Development and Verification of a Short-Range Ensemble Numerical Weather Prediction System for Southern Africa

by
Ruth Jean Park

Submitted in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

in the
Faculty of Natural and Agricultural Sciences
University of Pretoria

February 2014

Development and Verification of a Short-Range Ensemble Numerical Weather Prediction System for Southern Africa

Ruth Jean Park

Supervisor: Prof. Willem Landman

Department: Department of Geography, Geoinformatics and Meteorology

Faculty: Faculty of Natural and Agricultural Sciences

University: University of Pretoria

Degree: Master of Science

Abstract

This research has been conducted in order to develop a short-range ensemble numerical weather prediction system over southern Africa using the Conformal-Cubic Atmospheric Model (CCAM). An ensemble prediction system (EPS) combines several individual weather model setups into an average forecast system where each member contributes to the final weather forecast. Four different EPSs were configured and rainfall forecasts simulated for seven days ahead for the summer months of January and February, 2009 and 2010, for high (15 km) and low (50 km) resolution over the southern African domain. Statistical analysis was performed on the forecasts so as to determine which EPS was the most skilful at simulating rainfall. Measurements that were used to determine the skill of the EPSs were: reliability diagrams, relative operating characteristics, the Brier skill score and the root mean square error. The results show that the largest ensemble is consistently the most skilful for all forecasts for both the high and the low resolution cases. The higher resolution forecasts were also seen to be more skilful than the forecasts made at the low resolution. These findings conclude that the largest ensemble at high resolution is the best system to predict rainfall over southern Africa using the CCAM.

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

Preface

The field of short-range numerical weather prediction in South Africa has generally considered few integrations at a fairly coarse spatial resolution to be standard practice. In order to continue to improve the quality of weather forecasts being produced nationally, ensemble prediction systems run at higher resolution and of larger size than what is considered normal must be developed. This thesis focuses on developing a large ensemble, high resolution short-range forecast system that produces skilful precipitation forecasts over the the southern African domain.

Precipitation is a significant weather phenomenon that affects the entire population, especially the poverty-stricken and rural regions, and the largely agriculture-based economy of southern Africa. Since precipitation has such an impact on society, it is paramount that precipitation forecasts be skilful in order to prevent possible adverse effects of precipitation events. Flooding and heavy rainfall have far-reaching consequences financially and socially, with the potential for even loss of life. In order to improve the quality of precipitation forecasts made for southern Africa, new short-range weather forecasting techniques must be explored. This research aims to improve the quality of precipitation forecasts by focusing on spatial resolution and ensemble size.

The hypotheses to be tested are:

1. Whether a larger ensemble prediction system is more skilful at predicting 7-day precipitation over southern Africa than one with fewer members.
2. Whether a high resolution ensemble prediction system produces more successful 7-day precipitation forecasts over southern Africa than a coarse resolution one.

In order to test these hypotheses, the following steps must be taken:

1. Eight ensemble members must be generated.
2. Four ensemble prediction systems must be configured.
3. The four ensemble prediction systems must be run to simulate precipitation at fine and coarse spatial resolution over southern Africa.
4. Verify the four ensemble prediction systems against observational precipitation data deterministically and probabilistically

This dissertation contains five chapters. Chapter 1 describes the weather and climate of southern Africa, an overview of numerical prediction and ensemble forecasting. This is followed by a discussion of precipitation forecasting techniques and an analysis of weather modelling in South Africa. The methodology of the research and details of the data is discussed in Chapter 2, which also includes a description of the verification methods employed. Chapter 3 and Chapter 4 details the results of the deterministic and probabilistic verification. The final results are discussed, and conclusions are drawn in Chapter 5, along with recommendations for future research.

Acknowledgments

- I would like to acknowledge my supervisor Prof. Willem Landman for his generosity, understanding and endless guidance for without him this work would not have been possible.
- I would like to express my thanks to Dr. Francois Engelbrecht who taught me all I needed to know about the CCAM system, and for his patience.
- Thank you to Dr. John McGregor and Dr. Marcus Thatcher for being so welcoming and generous with their data and help.
- My family, for their unwavering support and faith in me.
- Ivo Donev for his long-suffering patience, optimism, and sense of humour.
- Lastly I would like to acknowledge my five dogs and kitten for their unconditional love and devotion. To err is human, to forgive is canine.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Southern African Weather and Climate

Southern Africa is positioned in the sub-Saharan region of the African continent and includes several different climatic regions ranging from semi-tropical in the north (Angola, Malawi, Zambia, Mozambique and Madagascar); to semi-arid in South Africa, Botswana and Zimbabwe; to arid in Namibia (Davis, 2010; Tyson & Preston-Whyte, 2000). Southern Africa receives most of its rainfall during summer from convective systems and mid-latitude cyclones from the east, and these can bring heavy rainfall and large storm systems (Cook *et al.*, 2004). In contrast, the westerly side of southern Africa remains dry most of the year with the exception of the southern Cape areas in South Africa which experience winter rainfall from cold front systems (Harrison, 1984).

Namibia and the west coast of South Africa are desert and semi-desert regions, and experience little rain throughout the year. These arid conditions are caused by the sub-Tropical High Pressure Belt, or Subtropical Ridge, as well as the north-flowing, cold upwelling in the Benguela current in the coastal Atlantic Ocean, both of which cause very

low rainfall (Engelbrecht & Landman, 2011). The high pressure systems that dominate the region cause sinking air which results in clear, sunny conditions and the anti-cyclonic motion also causes hot, dry winds the flow from the arid inland toward the coast which account for low levels of humidity which makes rainfall unlikely (Garstang *et. al.*, 1996). Botswana also has a semi-arid climate which is caused by the presiding high pressure belt. Although Botswana does not have a coastline and therefore the climate is not influenced by the Benguela Current, the high pressure systems cause wide-spread dry and sunny conditions with very little rainfall, and often drought (Bhalotra, 1987; Garstang *et. al.*, 1996).

The South African western and southern Cape coast is dominated by winter frontal weather systems, resulting in dry summer months and wet winter months caused by antarctic cold fronts moving eastward over the Atlantic Ocean (Schultz, 2005). This region experiences a more Mediterranean climate and has dry and hot, summers. This is due to inland winds blowing from the semi-desert regions toward the coast, which causes low humidity and hot, sunny conditions. The South-Atlantic high pressure system also influences the Cape summer weather, causing strong, dry south-easterly winds to blow over the western Cape (Davis, 2011).The eastern Cape coast receives rainfall all year round, and lies between the Mediterranean climate of the west and south Cape, and the sub-tropical climate of the Indian Ocean coast (Engelbrecht & Landman, 2011). During summer the eastern Cape coastline experiences sub-tropical rainfall from the warm upwelling in the south-west Indian Ocean Agulhus Current, which causes convective rainfall over the region, and sometimes heavy rainfall and flooding. This area can also receive rain during winter from frontal systems that sweep along the coastline, as well as extremely cold conditions and snow inland.

The northern countries (Angola, Zambia, Malawi, Mozambique) receive rain throughout the year, with especially heavy rain during the summer months. This climate is due to tropical convective systems that develop over the equatorial tropical regions of central Africa. The Inter-Tropical Convergence Zone (ITCZ) which is a band 5^{o} around the equa-

tor where the north-east and south-east trade winds converge and cause a low-pressure zone and extreme convection. Southern Africa is most influenced by the ITCZ during summer when it moves south of the equator (Davis, 2011). The deep convective activity results in large storm systems that bring heavy rainfall to Angola, Zambia, Malawi, and the northern parts of Mozambique, Botswana and Zimbabwe. The ITCZ contributes to the annual flooding of the Okavango Delta in Botswana, but the presence of a high-pressure system over Botswana pushes the ITCZ away which results in prevailing dry conditions throughout Botswana and parts of Zimbabwe, often leading to drought (Batisani, 2011).

The climate of Madagascar is dominated by orographic rainfall caused by the mountain ranges in the east, and tropical cyclones that develop in the Indian Ocean and move westward. These tropical cyclones become mid-latitude cyclones when they enter the mid-latitude regions. These cyclones are present mostly during the spring and summer months, and can result in extremely heavy rainfall over Madagascar, also traveling as far westward as the Mozambican coast. Mid-latitude cyclones can even move inland, extending into north-eastern South Africa (Reason & Kiebel, 2004; Blamey & Reason, 2012). And example of this is the Cyclone Leon-Eline, that landed in Madagascar on the 17th of February 2000, crossed Madagascar to the Mozambican coast, traveled westward into Mozambique and continued through South Africa and Zimbabwe along the Limpopo River region (Reason & Kiebel, 2004; Vitard *et. al.*, 2003). Figure 1.1 shows the average annual rainfall distribution over southern Africa. Areas of low annual rainfall are located around the desert and semi-desert regions of Namibia, Botswana and South Africa; and the heaviest annual rainfall concentrated near the tropical regions and over Madagascar.

1.1.1 Southern African Summer Rainfall

This research focused on southern Africa during the months of January and February, 2009 and 2010, which are austral summer months. The southern African domain is shown in Figure 1.1.2.

