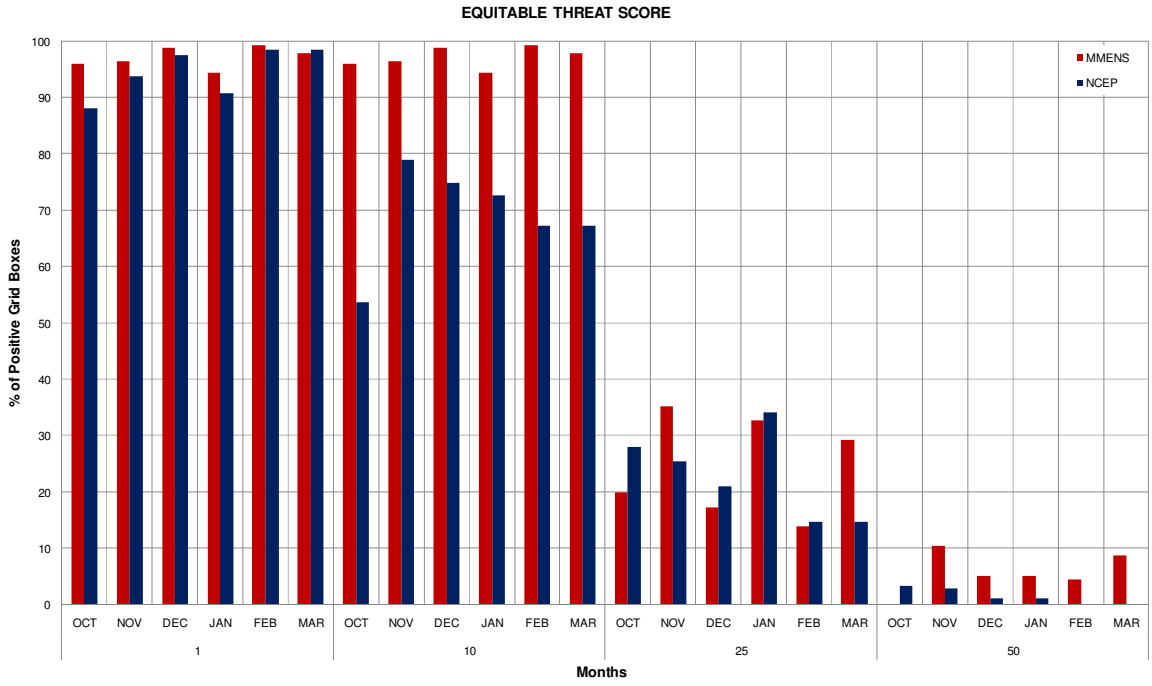


FIGURE 4.17: ETS for the MMENS and NCEP systems for all six months under investigation

4.4 VERIFICATION OF PROBABILITY FORECASTS

4.4.1 ROC CURVE AND AREA VALUES

The ROC curves in Figure 4.18 reveal that the MMENS system is superior over the NCEP forecast in distinguishing between 1 mm/day rainfall events and non-events for all six months under investigation. Indeed, the corresponding ROC areas shown in Table 4.1 for the MMENS, for all six months, are greater than the NCEP forecast areas. This attribute of the MMENS to outperform the NCEP system is also seen true in general for the larger thresholds.

The reliability diagrams (Figure 4.19) display an over-confident performance for both ensemble systems at higher thresholds. Considering the 1 mm/day threshold (solid lines), the blue line (NCEP) indicates a slightly better reliability and resolution than that of the MMENS system (red line). Also, both these systems have a tendency to under-forecast the 1 mm/day events.

The contrary is true for the 10 mm/day events when the MMENS system (red dashed lines) seems slightly better than that of the NCEP system. The same is true for the 25 mm/day events (dash-dot-dash lines), especially for the higher probability bins.

Looking at the sharpness diagrams for the MMENS and NCEP systems in Figure 4.19 a –f, it is seen that the two systems are similar in confidence and sharpness, with both systems having in increase in forecast occurrences at the higher probability values with the 10 mm/day threshold events.

For the 50 mm/day events (dotted lines), both the systems have little reliability, but the MMENS is slightly more reliable in predicting these events.

TABLE 4.1: ROC Area values for the MMENS and NCEP ensemble systems

	OCTOBER				**JANUARY**			
Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
NCEP	0.7275	0.7164	0.6692	0.0009	0.7023	0.7668	0.6803	0.5681
MMENS	0.7961	0.7493	0.7177	0.597	0.7555	0.7735	0.7145	0.5761
	NOVEMBER				**FEBRUARY**			
Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
NCEP	0.725	0.761	0.679	0.5538	0.6599	0.7201	0.6506	0.5035
MMENS	0.7732	0.771	0.6974	0.6442	0.7665	0.7504	0.6092	0.5741
	DECEMBER				**MARCH**			
Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
NCEP	0.7074	0.7246	0.6375	0.5634	0.7019	0.7184	0.6397	0.5757
MMENS	0.7637	0.7652	0.6435	0.5876	0.7387	0.7582	0.7295	0.6815

In order to accurately determine the difference between the two systems, the reliability, resolution and uncertainty of both the forecast systems were calculated for all four threshold values and represented in Table 4.2.

For most of the events, the MMENS has a better resolution than that of the NCEP system, but the MMENS is more reliable than the NCEP for all threshold values for all six months.

The uncertainty was similar for all the threshold values and months for both forecast systems.

TABLE 4.2: Reliability, Resolution and Uncertainty for the MMENS and NCEP ensemble systems

		OCTOBER				JANUARY			
SYSTEM	Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
MMENS	RESOLUTION	0.1476	0.0198	0.0023	-	0.123	0.032	0.005	-
	RELIABILITY	0.0592	0.0225	0.0025	-	0.044	0.032	0.006	-
	UNCERTAINTY	0.1878	0.1633	0.0418	-	0.172	0.203	0.081	-
NCEP	RESOLUTION	0.1266	0.0136	0.0018	-	0.1079	0.0259	0.0068	-
	RELIABILITY	0.1273	0.0387	0.0085	-	0.0841	0.0524	0.0147	-
	UNCERTAINTY	0.1892	0.161	0.0405	-	0.1721	0.2085	0.0823	-
		NOVEMBER				FEBRUARY			
SYSTEM	Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
MMENS	RESOLUTION	0.133	0.029	0.005	0.0003	0.21	0.027	0.003	0.0001
	RELIABILITY	0.051	0.028	0.005	0.0015	0.04	0.031	0.006	0.0011
	UNCERTAINTY	0.172	0.204	0.077	0.0156	0.182	0.194	0.065	0.0111
NCEP	RESOLUTION	0.1159	0.0302	0.0088	0.0011	0.1139	0.0215	0.0046	-
	RELIABILITY	0.1107	0.0543	0.0214	0.004	0.1061	0.0513	0.0119	-
	UNCERTAINTY	0.1734	0.202	0.0745	0.0158	0.1823	0.1933	0.0643	-
		DECEMBER				MARCH			
SYSTEM	Threshold	1 mm	10 mm	25 mm	50 mm	1 mm	10 mm	25 mm	50 mm
MMENS	RESOLUTION	0.141	0.034	0.004	-	0.121	0.02	0.003	-
	RELIABILITY	0.043	0.038	0.007	-	0.046	0.026	0.004	-
	UNCERTAINTY	0.162	0.207	0.07	-	0.199	0.176	0.06	-
NCEP	RESOLUTION	0.1152	0.0283	0.0109	-	0.1092	0.0171	0.0041	-
	RELIABILITY	0.1043	0.0761	0.0305	-	0.1848	0.0408	0.0096	-
	UNCERTAINTY	0.1605	0.2085	0.0706	-	0.1983	0.1763	0.0588	-

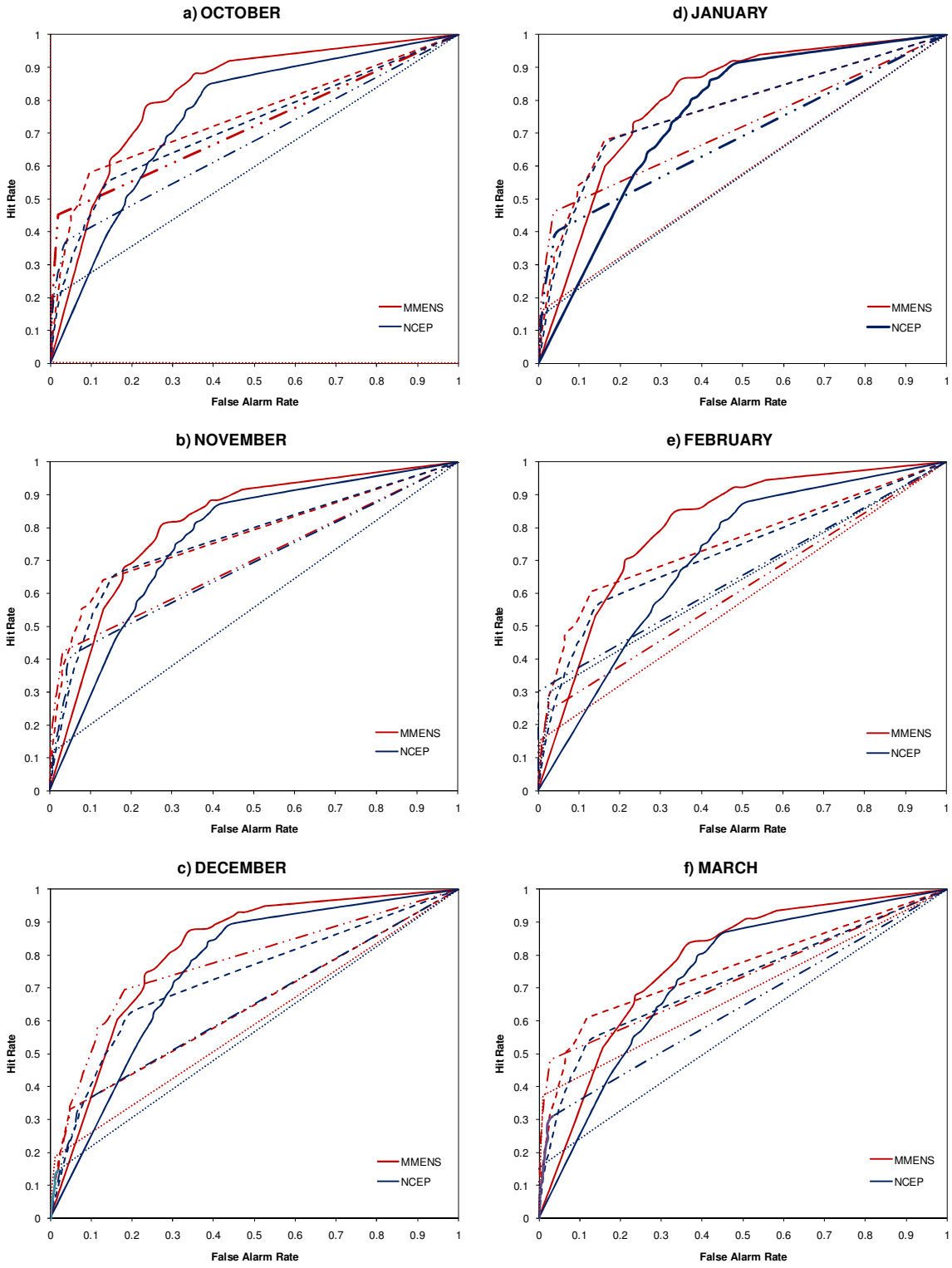
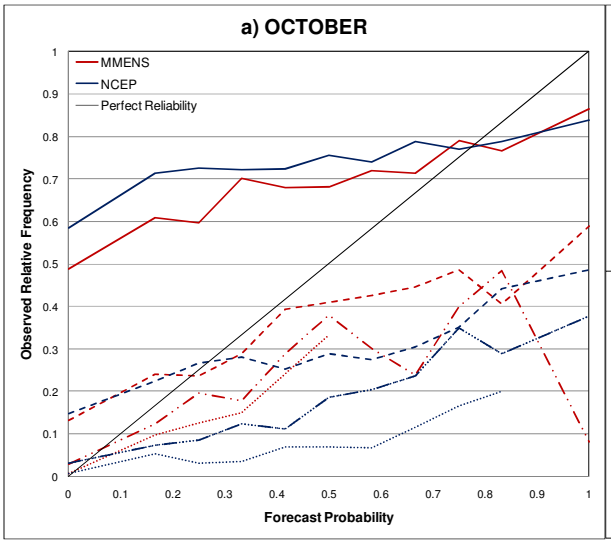
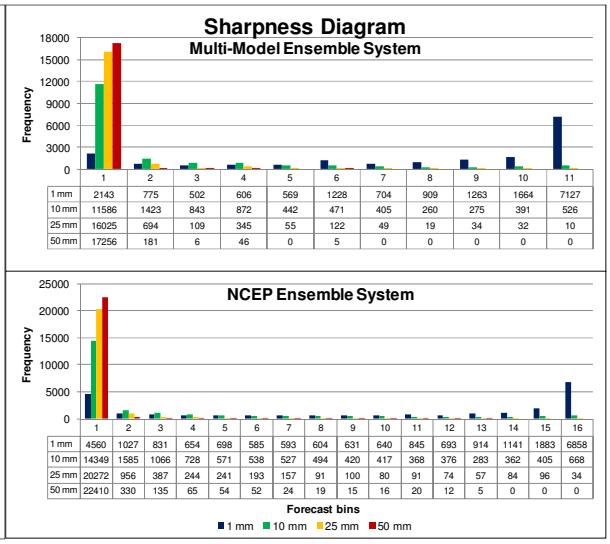