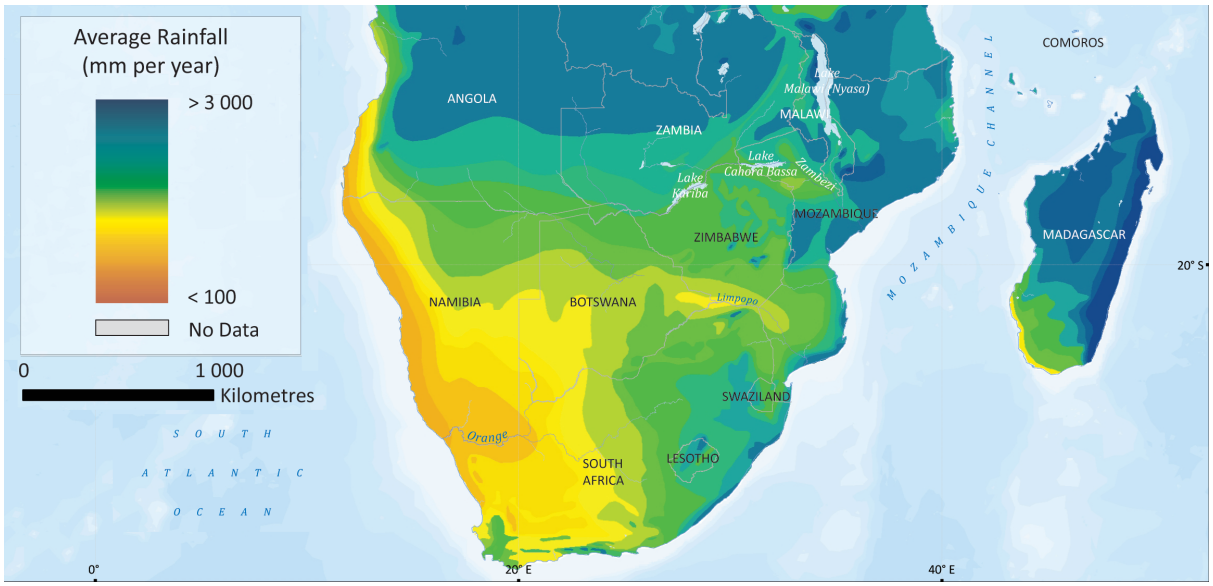

Figure 1.1.1: Southern African Annual Rainfall (UNEP, 2010)

South African summers have characteristic rainfall patterns. The Cape coastal and interior regions remain dry and receive very little rain then – this is because most Cape rain is due to Southern Ocean frontal systems which are not present during the summer months, as they shift further south. The western interior remains arid, as it receives very little rain throughout the year. The east coast of South Africa as well as the Lowveld receive heavy rain and are very humid as a result of on-shore flow from warm Indian Ocean systems. Summer rainfall along the eastern escarpment is characterised by orographic rain, as warm air containing water vapour is forced upward and cooled due to the up-slope flow brought about by ridging atmospheric circulation (Hastenrath *et. al.*, 1995; Tyson &

Preston-Whyte, 2000). An example of orographic rain phenomena is the summer rainfall over the mountainous eastern region of South Africa and Lesotho caused by up-slope flow of moist air up the steep slopes of the Drakensberg. Unstable atmospheric conditions and convective low-pressure cells are prevalent throughout the Highveld interior (Engelbrecht *et. al.*, 2012). These mesoscale convective complexes cause short-lived, fast moving thunderstorm systems (Blamey & Reason, 2011; Eckel & Mass, 2005).

Further north into Zimbabwe, northern Botswana, Malawi, Zambia and Angola, summer rainfall patterns are dominated by the Inter-Tropical Convergence Zone (ITCZ), which moves south of the equator during this time. The ITCZ brings heavy tropical rain to these countries, and can even extend down into South Africa due to deepening low pressure troughs, bringing tropical rain to the Limpopo province. The ITCZ can also form a Tropical Temperate Trough (TTT), where low pressure troughs that extend across South Africa join the tail end of the ITCZ forming one long, deep trough from the tropics down to the temperate climate of South Africa. TTTs are the most dominant source of convective summer rainfall over southern Africa (Todd, 2004; Todd & Washington, 1999).

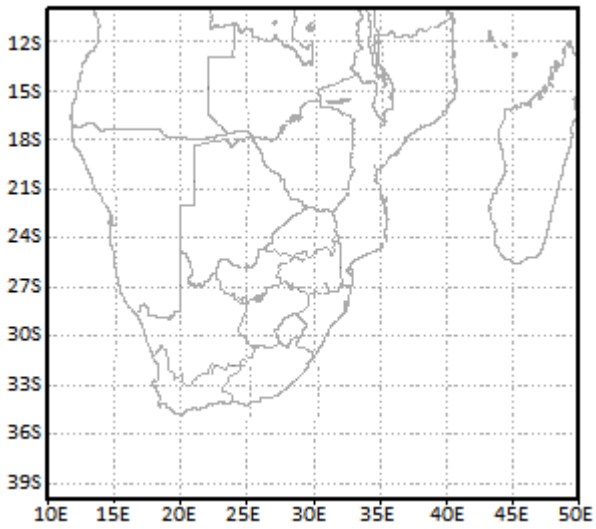

Figure 1.1.2: SADC Domain

It has been well documented that the El Niño Southern Oscillation (ENSO) has a strong influence on the Indian Ocean and southern African climate (Cook, 2001; Reason *et. al.*, 2000; Landman & Beraki, 2012) . ENSO is a measure of sea surface temperature (SST) anomalies in the tropical Pacific Ocean, as well as surface pressure increase over the eastern Pacific and the Indian Ocean. A warming in SSTs in the western Pacific combined with a rise in surface pressure, transporting rainfall toward the eastern Pacific indicates an El Niño episode (NCEP, FAQ; Trenberth *et. al.*, 2007; IPCC Fourth Amendment, 2007). These conditions result in anomalously dry conditions for the eastern southern African area during the summer season. Alternatively, La Niña is associated with cooling of the SSTs in the western Pacific, with drier, cooler conditions extending into the tropical eastern Pacific. This causes lower pressure over the Indian Ocean which results in wetter than normal conditions over the east of southern Africa during the summer months.

The ability to accurately predict rainfall events in the summer season over southern Africa is important as well as strategic. In areas adjacent to the Indian Ocean and Mozambican Channel there is strong onshore flow of warm, moist air that causes considerable rainfall and floods over these regions. From a social aspect, these coastal regions are often found to be impoverished (Mozambique, South Africa, Madagascar and even as far inland as Zimbabwe), and do not have the resources or capacity to be able to manage destructive rainfall events. In order to help government or other involved parties assist these communities in enduring these potentially devastating phenomena, there need to be accurate weather forecasts that can capture future flood and extreme episodes of precipitation.

Accurate summer rainfall forecasts are able to help impoverished regions avoid destruction of property, possible drowning deaths and community members being left homeless. More knowledge regarding summer rainfall events can capacitate vulnerable areas to become aware of and be able to put in place protective measures for future destructive weather.

1.2 Numerical Weather Prediction

Numerical Weather Prediction (NWP) is a process where computational mathematical models are used to simulate weather patterns, based on fundamental physical principles of the atmosphere (The History of Numerical Weather Prediction, NOAA, 2012). Sets of physical equations are used to represent ideal atmospheric behaviour. NWP models are used to predict short-range future weather phenomena by having reliable initial conditions to run the model, as well as an accurate representation of the atmosphere (Kalnay, 2003).

An accurate description of the state of the atmosphere is incredibly difficult to model since the atmosphere is inherently chaotic, and constantly changing (Lorenz, 1963). As stated in Kalnay, 2003: "No simple set of causal relationships can be found which relate the state of the atmosphere at one instant of time to its state at another". There are numerous variables to take into consideration when describing the atmosphere, and since this is the case there is a huge amount of information and extremely large data sets that have to processed by these NWP models in order to predict future atmospheric states. For this reason the analysis thereof requires super-computers due to their computational power and capacity to manipulate such a vast quantity of data.

NWP was originally developed in order to resolve the fundamental atmospheric equations known to dictate weather phenomena; for example: the hydrodynamic and thermodynamic equations. In 1940, Rossby published a paper analysing his barotropic vorticity equation, which can be considered the first major development of a numerical weather prediction system (Wiin-Nielsen, 1991). Soon after, in 1947, Charney pioneered a simple quasi-geostrophic model that would become part of the basis of NWP. The National Centers for Environmental Prediction (NCEP) has, since 1955, been producing short-range weather forecasts using NWPs (Kalnay, 2003; Wiin-Nielsen, 1991). NWPs function best at high resolution, since an NWP model becomes more reliable with increased spatial

resolution (Kalnay, 2003). Since then, other meteorological centers have been developing sophisticated NWP models. For example: the Unified Model (UM) of the Met Office has been developed to produce high-resolution NWP forecasts (Davies *et. al.*, 2005).

1.2.1 Ensemble Prediction Systems

Ensemble prediction systems (EPS) are considered to be an effective way to probabilistically forecast a range of possible future states of the atmosphere, given its current state; while also addressing the problems of uncertainties in the initial conditions, and inaccuracies in the model approximations (Buizza, *et. al.*, 1999; Du, 2007; Tennant *et. al.*, 2006). Monte Carlo analysis is used in ensemble forecasting because Monte Carlo simulations are good at modelling phenomena with large uncertainty in the input (Leith, 1974). In the case of ensemble forecasting, Monte Carlo methods are particularly useful because they are able to overcome the uncertainties in the initial conditions and produce forecasts as well as the mean and variance thereof (Houtekamer & Derome, 1995).