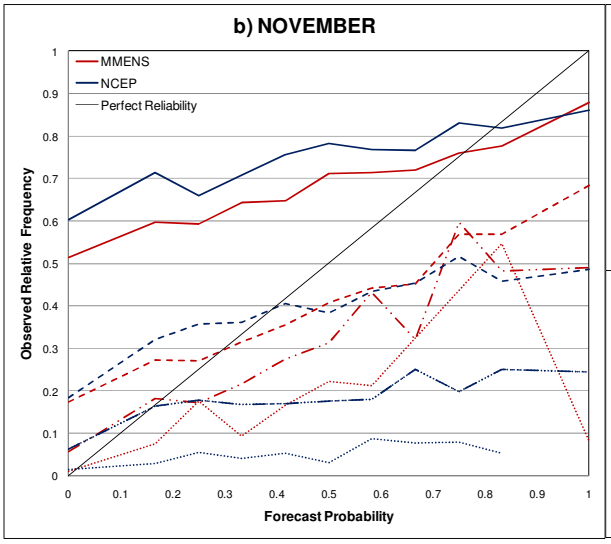
FIGURE 4.18: ROC curves for the MMENS and NCEP systems. (a) October, (b) November, (c) December, (d) January, (e) February and (f) March.

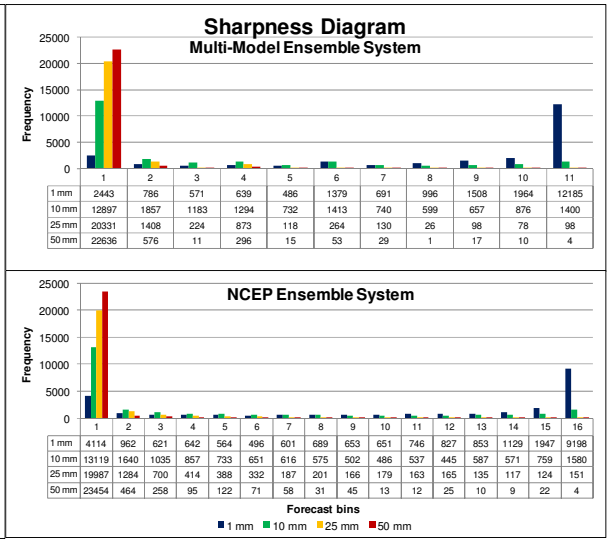




	1	2	3	4	5	6	7	8	9	10	11
1 mm	2143	775	502	606	569	1228	704	909	1263	1664	7127
10 mm	11586	1423	843	872	442	471	405	260	275	391	526
25 mm	16025	694	109	345	55	122	49	19	34	32	10
50 mm	17256	181	6	46	0	5	0	0	0	0	0

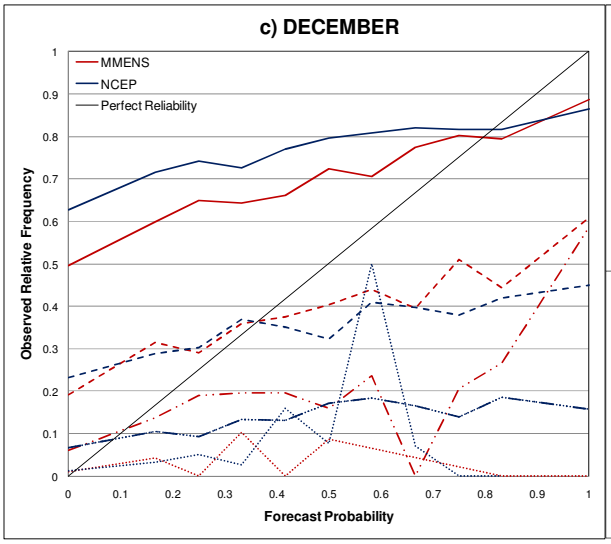
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 mm	4560	1027	831	654	698	585	593	604	631	640	845	693	914	1141	1883	6858
10 mm	14349	1585	1066	728	571	538	527	494	420	417	368	376	283	362	405	668
25 mm	20272	956	387	244	241	193	157	91	100	80	91	74	57	84	96	34
50 mm	22410	330	135	65	54	52	24	19	15	16	20	12	5	0	0	0

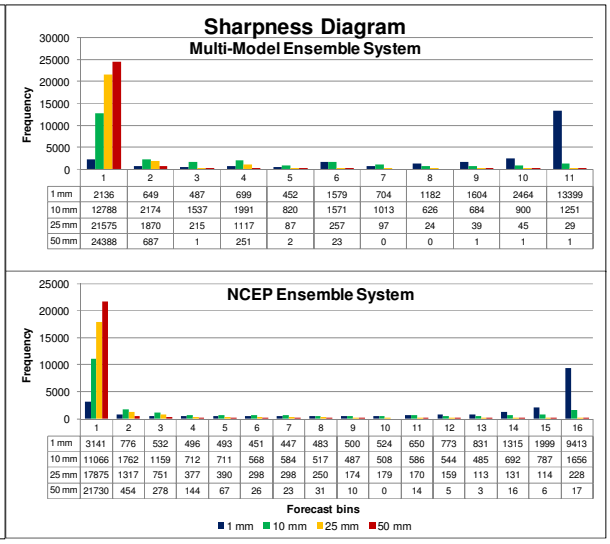




	1	2	3	4	5	6	7	8	9	10	11
1 mm	2443	786	571	639	486	1379	691	996	1508	1964	12185
10 mm	12897	1857	1183	1294	732	1413	740	599	657	876	1400
25 mm	20331	1408	224	873	118	264	130	26	98	78	98
50 mm	22636	576	11	296	15	53	29	1	17	10	4

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 mm	4114	962	621	642	564	496	601	689	653	651	746	827	853	1129	1947	9198
10 mm	13119	1640	1035	857	733	651	616	575	502	486	537	445	587	571	759	1580
25 mm	19987	1284	700	414	388	332	187	201	166	179	163	165	135	117	124	151
50 mm	23454	464	258	95	122	71	58	31	45	13	12	25	10	9	22	4





	1	2	3	4	5	6	7	8	9	10	11
1 mm	2136	649	487	699	452	1579	704	1182	1604	2464	13399
10 mm	12788	2174	1537	1991	820	1571	1013	626	684	900	1251
25 mm	21575	1870	215	1117	87	257	97	24	39	45	29
50 mm	24388	687	1	251	2	23	0	0	1	1	1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 mm	3141	776	532	496	493	451	447	483	500	524	650	773	831	1315	1999	9413
10 mm	11066	1762	1159	712	711	568	584	517	487	508	586	544	485	692	787	1656
25 mm	17875	1317	751	377	390	298	298	250	174	179	170	159	113	131	114	228
50 mm	21730	454	278	144	67	26	23	31	10	0	14	5	3	16	6	17