EPS were introduced in order to allow forecasters to produce reliable forecasts with lead times of greater than a few days. As opposed to deterministic forecasts which is a process of predicting a single output, probabilistic predictions present a range of possible future outcomes, all intialised by the different initial conditions (Du *et. al.*, 1997). Therefore, probabilistic forecasts are more likely to capture specific events than deterministic forecasts, since they are more accommodating of possible outcomes. This is especially important when it comes to predicting rainfall.

Deterministic rainfall forecasts are reliable up to a 3-day lead time (World Meteorological Organisation, 2012). Forecasts made for more than two days ahead in time will begin to be influenced by the uncertainty inherent in the atmosphere, and this uncertainty is not captured in deterministic forecasting. They can be presented in binary form (will it

rain? – yes or no) or categorical form (how much will it rain? – 0-5 mm, 5-9mm, etc.). From this it can be seen that even if the observed rainfall differs slightly from what is forecast, the deterministic forecast will still have failed.

Probabilistic rainfall predictions, on the other hand, do not produce forecasts of exact nature (i.e.: 'how much will it rain'), but rather represent the likelihood of a rainfall event. Even when forecasts are being made for more than two days ahead, probabilistic rainfall forecasts can still be considered reliable. This is because probabilistic forecasts represent the likelihood of a range of possible outcomes to occur, thus providing a forecast that accommodates uncertainties in the initial conditions, and model errors.

The development of EPSs began to contribute significantly when the National Centers for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasts (ECMWF) began to incorporated previous research (Leith, 1974) by moving away from producing only high-resolution deterministic forecasts to a combination of this and a new probabilistic ensemble prediction systems (Buizza & Palmer, 1997). Since then, the Met Office Global and Regional Ensemble Prediction System (MOGREPS) model and the National Center for Atmospheric Research (NCAR) model (to mention a few) have been developed to produce successful short-range ensemble forecasts (Bowler *et. al.*, 2008; Du *et. al.*, 1997).

The number of members that make up an EPS contributes greatly to the performance of the system (Buizza & Palmer, 1997; Buizza, 2008; Brankovic & Palmer, 1997; Houtekamer & Derome, 1994). Very small ensembles, or "Poor Man's" ensembles, however, are found to be an efficient way to improve the performance of short-range forecasts (Arribas, *et al.*, 2005; Bowler, *et al.*, 2008; Ebert, 2001). Probabilistically, an increase in ensemble size provides a larger spread, a larger distribution of forecast outcomes that all contribute to the final probabilistic EPS forecast. The average forecast of an EPS has been verified to be more skilful than any one of the individual members (Buizza, 1997; Grimitt & Mass, 2002; Hamill & Collucci, 1997). A wider spread of forecasts makes it more likely that the

final EPS forecast will include more extreme forecast values that will result in the model being able to predict extreme weather. Computational capacity imposes a limit on the number of members that can be combined in the EPS since calculating an endless number of members is simply impractical (Kalnay, 2003).

Different ensemble members can be generated through perturbation methods, or by lagged-average methods (Kalnay, 2003). Perturbation methods to increase ensemble size is a relatively simple process. It involves adding small perturbations to the initial conditions, that result in each member producing different forecast values within a range of uncertainty (Ensemble Forecasting at NCEP, 1995). This method can be used to create a large number of forecast members, all with slightly perturbed initial conditions which will result in different forecasts. The lagged method of increasing ensemble size involves initialising each member with more recent observations: so each member is initialised at a different time (hence the name "lagged-average forecasting") and then averaged to form the EPS. However, adding weights to each member in a lagged forecast system does not add skill to the system (Kalnay, 2003) and so the members are simply average to produce the final EPS forecasts. Lagged-average forecasting is equally effective in multi-model and single-model systems (Lawrence & Hansen, 2007). The perturbation breeding method can produce a large number of ensemble members, whereas the lagged-average method can only produce as many forecasts as there are updates in the initial conditions.

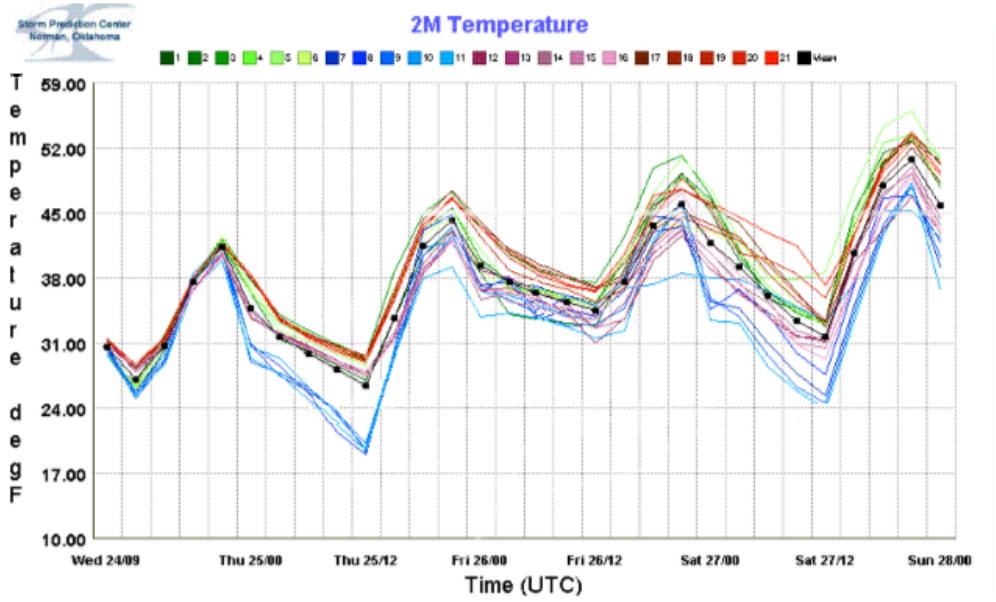

Figure 1.2.1: Short-Range Forecast Plume where the black line is the ensemble mean (NOAA Storm Prediction Center, 2010)

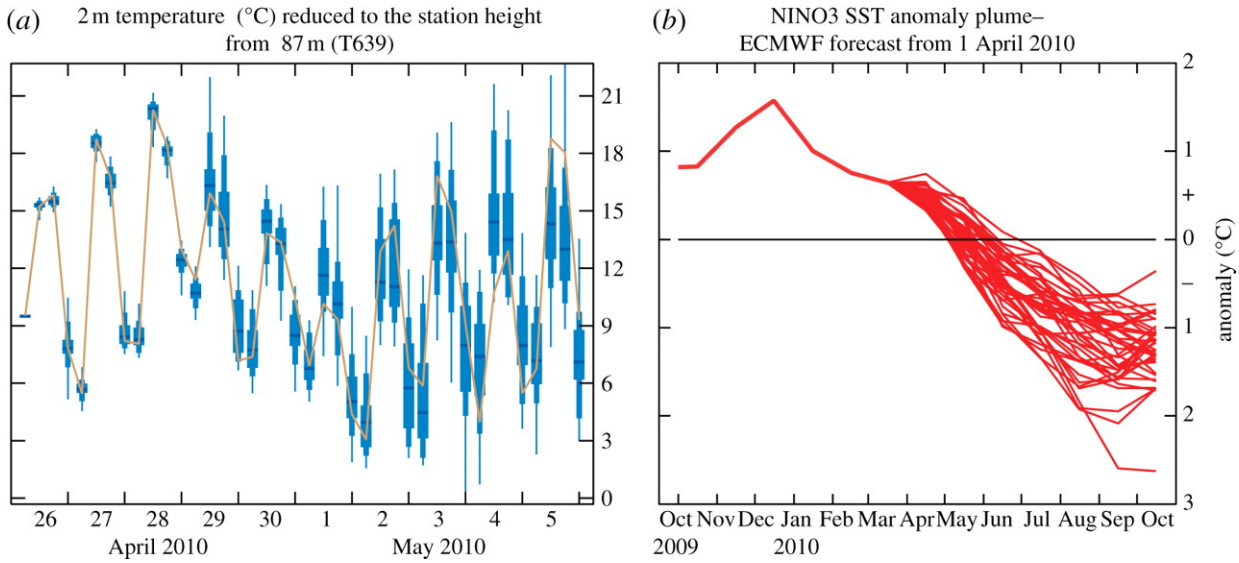

Figure 1.2.2: Medium- and Long-Range forecasts: ECMWF Ensemble Members Forecast Plume & Average EPS Forecast (Palmer & Hardaker, 2011)

The individual members of an EPS produce different forecast projections. At the small time-steps, the individual forecasts are very similar, but as the time-step increases the forecasts begin to diverge. This is because of the differences in the initial conditions or the model configuration. Plotting all the individual projections results in a plume forecast. This plume represents the spread of the EPS by showing all the possible outcomes together on one graph. Figure 1.2.1 is an illustration of a forecast plume for a short-range EPS with 12 hour time-steps. Figure 1.2.2 shows two plume forecasts: the left diagram is of a medium-range EPS and the right diagram is of a long-range EPS.

Short-range ensemble forecasting was developed after seeing encouraging results in global EPSs (Grimitt & Mass, 2001; Hamill & Collucci, 1997). Since ensemble forecasting is capable of overcoming forecast uncertainty, applying ensemble methods to short-range forecasts can help overcome forecast uncertainty in a finer resolution, more detailed domain (Buizza *et. al.*, 1998). Short-range EPSs require a higher resolution than the medium-range; because the forecasts are only over a few days, their purpose is to capture more detailed weather systems, such as daily temperature and thunderstorm events (Bowler *et. al.*, 2008). These systems are initialised with the boundary conditions on the edges of the high-resolution domain.