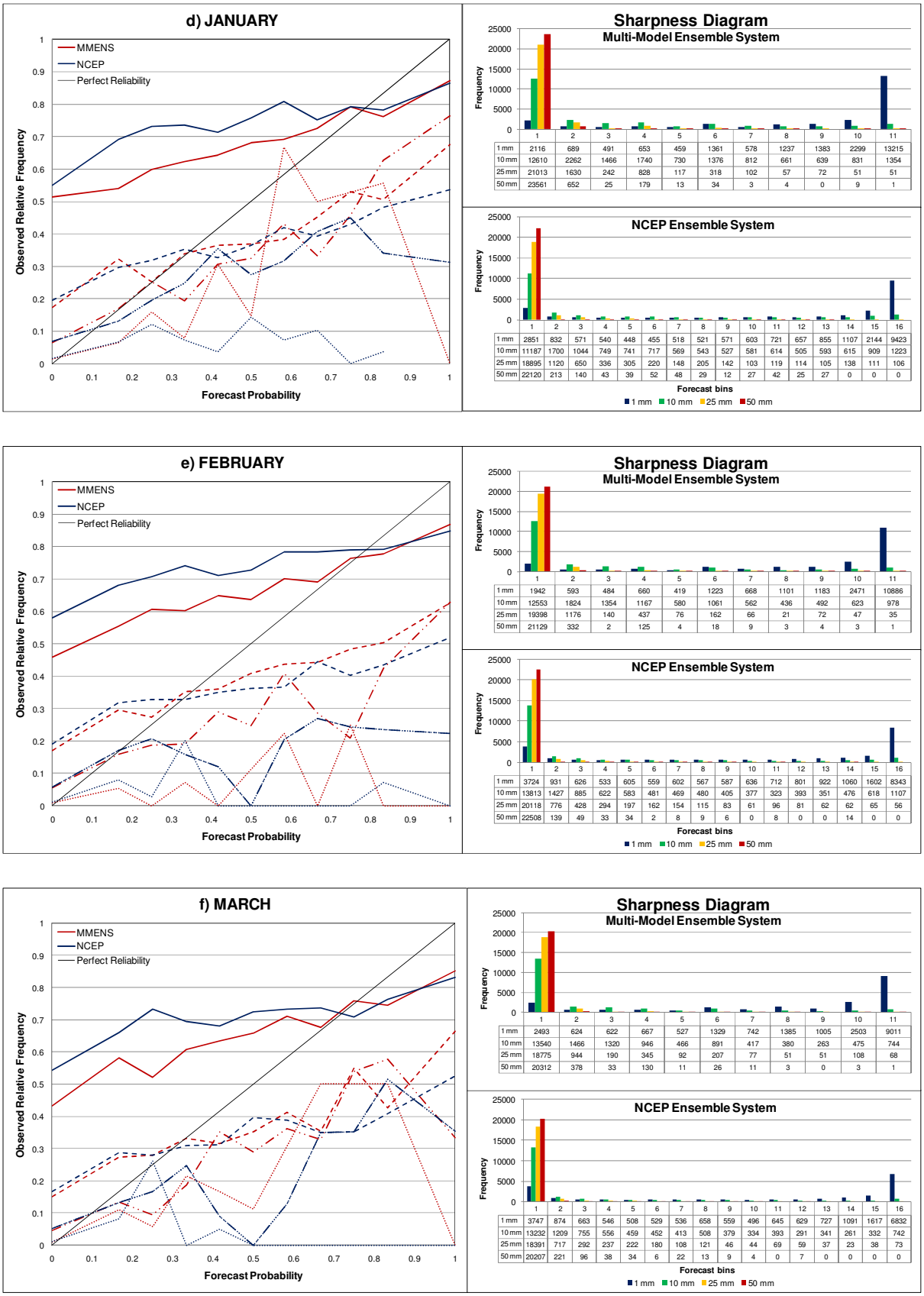


FIGURE 4.19: Reliability and sharpness diagrams for the MMENS and NCEP systems. (a) October, (b) November, (c) December, (d) January, (e) February and (f) March

4.5 SUMMARY

The MMENS outscores and is more skillful than the coarse resolution global model for threshold 10 mm/day and higher. The high resolution MMENS does however have higher area average biases but has a higher probability of detecting events. Similarly, the MMENS has better reliability for events with thresholds greater than 1 mm/day. The MMENS has a tendency to over-forecast the amount and the frequency of 1 mm/day rainfall events whereas the NCEP system is closer to the observed.

This chapter shows that the MMENS is able to perform as well as the NCEP system but has greater reliability and discrimination. This result is significant in that it shows that there is an advantage in the skill of the higher resolution regional member to the MMENS to that of the greater ensemble size of the NCEP system.

CHAPTER 5

SUMMARY AND CONCLUSIONS

A single-model deterministic forecast does not provide guidance on the uncertainty range associated with forecasting the future state of the atmosphere. At the short-range time-scale, these uncertainties stem mainly from the imperfect description of the initial state of the atmosphere, and from the imperfect description of atmospheric dynamics and physics within the numerical model. Until recently, the probabilistic rainfall forecasts issued at the short-range timescale in South Africa have been mainly subjective. These subjective forecasts are interpreted in many different ways by the public, as well as by forecasters themselves. Subjective forecasts are produced by forecasters interpreting the output from numerous numerical forecasts from different national and international centers, as well as real-time observational data available at the time of the forecast being issued. With the advent of ever faster computers in recent years, it has become possible for NMS's to run multiple forecast simulations in parallel for the same forecast period. An investigation of the skill of a short-range ensemble prediction system for South Africa using an objective method of issuing probabilistic rainfall forecasts is therefore not only essential, but has also become feasible.

In this manuscript an objective probabilistic rainfall prediction system for South Africa was developed. This was achieved by obtaining the rainfall forecast of two high-resolution regional models operational in South Africa. The first model is the UM, which is operational at the SAWS. The UM contributed three ensemble members that differ in physics, data assimilation techniques and horisontal resolution. The second model is the CCAM operational at the CSIR, which in turn contributed two members (differing in horisontal resolution) to the ensemble system. A single-model ensemble was constructed for each of the individual models, with each contributing member (three in the case of the UM and two in the case of the CCAM) having equal weights. Finally, the two single-model ensemble systems were combined to form the multi-model ensemble system, to establish whether the skill of this multi-model system outscores that of the single-model ensemble systems. The multi-model ensemble was constructed by assigning equal weights to each of the two contributing single-model ensembles. A single-model ensemble from a coarse horisontal resolution global forecast system was additionally evaluated, to determine a baseline level of

skill the multi-model system should exceed, in order to be considered a value-added product.

The developed multi-model short-range forecast system was rigorously tested over three austral summer half-years (October to March). The main findings are summerised as follows:

A. Quantifying the skill of the individual short-range forecast systems

1. Both the UM and CCAM single-model ensembles have a wet bias over eastern South Africa (with regards to predicting daily rainfall totals). This bias is smaller in spring than in the summer and late-summer months.

2. The CCAM ensemble in general has a smaller bias than the UM ensemble.

3. The individual single-model short-range forecast systems are skillful in predicting daily rainfall over the summer rainfall region of South Africa (using persistence as the reference forecast).

4. Relative to persistence as the reference forecast, the forecast systems have the lowest skill in predicting rainfall for the lowest threshold (that is, the occurrence or non-occurrence of rainfall) and in predicting the occurrence of extreme rainfall events (threshold 50 mm/day). Both the single-model systems have a tendency to over-forecast the lower threshold rainfall events and under-forecast the extreme rainfall events.