1.3 Forecasting Precipitation

Predicting quantitative precipitation is important and very useful, yet it is significantly difficult to do so (Fritsch *et. al.*, 1998) and it is generally accepted that deterministic precipitation forecasts have low skill (Olson *et. al.*, 1995). Measurable rainfall forecasts are essential to produce, as they are of great interest to the public and serve a major role in risk and disaster management since they are able to forecasts extreme rainfall and floods. Probabilistic forecasts produced by EPSs do not measure rainfall quantitatively, but rather probabilistically. Because EPSs are shown to surmount part of the problem of forecast uncertainty, probabilistic rainfall forecasts are considered to be more skilful, even though they do not provide numerical measurements of rainfall (Bright & Mullen, 2002). High resolution grids are required to predict rainfall systems in the short range timescales; it has been shown that ensemble rainfall prediction skill increases with an increase in horizontal resolution (Mass & Ovens, 2002).

Small-scale precipitation phenomena need to be predicted at high-resolution to produce accurate simulations thereof (Kalnay, 2003). Convective-type precipitation occurs at a short-range timescale limit and is predicted most effectively when the grid spacing has been reduced to the 1-10 km range (Weisman *et. al.*, 1997; Mass & Ovens, 2002). The same settings should be applied to orographic and topographic rainfall, with an added increase in vertical resolution to accurately simulate the land surface and therefore the rainfall events (Colle & Mass, 2000; Davis *et. al.*, 1999). Frontal rainfall, on the other hand, can be modeled at a resolution of 50 km; although, a resolution of down to 5 km can model the details of the phenomenon.

1.4 Southern African Weather Modelling

Southern African weather and climate can impact the social and economic infrastructure of the region (Fauchereau *et. al.*, 2008; Friedrichs & Hense, 2008). The areas of heavy summer rainfall are particularly susceptible to flooding and other rain damage. The southern parts of South Africa can experience extreme winter rainfall, which also affects living conditions and quality of life as well as small-scale (e.g.: independent businesses) and large-scale (e.g.: insurance companies and food export companies) economic development. Rain-fed crop production is highly sensitive to fluctuations in seasonal rainfall, and therefore any anomalous rainfall events can be detrimental to crop production and food resources. In order to prevent or mitigate the effects of weather damage, accurate and reliable weather forecasts need to be produced in order to supply decision-makers, farmers and other affected parties with these forecasts regularly.

Accurate weather models are critical in forecasting short-range weather. The ability to predict the weather with skill also depends on the computational capacity available to run the NWP models. The more computational power available, the more detailed NWP models can be run. And the more detailed the models, the more detailed the forecasts.

South African atmospheric modelling began in the 1960's, with the development of the first South African numerical weather model (Triegaardt, 1965). This model became operational in the 1970's but was very limited due to the lack of computational power (Riphagen & Burger, 1986). In the 1980's, a new model was developed that was far more advanced and was found through verification to be equivalent to other international numerical weather models at the time (Riphagen, 1985; Riphagen & Burger, 1986). In the 1990's there was major development in the field of numerical weather prediction internationally, in particular the National Centers for Environmental Prediction's Eta model, which was adopted by the South African meteorologists to produce weather forecasts

(Riphagen *et. al.*, 1996). Recently there has been a surge in South African weather modelling, with institutions such as the South African Weather Service (SAWS), the University of Pretoria (UP), the University of Cape Town (UCT) and the Council of Scientific and Industrial Research (CSIR) implementing international models and pursuing model development research (Engelbrecht *et. al.*, 2007; Rautenbach, 2003; Singleton & Reason, 2006; Landman *et. al.*, 2005).

Several short-range weather models have been, and are, used to forecast southern African weather. The Conformal-Cubic Atmospheric Model (CCAM) is used at the Council of Scientific and Industrial Research (CSIR) to produce short-range forecasts , seasonal forecasts, and climate change projections (Engelbrecht *et. al.*, 2011; Engelbrecht et.al., 2012; Potgieter, 2009). The Limited Area Model (LAM) developed by the United Kingdom Meteorology Office (UKMO) is run at 20 km resolution over South Africa at the South African Weather Service and produced short-range weather forecasts (Dyson, 2004; Idowu, 2008). The National Centers for Environmental Prediction (NCEP) Regional Eta model is used at the South African Weather Service (Riphagen, 2002; de Coning, 2011; Dyson, 2000; Schultze, 2007). The SYNERGIE model developed by Meteo France International is used in the La Reunion Forecast Office to forecast tropical cyclones over the southern Indian Ocean (Caroff, 2002).

Recently studies have been done on a multi-model approach to forecasting rainfall over South Africa. One such study was done using the Unified Model (UM) and the CCAM, and shows that the multi-model ensemble outscores each individual model (S. Landman *et. al.*, 2012). In other studies, multi-model forecasting techniques have been applied to seasonal rainfall forecasting, where the skill of a multi-model approach is analysed in the case of mid-summer rainfall (Landman & Beraki, 2012).

1.5 Objectives of Research

The objectives of this research were to develop a probabilistic short-range, single-model ensemble prediction system (EPS) for southern Africa. This was achieved by constructing four different ensemble prediction systems using the CCAM that were then verified against observational data. These four ensemble systems differed in size, and the simulations were executed at both a low and a high spatial resolution over southern Africa. The results of this verification were compared in order to determine which ensemble system performed the best. This research determined whether a larger or smaller ensemble, at low or high resolution, was the most successful at predicting rainfall over southern Africa.

The results of this study will determine the most accurate EPS that will then be implemented at the Council of Scientific and Industrial Research (CSIR). Since probabilistic forecasts are more able to overcome inherent model errors, they almost always produce more reliable forecasts than deterministic prediction systems that do not take model errors into account (assuming that there are innate errors in the model). Currently at the CSIR the CCAM is being used as a deterministic model. The probabilistic system developed in this study will use multiple members for weather forecasts made up to seven days ahead. This will enhance numerical weather prediction research with an optimised model to produce more skilful short-range weather forecasts, which will benefit both future research and society.

Chapter 2

Data and Methodology

2.1 Introduction

Forecast verification is an essential part of numerical weather prediction. Without confirming or monitoring weather forecasts, no assumptions can be made as to whether the forecasting system or weather model are producing valuable forecasts. Verification of the forecasts is performed in order to monitor how well the forecasts are performing and how accurate they are; to improve upon the forecasts by altering the model to better simulate the observations; and to inspect and compare different forecast models to conclude which produces forecasts of higher quality, and how to improve the forecast systems.

This research focused on daily rainfall during January and February 2009 and 2010. The CCAM was used to produce rainfall forecasts at low (50 km) and high (15 km) resolution, at four initialisation times, with two cumulus parameterisation schemes. Four different ensembles were configured using these 8 possible members, and rainfall forecasts were produced at both resolutions. Verification statistics was performed to analyse the skill and accuracy of these ensembles in relation to the observational rainfall data. The

verification methods that were used were: the root mean square error, Brier score, relative operating characteristic and reliability diagrams. The high resolution ensembles were individually assessed and then compared to each other in order to establish which ensemble configuration was the most skilful overall; the same was done for the low resolution forecasts. The high and low resolution ensemble pairs were also compared to each other so as to establish whether the high or low resolution forecasts were more skilful. It was finally concluded which ensemble prediction system is the most skilful system.

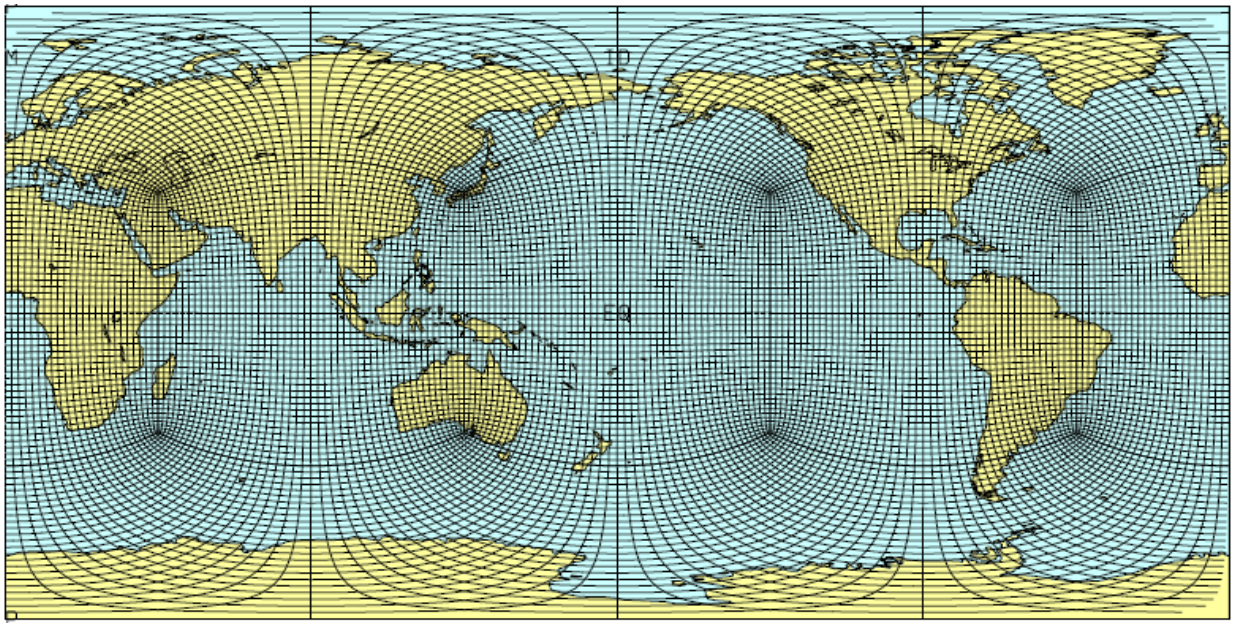

Figure 2.1.1: CCAM model superimposed over the entire globe at horizontal resolution of 200 km (Engelbrecht *et. al.*, 2011)