5. Using persistence forecasting as reference, the CCAM ensemble is more skillful than the UM ensemble in predicting daily rainfall for thresholds of 10 mm/day and higher, whilst the UM ensemble is more skillful in predicting the occurrence or non-occurrence of rainfall (1 mm/day threshold).

6. The UM ensemble has a greater probability of detecting predefined rainfall events but it also has a higher false alarm rate for the same events (compared to the CCAM ensemble).

7. For the UM, the data assimilation member has the best skill, whereas the 15 km horisontal resolution member has the lowest overall skill (relative to persistence

forecasting). The influence of the low skill of the 15 km member is evident in the skill of the UM ensemble.

8. The CCAM 15 km member is more skillful than the 60 km member in predicting the occurrence or non-occurrence of rain, but for the 10 mm/day and 25 mm/day thresholds, the 60 km forecast is more skillful.

B. Quantifying the skill of the multi-model short-range forecast system

1. The multi-model system has comparable skill to that of the single-model ensemble systems, outscoring the systems for the important threshold of 10 mm/day events.

2. The multi-model system has greater discrimination in rainfall events than any of the single-model ensemble systems.

3. The multi-model system has the highest critical success index average score for rainfall events of 10 mm/day threshold, as well for the 25 mm/day threshold for the months of January and March.

C. Comparing the skill of the multi-model short-range forecast system to a coarse resolution global model

1. The multi-model ensemble system outperforms and is more skillful than the coarse resolution global model for thresholds 10 mm/day and higher.

2. The multi-model ensemble does however have a higher area average bias.

3. The multi-model system has a tendency to over-forecast the amount and the frequency of 1 mm/day rainfall events, especially when compared to the coarse resolution ensemble.

The UM and CCAM short-range numerical weather prediction models have been used to construct a multi-model ensemble prediction system for South Africa, using simple un-weighted averaging. The multi-model system was used to predict the 24-hour rainfall totals for three austral summer half-year seasons of 2006/07 to 2008/09. The forecasts obtained from this system were subsequently compared to observed rainfall data for the same period. From the multi-model system it has been found that the probabilistic forecast has significant

skill in predicting rainfall. The coarse resolution NCEP system has shown to have comparable skill to the multi-model system, but the multi-model system has a better discrimination between events and non-events. Therefore, this multi-model system can be used in an operational environment.

The most significant contribution of this dissertation is that a multi-model short-range ensemble prediction system can be used for South Africa with useful skill. However, for this system to be optimized fully, it will be necessary for the model errors identified in this study to be corrected within each of the ensemble members, before constructing an improved multi-model system. Other possible construction techniques should also be investigated in order to establish the best possible combination of the members for the ensemble system.
Other possible methods include:

1. bias correction of the individual members which would help decrease the quantitative precipitation forecast systematic errors and also improve reliability of probabilistic forecasts;
2. lagged ensemble (include older forecasts) which would increase the precision of the probabilities;
3. combine NCEP members and the MMENS;
4. weight the more accurate ensemble members more. In this way, some of the less skillful members will have a smaller influence on the skill of the multi-model system, and
5. calibrate the probability forecasts to be a function of the forecast rain; i.e. logistic regression.

Another way forward will be to investigate the effect of ensemble size on probabilistic short-range weather prediction versus the increase in horisontal resolution of the individual members.

This study has shown that it is possible to make an objective probabilistic rainfall forecast by constructing a multi-model system with high resolution regional models currently operational in South Africa. The multi-model short-range ensemble prediction system developed can provide forecasters with an objective, probabilistic and skillful rainfall forecast for the following 24-hours over South Africa. Such a system is not currently operational in the country.

REFERENCES

Arribas, A., Robertson, K.B. and Mylne, K.R., 2005: Test of a Poor Man's Ensemble Prediction System for Short-Range Probability Forecasting, *Monthly Weather Review*, **133**, 1825-1839

Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlin. Proc. Geophy,.* **8**; 401-417

Bakhshaii, A. and Stull, R., 2009: Deterministic Ensemble Forecasting Using Gene-Expression Programming. *Weather and Forecasting*, **24**, 1431-1451

Barnston, A.G., Mason, S.J., Goddard, L., DeWitt, D.G. and Zebiak, S.E., 2003: Multimodel ensembling in seasonal climate forecasting at IRI. Bull. *Amer. Meteor. Soc.*, **84**, 1783-1796

Bishop, C.H., Holt, T.R., Nachamkin, J., Chen, S., McLay, J.G., Doyle, J.D. and Thompson, W.T., 2008: Regional Ensemble forecasting Using the Ensemble Transform Technique. *Monthly Weather Review*, **137**, 288-298

Bowler, N.E., Arribas, A. and Mylne, K.R., 2008: The Benefits of Multianalysis and Poor Man's Ensembles, *Monthly Weather Review*, **136**, 4113-4129

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. And Beare, S.E., 2008: The MOGREPS short-range ensemble prediction system. *Q.J.R. Meteorol. Soc,* **134**, 703-722

Bright, D.R. and Mullen, S.L., 2002: Short-Range Ensemble Forecasts of Precipitation during the Southwest Monsoon. *Weather and Forecasting.***17**, 1080-1100

Clark, A.J., Gallus, W.A. and Chen, T., 2008: Contributions of Mixed Physics versus Perturbed Initial/Lateral Boundary Conditions to Ensemble-Based Precipitation Forecast Skill. *Monthly Weather Review*, **136**, 2140-2156

Clark, A.J., Gallus Jr, W.A. and Chen, T., 2008: Contributions of Mixed Physics versus Perturbed Initial/Lateral Boundary Conditions to Ensemble-Based Precipitation Forecast Skill. *Monthly Weather Review*, **136**, 2140-2156

Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawsom, M.H., Staniforth, A., White, A.A. and Wood, N., 2005: A new dynamical core for the MetOffice's global and regional modelling of the atmosphere. *Q. J. R., Meteor. Soc.,* **131***, 1759-1782*

De Elía, R., Laprise, R. and Denis, B., 2002: Forecasting Skill Limits of Nested, Limited-Are Models: A Perfect-Model Approach. *Monthly Weather Review,* **130**, 2006-2023

Du, J., Mullen, S.L. and Sanders, S., 1997: Short-Range Ensemble Forecasting of Quantitative Precipitation. *Monthly Weather Review*, **125**, 2427-2459

Ebert, E. E. and McBride, J.L., 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology,* **239***,* 179-202

Ebert, E. E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Monthly Weather Review,* **129**, 2461-2480

Engelbrecht, F.A., McGregor, J.L., and Rautenbach, C.J. deW., 2007: On the development of a new nonhydrostatic atmospheric model in South Africa. *S.A. J. Sci.*, **103**, 127-134

Engelbrecht F.A., McGregor J.L. and Engelbrecht C.J. (2009). Dynamics of the conformal-cubic atmospheric model projected climate-change signal over southern Africa. International Journal of Climatology **29**, 1013-1033. DOI: 10/1002/joc.1742. 29 1013-1033.

Engelbrecht C.J., Engelbrecht F.A. and Dyson L.L. (2011). Projected changes in closed-low frequencies and extreme rainfall events over southern Africa. International Journal of Climatology. Accepted.

Eckel, F.A. and Mass, C.F., 2005: Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting. *Weather and Forecasting*, **20**, 328-350

Fauchereau, N., Pohl, B., Reason, C.J.C., Rouault, M. and Richard, Y.,2008: Recurrent daily OLR patterns in the Southern Africa/Southwest Indian Ocean region, implications for South African rainfall and teleconnections. *Clim Dyn*. DOI 10.1007/s00382-008-0426-2

Fawcett, R., 2008: Verification Techniques and Simple Theoretical Forecast Models. *Weather and Forecasting*, **23**, 1049-1068

Friederichs, P. and Hense, A., 2008: A Probabilistic Forecast Approach for Daily Precipitation Totals, *Weather and Forecasting.* **23**, 659-673

Ghile, Y.B., and Schulze, R.E., 2009: Evaluation of Three Numerical Weather Prediction Models for Short and Medium Range Agrihydrology Applications. *Water Resource Management*, **24**, 1005-1028

Grimit, E.P and Mass, C.F., 2002: Initial Result of a Mesoscale Short-Range Ensemble Forecasting System over the Pacific Northwest, *Weather and Forecasting*, **17**, 192-205

Hamill, T.M., and Colucci, S.J., 1997: Verification of Eta-RSM Short-Range Ensemble Forecasts. *Monthly Weather Review*, **125**, 1312-1327

Harrison M.S.J., 1984: A generalized classification of South African rain-bearing synoptic systems, *J. of Climatol.* **4**: 547-560.

He, Y., Wetterhall, F., Cloke, H.L., Pappanberger, F., Wilson, M., Freer, J. and McGregor, G., 2009: Tracking the Uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteorological Applications,* **16,** 91-101

Hoffman, R.N. And Kalnay, E., 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100-118

Joliffe, I.T. and Stephenson, D.B., 2003: Forecast Verification: A Practitioner's Guide in Atmospheric Sciences. John Wiley & Sons Ltd, England, 1-1630pp

Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, UK

Krishnamurti, T.N., Kishtawal, C.M., Zhang, Z., Larow, T., Bachiochi, D., Williford, E., Gadgil, S. and Surendran, S., 2000: Mulitmodel Ensemble Forecasts for Weather and Seasonal Climate. *Journal of Climate,* **13**, 4196-4216

Landman, S., Engelbrecht, F.A. and Engelbrecht, C.J., Landman, W.A., Dyson, L., 2010: A Multi-Model Ensemble System for Short-Range Weather Prediction in South Africa, 26th Annual Conference of the South African Society for Atmospheric Sciences, September 2010, Gariep Dam. ISBN 978-0-620-47333-0

Landman, W.A., and Beraki, A., 2010: Multi-model forecast skill for midsummer rainfall over southern Africa. *International Journal of Climatology*, DOI: 10.1002/joc.2273.

Leith, C.E., 1974: Theoretical Skill of Monte Carlo Forecasts. *Monthly Weather Review,* **102**, 409-418

Lewis, J.M., 2005: Roots of Ensemble Forecasting. *Monthly Weather Review*, **133**, 1865-1885

Lorenz, E.N., 1963: Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, **20**, 130-141

Mason, S.J., 2004: On Using "Climatology" as a Reference Strategy in the Brier and Ranked Probability Skill Scores. *Notes and Correspondence*, **132**, 1891-1895

McGregor, J.L., 2003: A new convection scheme using simple closure. *BMRC Research Report*, **93**, 36-36

McGregor, J.L, 2005: Geostrophic adjustment for reversibly staggered grids. *Monthly Weather Review,* **133**, 1119-1128

Palmer, T.N., Allessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delecluse, P., Deque, M., Diez, E., Doblas-Reyes, F.J., Feddersen, H., Graham, R., Gualdi, S., Gueremy, F.J., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazer, A., Maisonnave, E., Marletto, V., Morse, A.P., Orfila, B., Rogel, P., Terres, J.M. and Thomson, M.C., 2004: Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER). Unpublished internal report, Technical Memorandum No 434, European Centre for Medium-Range Weather Forecasts, England

Peel, S. and Wilson, L.J., 2008: A Diagnostic Verification of the Precipitation Forecasts Produced by the Canadian Ensemble Prediction System. *Weather and Forecasting*. **23**, 569-616