2.2 Conformal-Cubic Atmospheric Model

The Conformal-Cubic Atmospheric Model (CCAM) was devised by Rancic *et. al.*,1996, and has since been developed at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) (McGregor, 2005). It is a variable-resolution cube projected onto the globe at a resolution of 200 km. Because the CCAM is described as a cube, it is orthogonal and isotropic (McGregor & Dix, 2001; McGregor *et. al.*, 2008). Orthogonality results in the grid lines being perpendicular, therefore allowing vectors (e.g.: wind) to be decomposed into their orthogonal parts. Since CCAM is isotropic, the coordinates are non-polar (Euclidean), and hence the three dimensions can be treated equally. This is particularly useful for atmospheric modelling since isotropic coordinates are good at modelling gravitational fields (Collins & Hawkins, 1973; University of Winnipeg, 2010).

The CCAM is configured such that it is a variable resolution model, and therefore can function both as a global climate model (GCM) and be downscaled to a regional climate model (RCM). The variability ensures that the boundary conditions from the GCM into the RCM are continuous, thus preserving the integrity of the atmospheric characteristics over the boundaries (McGregor *et. al.*, 2008). Because the CCAM can be used as a GCM and a RCM, the CCAM can model global low resolution as well as extremely high spatial resolution over a small area (Engelbrecht *et. al.*, 2011). In addition to this range of functions, the CCAM can also produce predictions from short-range weather to multi-decadal climate change projections (Engelbrecht *et. al.*, 2011).

2.2.1 Changing the Resolution

The CCAM uses a dynamical spectral downscaling technique in order to scale from low to medium to high resolution, and stretch the grid to focus on a specific domain (strech-grid mode). This technique is known at the Schmidt Transformation (McGregor, 2008), proposed by Schmidt in his 1977 paper "Variable Fine Mesh in a Spectral Global Model" (Fox-Rabinovitz *et. al.*, 2006). The Schmidt transformation is a conformal process of both stretching and rotation (Hardiker, 1997). The stretching transformation is an iteration process of the linear equations that dictate the GCM, in order to transform to a higher resolution. The rotation transformation is to move the area of high resolution to be over the domain of interest.

Figure 2.2.1: Illustration of the cube projected onto the globe at 200 km resolution (McGregor, 2006)

In the case of the CCAM, the global low resolution of 200 km has a Schmidt factor of 1, as seen in Figure 2.2.1 and Figure 2.2.2. The medium regional resolution of 60 km has a Schmidt factor of 2.5, as seen in Figure 2.2.3. Extremely high resolutions of 8 km and 1 km have Schmidt factors of 3.3 and 100 respectively (Engelbrecht *et. al.*, 2011).

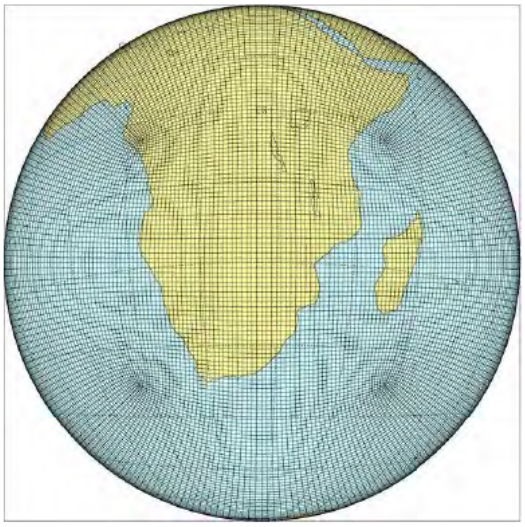

Figure 2.2.2: CCAM scaled from 200 km to 60 km resolution (Engelbrecht *et. al.*, 2011)

2.2.2 Forecasting Rainfall with CCAM

Studies have been done regarding CCAM's ability to forecast precipitation over southern Africa. One such study was of CCAM's ability to predict cut-off lows, and the rainfall produced by them; with a focus on extreme rainfall events ($>$ 20 mm in 24h) (Engelbrecht *et. al.*, 2012). The CCAM was downscaled to a resolution of 60 km over southern Africa, and used to model the frequency of seasonal cut-off low occurrences. The results from this analysis were that the CCAM performed fairly well, especially over the coastal regions, but has a bias to overestimate the frequency of these events. (Engelbrecht *et. al.*, 2012). This indicates that seasonally, the CCAM can be used to recognise extreme rainfall events, with most accuracy over the coastal regions. The biased results over the rest of southern Africa should be taken into consideration when making forecasts for the region.

In previous studies of precipitation over southern Africa the CCAM was able to produce satisfactory predictions of annual rainfall, as well as intra-annual and realistic daily rainfall patterns (Engelbrecht *et. al.*, 2005; Potgieter, 2007). These forecasts were all made on a climate change time scale. In the 2011 study, climate and seasonal time scales were verified to be reliable, but high resolution predictions were largely unverifiable due to the lack of observational data. in the SADC domain, the most accurate forecasts were were recorded over Zambia and Zimbabwe. (Engelbrecht *et. al.*, 2011).

Several verification analyses of the CCAM have been conducted over the rest of the globe. These include analyses of the CCAM's ability to predict monsoons over Indonesia (McGregor & Nguyen, 2010); verification of CCAM simulating atmospheric circulation and climate change projections over Australia (Thatcher & McGregor, 2009); a study of CCAM's ability to predict the Asian monsoon as well as a climate change simulation over Fiji (McGregor *et. al.*, 2010); all of which show significant skill in the model.

2.2.3 Forecast Resolution

The CCAM can run at both low 50 km resolution and high 15 km resolution. High and low resolution rainfall simulations were made with the intention of creating ensembles at both resolutions, performing verification statistics at both resolutions and then comparing the skill of the high resolution ensembles to the low resolution ensembles. It is generally assumed that high resolution forecasts are more accurate than low resolution forecasts, since the higher the resolution, the closer to reality the model is. But there is a possibility that forecasting at high resolution could amplify inaccuracies within the model, such as topographical errors, small anomalies or data errors. Comparing the verification statistics and skill scores of the high resolution ensembles to the low resolution ones will allow conclusions to be drawn as to which is a more skilful system.

2.3 Research Data

Before the research can begin, the data used to initialise the model and making the forecast must exist. In this research, the data used to initialize the CCAM is from the National Oceanic and Atmospheric Administration (NOAA) Global Forecast System (GFS) database. The GFS is a NWP model that assimilates global weather data via satellite and produces global forecasts four times daily.

Observational data at the same resolution as the model output must also be collected. Since we are researching rainfall, the daily rainfall observations are needed. The model was run at resolutions of 50 km and 15 km over southern Africa. The observation data used for this research is from the Tropical Rainfall Measuring Mission (TRMM) data archive (http://www.trmm.gsfc.nasa.gov/). TRMM is an initiative run by NASA to monitor tropical rainfall, and is a part of the Global Precipitation Analysis which produces real-time three-hourly global precipitation data from analysed TRMM data. Verification studies have been performed on the TRMM data. TRMM rainfall estimates were validated against rainfall gauge stations over West Africa and it was seen that the TRMM data overestimates rainfall at medium-range time scales (Nicholson *et. al.*, 2003). This overestimation, however, does not affect this study as the focus was on short-range time scales only. A validation of TRMM data over Bolivia found that the estimated daily rainfall data was unbiased (Gómez, 2007).

2.4 Creating the Ensembles

Firstly, the model domain must be chosen: in this case being the SADC region. The CCAM is a variable resolution model, and the specific domain can be focused in on with the conformal cubic grid through spectral nudging techniques. (Thatcher & McGregor, 2009). The digital filter is applied at a length scale of 4000 km in order to force the large-scale synoptic features (Thatcher & McGregor, 2011). Model dynamics simulate the small-scale events, such as thunderstorm systems. For this study, the model wasspectrally nudged from the Schmidt Factor 1 resolution of 200 km, to a resolution of 50 km which wasconsidered in this study to be coarse resolution; and further nudged to a high resolution of 15 km. In this strech-grid mode, the CCAM can function as a limited area model over southern Africa.

The initial data is assimilated and issued four times a day, at 00h00, 06h00, 12h00 and 18h00 GMT, and so the CCAM can be initialised four times a day, producing four forecasts a day. The model has two different cumulus convection parameterisation schemes which can be used to initialise the model. There is a cloud microphysics scheme that has been developed by Rotstayn (1997), and a cumulus convection scheme that has been created by McGregor (2003). These schemes describe the complicated and detailed cloud microphysics and dynamical processes (Engelbrecht *et. al.*, 2007; Thatcher, 2010). The ensembles were constructed using these time-lagged members – what is known as the 'lagged average forecasting' technique. Time-lagged ensemble forecasting has been used successfully for short- and medium-ranged forecasting (Hoffman & Kalnay, 1983; Lu, 2007). Lagged average forecasting is used instead of the perturbation technique, where small perturbations are made to a the real initial conditions, also know as 'breeding' (Lewis, 2005), to create a large set of perturbed initial conditions that are then used to initialise the model (Buizza, 2008; Lawrence & Hansen, 2006). This method is successful in creating very large ensembles as there are an infinite number of ways to perturb the

original conditions (Kalnay, 2003). The lagged average forecast method is used in this research because multiple sets of real initial data are available, which makes time-lagged forecasting possible. However, this does place a limit on the number of ensemble members that can be created.

Configured thusly, the CCAM can be run a maximum of eight times a day: four times for each cumulus parameterisation scheme. This means there is a maximum of eight different ensemble members. Usually ensemble members obtained through lagged average forecasting are all given equal weight, since estimating the relative weights according to how recent the initial conditions are has proven very difficult (Kalnay, 2003). Therefore, the members in all of these ensemble systems are to be given equal weighting.

Four ensembles are going to be analysed in this dissertation, and they are to be constructed as follows:

1. **Ensemble 1:** The two most recent members – 18h00 from both parameterisation schemes
2. **Ensemble 2:** The four most recent members – 12h00 and 18h00 from both parameterisation schemes.
3. **Ensemble 3:** The six most recent members – 06h00, 12h00 and 18h00 from both parameterisation schemes.
4. **Ensemble 4:** All eight members - 00h00, 06h00, 12h00 and 18h00 from both parametrisation schemes.

These four different configurations were chosen in order to see the effect of ensemble size on ensemble skill, and also to see whether including older members in an ensemble affects the ensemble skill. It is possible that the more members an ensemble includes, the more skilful the system becomes, since it is more likely to capture outlying events. It is also possible that ensembles which include older members (initialised at times 00h00 and

06h00) could be less skilful than ensembles made up of only recent members, because the older members may amplify inaccuracies that have been minimised in the more recent members. (It is generally accepted that forecasts with longer lead-times are usually less accurate than forecasts made with shorter lead-times. i.e.: a forecast for four days ahead that is made at 00h00 is expected to be less accurate than one made at 18h00).

For example: comparing Ensemble 1 to Ensemble 3 will show how the skill of an eight-member ensemble compares to that of a four-member ensemble; and also how the skill of an ensemble that includes older members (initialisation times 00h00 and 06h00) differs to that of an ensemble made up of only recent members (12h00 and 18h00).

2.5 Forecast Verification

The verification of the four ensembles were evaluated by studying four skill measurements. Multiple assessments must be calculated, since no one single measurement can completely describe the attributes of the forecast (Casati *et. al.*, 2008; Stanski *et. al.*, 1989). The four methods that were used are: the root mean square error, Brier skill score, relative operating characteristics, and reliability diagrams. These verification measures can be split in to two categories: deterministic, and probabilistic verification measures. The root mean square error is a deterministic verification score, whereas the Brier skill score, the relative operating characteristics and the reliability diagrams are probabilistic verification methods. The results of these assessments will determine the quality of the forecasts made by each ensemble system.

Several terms relevant to forecast verification must be defined:

1. **Accuracy**: A measurement of the average difference in individual forecast-observation pairs.
2. **Reliability**: Assessment of how well the forecast probabilities correspond with the observed frequencies.
3. **Resolution**: Summarises the ability of the model to differentiate between the conditional probabilities of the observations and the climatological average.
4. **Skill**: The average accuracy of the forecasts compared to the average accuracy of a reference forecast (e.g.: climatology, persistence, etc.).
5. **Uncertainty**: The degree of variability in the observations.

2.5.1 The Root Mean Square Error (RMSE)

The RMSE is a measure of accuracy, and is represented by the average magnitude of error in the forecast (Kalnay, 2003). This is calculated by taking the difference between the forecast and observed values, squared; these values for each point are then summed and averaged to produce the Mean Square Error. The square root of this is taken to produce the RMSE. The RMSE values are plotted against the time intervals of the forecast.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{2.5.1}$$

Equation (2.5.1): Formula of the Root Mean Square Error, where $\hat{y}_i$ is a vector of predictions and y_i is a vector of observations.

The RMSE values are plotted against the time intervals of the forecast. The more accurate the forecast, the closer to zero the RMSE values should be. The closer the forecast is to the observations, the smaller the RMSE. As the time steps increase the RMSE increases in value (Figure 2.5.1), since the further away the time step is from the forecast initialisation, the less skilful the forecast will be. Uncertainty inherent in the model (and in the weather) make forecasting future time-steps progressively more difficult. This makes the RMSE of each prediction system increase with time.

Spatial representation of the RMSE is useful in order to see the spatial distribution of error over the forecast domain. Figure 2.5.2 is an example of a spatial distribution map of RMSE for mean surface temperature.

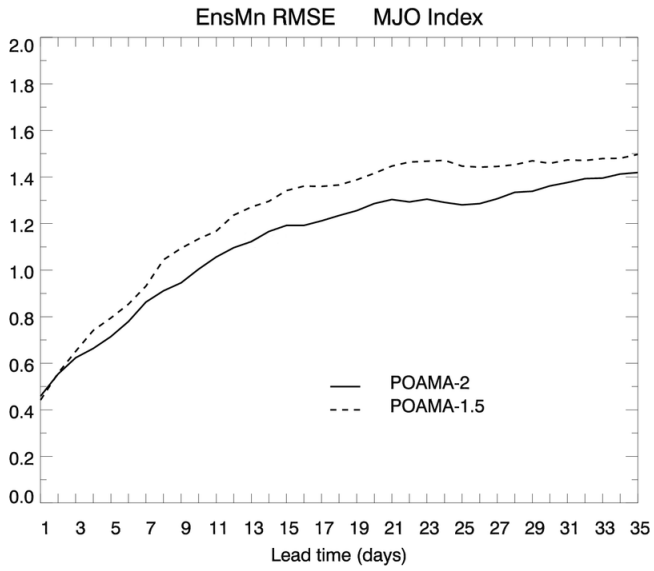

Figure 2.5.1: Graph illustrating the RMSE calculated and plotted for two ensemble systems (Centre for Australian Weather and Climate Research, 2011)

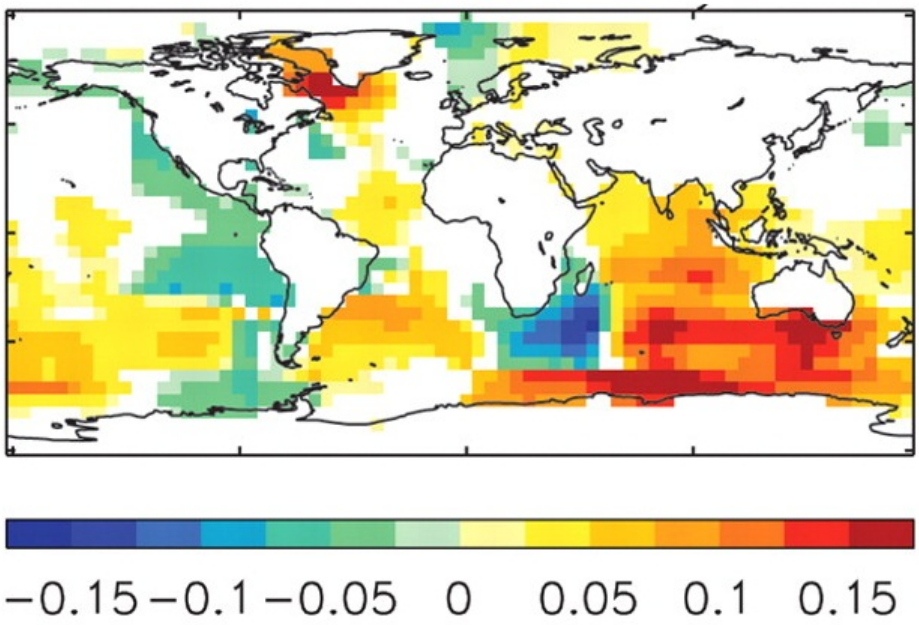

Figure 2.5.2: RMSE Spatial Distribution (Smith *et. al.*, 2007)

2.5.2 The Brier Score (BS)

The Brier Score (BS) is calculated by averaging the square differences between forecast probability p_i and observation o_i pairs (Brier, 1950). The observation value is measured as either 1 if the event occurred or 0 if the event did not occur. (Stanski *et. al.*, 1989). Therefore, the BS is negatively-oriented and a perfect forecast has BS = 0. When evaluating the skill of an ensemble, the smaller the value of BS, the more accurate the forecast (Ferro, 2006). As seen in Figure 2.5.3, the value of the BS increases with an increase in time-step. This is because the further the forecast in from the initialisation time, the more difficult it is to simulate the atmosphere and produce forecasts close to the observations.

$$BS = \frac{1}{n} \sum (\overline{p}_i - \overline{o}_i)^2 \tag{2.5.2}$$

Equation (2.5.2): Where k indicates the indexing of the n forecast-observation pairs, $\overline{p}_i$ is the forecast probability and $\overline{o}_i$ is the observed value.