Potgieter, C.J., 2007: Short-range weather forecasting over southern Africa with the conformal-cubic atmospheric model. Unpublished Msc. Dissertation, University of Pretoria, Pretoria

Richardson, D.S., 2001: Ensembles using multiple models and analyses. *Q.J.R. Meteorol. Soc.*, **27**: 1847-1864

Rotstayn, L.D., 1997: A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models, I: Description and evaluation of the microphysical processes. *Quart. J. Roy. Meteor. Soc.*, **123**, 1227-1282

Roy Bhowmik, S.K. and Durai, V.R., 2010: Application of multimodel ensemble techniques for real time district level rainfall forecasts in short range time scale over Indian region. *Meteorol. Atmos. Phys.*, **106**, 19-35

Ruiz, J., Saulo, C. and Kalnay, E., 2009: Comparison of Methods Used to Generate Probabilistic Quantitative Precipitation Forecasts over South America. *Weather and Forecasting*, **24**, 319-336

Santoz-Muños, D., Martin, M.L., Morata, A., Valero, A. and Pascual, A., 2010: Verification of a short-range ensemble precipitation prediction system over Iberia. *Adv. Geosci.* **25**, 55-63

Schulze, R.E, Maharaj, M., Warburton, M.L., Gers, C.J., Horan, M.J.C., Kunz, R.P, and Clark, D.J., 2008: South African Atlas of Climatology and Agrohydrology.

Schulze, R.E., and Maharaj, M., 2007: Rainfall Seasonality. In: Schulze, R.E. (Ed). 2007, South African Atlas of Climatology and Agrohydrology, Water Research Commission, Pretoria, RSA, WRC Report 1489/1/06, Section 6.5

Singleton, A.T. and Reason, C.J., 2007: Variability in the characteristics of cut-off low pressure systems over subtropical southern Africa. *Int. J. Clim.*, **3**, 295-310

Stael von Holstein, C.S., 1971: An Experiment in Probabilistic Weather Forecasting. *J. Applied Meteor.*, **10**, 635-645

Stanski, H.R., Wilson, L.J. And Burrows, W.R., 1990: Survey of common verification methods in meteorology. WMO World Weather Watch Tech. Rep. 8 WMO TD 358, 114 pp.

Stensrud, D.J. and Yussouf, N., 2005: Bias-corrected short-range ensemble forecasts of near surface variables. *Meteor. Appl.*, **12**, 217-230

Stensrud, D,J., Brooks, H,E., Du, J., Tracton, S. and Rogers, E., 1999: Using Ensembles for Short-Range Forecasting. *Monthly Weather Review*, **127**, 433-446

Tadross, M.A., Gutowski Jr, W.J., Hewitson, B.C., Jack, C. and New, M., 2006: MM5 simulations of interannual change and the diurnal cycle of southern African regional climate. *Theor. Appl. Climatol.* **86**, 63-80

Taljaard, J.J., 1996: Atmospheric circulation systems, synoptic climatology and weather phenomena of South Africa. Part 6, Rainfall in South Africa. *Technical Report*; no 32. South Africa. Weather Bureau, Govt. Printer, Pretoria, 98p

Tennant, W,J., Toth, Z. and Rae, K.J., 2007: Application of the NCEP Ensemble Prediction Syustem to Medium-Range Forecasting in South Africa: New Products, Benefits, and Challenges. *Weather and Forecasting*, **22**, 18-35

Theis, S.E., Hense, A. and Damrath, U., 2005: Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteor. Appl.*, **12**, 257-268

Todd, M. and Washington, R., 1998: Extreme daily rainfall in southern Africa and souhtwest Indian Ocean tropical-temperate links. S. Afr. J. Sci., 94, 64-70

Toth, Z. and Kalnay, E., 1993: Ensemble Forecasting at NMC: The Generation of Perturbations. *Bulletin of the American Meteorological Society,* **74**, 2317-2330

Toth, Z., Xhu, Y. and Marchok, T., 2001: The Use of Ensembles to Identify Forecasts with Small and Large Uncertainties. *Weather and Forecasting*, **16**, 463-477

Tyson, P.D., 1986: Climate Change and Variability over southern Africa. Oxford University Press, Cape Town

Tyson, P.D. and Preston-Whyte, R.A., 1993: The atmosphere and weather of southern Africa. Oxford University Press, New York

Van Heerden, J. and Hurry, L., 1987: Southern Africa's Weather Patterns: An Introductory Guide, Acacia Books, Pretoria

Walker, N.D., 1990: Links between South African summer rainfall and temperature variability of the Agulhas and Benguela current systems. *J. Geophys. Res.*, **B 95**, 3297-3319

Wandishin, M.S., Mullen, S.L., Stensrud, D.J. And Brooks, H.E., 2001: Evaluation of a Short-Range Ensemble System. *Monthly Weather Review*, **129**, 729-747

Wang, W. and Seaman, N.L., 1997: A Comparison Study of Convective Parameterization Schemes in a Mesoscale Model. *Monthly Weather Review,***125**, 252-278

Wei, M., Toth, Z., Wobus, R. and Zhu, Y., 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus (A Dyn. Meteorol. Oceanogr.)*, **60-1**, 62-79

Wilks, D., 2006: Statistical methods in the atmospheric sciences. 2nd Edition, Elsevier Academic Press, California, United States of America

Wilson, D. and Bushell A., 2007: Annexe to Unified Model Documentation Paper No. 29. The PC2 Cloud Scheme, MetOffice, Exeter, UK

Zongjian, K.E., Dong, W. and Zhang, P., 2008: Multimodel Ensemble Forecasts for Precipitation in China in 1998. *Advance in Atmospheric Science*, **25**: 72-82