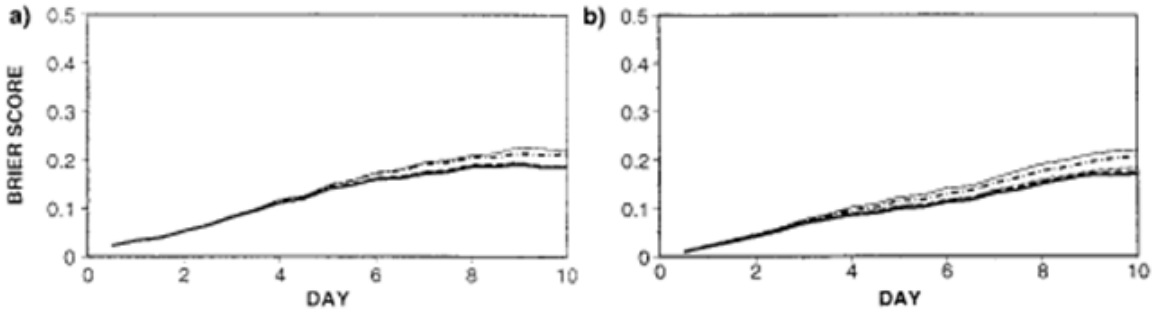

Figure 2.5.3: Two examples of Brier Scores calculated for 5 different ensembles. (Buizza & Palmer, 2008)

The BS can be decomposed into three terms describing reliability, resolution and uncertainty:

$$BS = \frac{1}{n} \sum_{i=1}^{I} N_i (\overline{p}_i - \overline{o}_i)^2 - \frac{1}{n} \sum_{i=1}^{I} N_i (\overline{o}_i - o_i)^2 + \overline{o}(1 - \overline{o}) \tag{2.5.3}$$

Reliability Term Resolution Term Uncertainty Term

Equation (2.5.3): Decomposition of the Brier Score, where I indicates the number of forecast categories, and N the number of times said forecast category is used in the collection of forecasts.

2.5.2.1 Reliability term: $\frac{1}{n}\sum_{i=1}^{I} N_i\left(\overline{p}_i - \overline{o}_i\right)^2$

The reliability term is the weighted average of the difference between the forecast probabilities ($\overline{p}_i$) and the observed frequencies ($\overline{o}_i$). This term measures the calibration of the system: how accurately the system forecasts probabilities. The more reliable a system is, the closer the reliability term is to zero.

2.5.2.2 Resolution term: $\frac{1}{n}\sum_{i=1}^{I} N_i\left(\overline{o}_i - o_i\right)^2$

The resolution term is the average square difference between the observed frequencies ($\overline{o}_i$) in each category, and the mean observed frequency (o_i) of the entire sample (N). This term measures how well the system is able to tell the different categories apart, irrespective of the forecast probabilities. (Atger, 1999). This signifies the sharpness of the system: when the resolution is a maximum, the system will forecast either 100% or 0%, which is a deterministic forecast. When the resolution is a minimum, the probabilities are all equal and the system becomes a climatological forecast. If the categories have very different relative frequencies then the resolution term will be large and the system resolves well.

2.5.2.3 Uncertainty term: $\overline{o}\left(1 - \overline{o}\right)$

The uncertainty term is clearly independent of the forecast system and measures the variance of the observations, using only the observed frequencies ($\overline{o}$).

2.5.2.4 The Brier Skill Score (BSS)

The BSS is calculated by comparing the BS of the forecast to the BS of a reference forecast (e.g.: persistence or climatology); and therefore BSS is a measure of the skill of the forecast with reference to the accuracy of the forecast determined by the BS. The BSS is the conventional way of measuring the forecast skill (Wilks, 2006, Buizza & Palmer, 1997). In this dissertation persistence is used as the reference forecast.

$$BSS = 1 - \frac{BS_{forecast}}{BS_{reference}} \tag{2.5.4}$$

Equation (2.5.4) shows the BSS in terms of the reference and forecast BS.

Since a perfect BS = 0, perfect BSS is equal to 1, with no skill having a BS equal to or less than 0. Figure 2.5.4 illustrates how the skill of the forecast decreases with an increase in time-step. The BSS can also be represented spatially in order to determine the spatial distribution of skill. Figure 2.5.5 shows a spatial distribution of Surface Temperature.

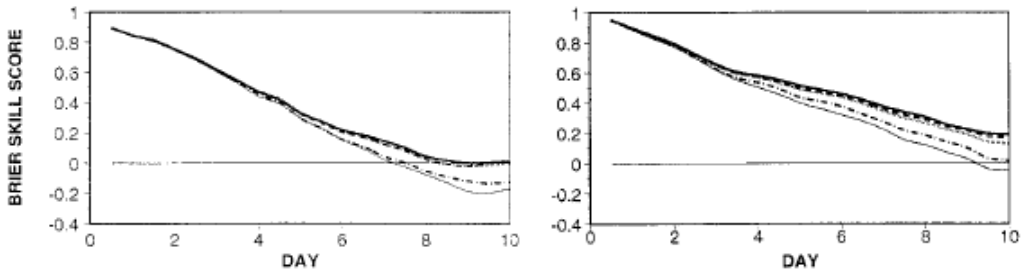

Figure 2.5.4: Brier skill score (Buizza & Palmer, 1998)

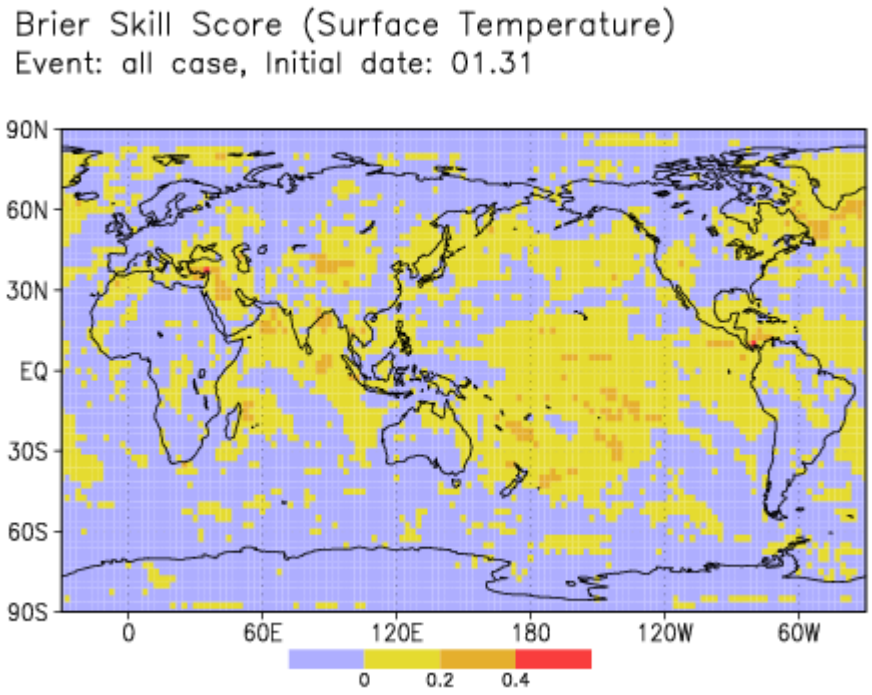

Figure 2.5.5: Spatial Representation of the Brier skill score (Japan Meteorological Agency (JMA), 2012)

2.5.3 The Relative Operating Characteristics (ROC)

The relative operating characteristics (ROC) is a measure of how well a forecast system is able to distinguish between an events and non-events. The ROC diagram illustrates the false alarm rate (FAR) versus the hit rate (HR) within a range of probability thresholds. The ROC is usually used to determine whether a forecast system can successfully discriminate between events and non-events at a certain threshold (e.g.: 5 mm of rainfall). The skill of a forecast when using the ROC is calculated using Table 2.1, the contingency table.

Table 2.5.1: Contingency Table

		Observed		
		Yes	**No**	
Forecast	**Yes**	Hit	False Alarm	Total Yes-Forecast
	No	Miss	Correct Negative	Total No-Forecast
		Occurrences	Non-Occurrences	**TOTAL**

Table 2.1 is a contingency table that contains the attributes of the forecasts based on a "yes" or "no" count. When a forecast and observation pair exceeds the desired threshold the count will be "yes", and if the threshold is not exceeded, the count will be "no" (Wilks, 2006; Jolliffe & Stephenson, 2003). Each forecast-observation pair is checked and depending on the "yes"/"no" decision, is deemed a "Hit", a "False Alarm", a "Miss" or a "Correct Negative". A "Hit" occurs when the event was forecast and observed. A "False Alarm"s when the even was forecast but was not observed. A "Miss" is what is recorded when the even was not forecast, but was observed. A "Correct Negative" happens when an event was neither forecast nor observed. The ROC score is calculated using the False Alarm Rate and the Hit Rate, which are determined from the four attributes.

The False Alarm Rate (FAR) is defined as the ratio of the false alarms to the total non-occurrences:

$$FAR = \frac{False\ Alarms}{False\ Alarms + Correct\ Negatives} \tag{2.5.5}$$

The Hit Rate (HR) is defined as the ratio of the hits to the total occurrences:

$$HR = \frac{Hits}{Hits + Misses} \tag{2.5.6}$$

The FAR as defined in Equation 2.5.5 is plotted against the HR as in Equation 2.5.6 to produce a graph with a curve from (0, 0) to (1, 1). The area under this curve is calculated using the trapezium method, where an area of 0.5 or less is a forecast with no skill and an area equal to 1 is perfect skill. If the area under the curve is found to be greater than 0.5, the forecast wasconsidered skilful. The closer the area is to being equal to 1, the more skilful the forecast is concluded to be. The diagonal line from (0, 0) to (1, 1) is the line of no skill.

Forecast systems that exhibit good discrimination have ROC curves that approach the upper-left corner of the diagram closely (point (0,1)), which is a perfect forecast with area of 1. Forecast systems have very little skill if the curve falls below the (0,0) to (1,1) diagonal line – the lower threshold for a useful forecast is the diagonal, and these curves have an area of <0.5. In Figure 2.5.6 it can be seen that Example 1 exhibits a greater ability to discriminate than Example 2 because the Example 1 curve has a an area closer to 1 than the Example 2 curve. (Wilks, 2006)

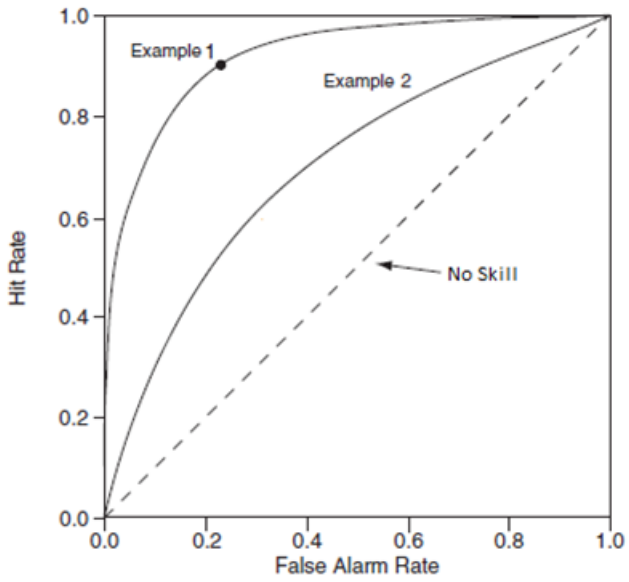

Figure 2.5.6: A ROC diagram for two different ensemble systems. (Wilks, 2006)

2.5.4 Reliability Diagram

A Reliability diagram illustrates how well-calibrated a forecast system is. The diagram is a plot of conditional event relative frequencies (Equation 2.5.5):

$$p\left(o_1 \mid y_i\right) \tag{2.5.7}$$

A perfect or well-calibrated forecast has a function which lies along the 1:1 diagonal of the diagram. The central diagram in Figure 2.5.7 is an example of a well-calibrated forecast. In this case, the conditional event relative frequency is almost equal to the forecast probability (Equation 2.5.6):

$$p\left(o_1 \mid y_i\right) \approx y_i \tag{2.5.8}$$

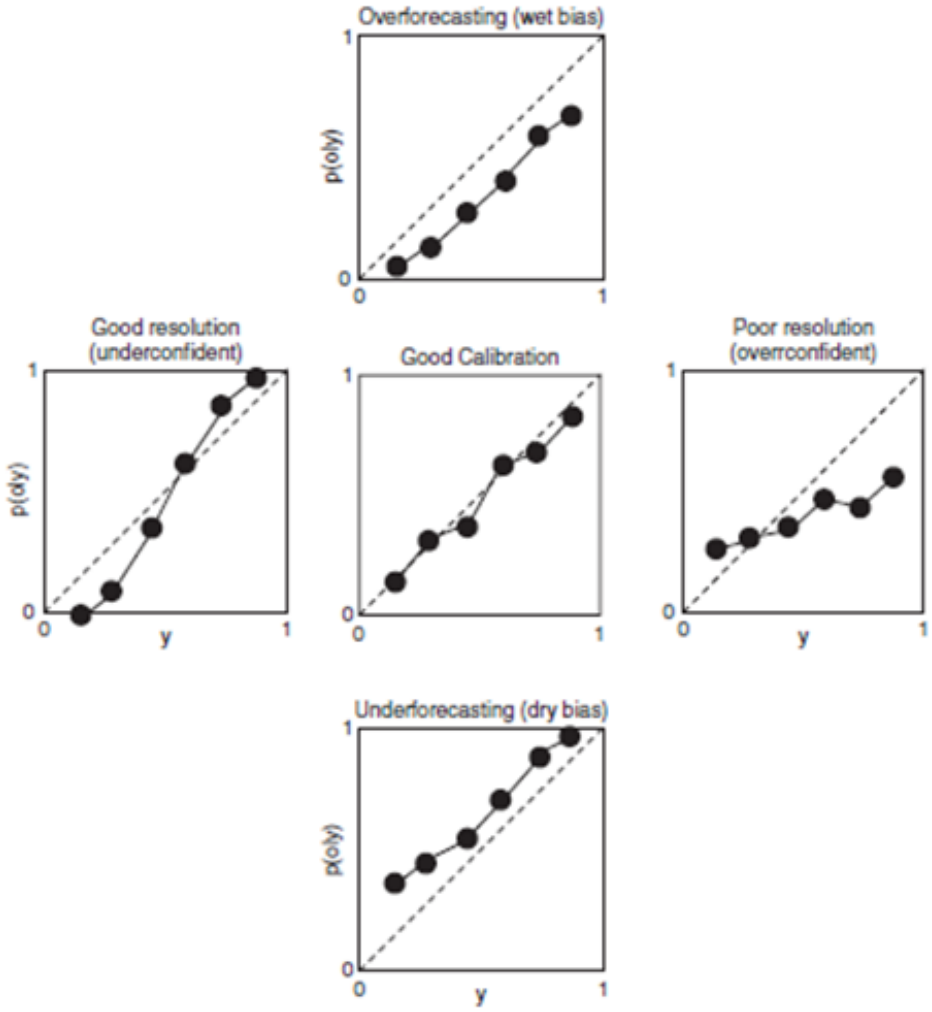

Figure 2.5.7: Examples of reliability diagrams, showing calibration functions $p(o \mid y)$ as function of forecast y. (Wilks, 2006)

From a reliability diagram, model biases and the quality of the resolution can be determined:

A forecast is said to be unconditionally biased if the function lies entirely to the left, or to the right of the 1:1 diagonal line. If the function is entirely to the right of the diagonal

(top diagram of Figure 2.5.7), the forecasts are considered to be constantly too large when compared to the conditional event relative frequencies, and hence the average forecast is greater than the average observation – this signifies overforecasting. If the function is entirely to the left of the diagonal (bottom diagram of Figure 2.5.7), the forecasts are considered to be constantly too small when compared to the conditional event relative frequencies, and hence the average forecast is smaller than the average observation – this signifies underforecasting.

A forecast is said to be conditionally biased when the slope of the function is either shallower or steeper than the 1:1 diagonal line. In these cases, the magnitudes of the forecast biases depend on the forecasts themselves. A function that has a slope shallower than the diagonal (right diagram of Figure 2.5.7) indicates an overconfident forecast. This implies poor resolution, since the conditional event relative frequencies are all near the climatological probability (they depend only weakly on the forecasts). A function that has a slope steeper than the 1:1 diagonal (left diagram of Figure 2.5.7) is an underconfident forecast. This implies good resolution, since the average square differences are large and therefore the system is able to distinguish different outcomes clearly.

Reliability diagrams are accompanied by frequency diagrams (also known as sharpness diagrams) that show the probability distribution of the forecasts. These are in the form of histograms and made by distinguishing different forecast probability bins, then calculating the sample size in each bin. Predictions that have probabilities further from the climatalogical average are considered to exhibit 'sharpness' and the forecast system has 'sharpness' (ECMWF, 2012). The climatological average is positioned at the middle of the x-axis of the sharpness diagram, and a forecast system with sharpness will have predictions that cluser around the extremes of the x-axis. In Figure 2.5.8 the sharpness diagram for the ensemble is seen shown on the reliability diagram, and the forecast is considered to exhibit sharpness.

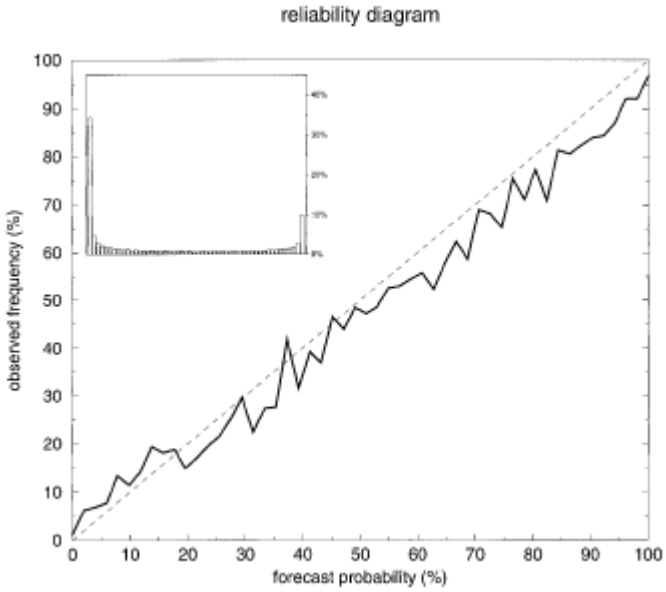

Figure 2.5.8: Reliability diagram accompanied by corresponding frequency diagram (Atger, 1999)

2.6 Synopsis

This chapter has provided the necessary background needed to proceed to the verification results, and a discussion thereof. The next chapters will be a detailed exploration of the results of the deterministic and probabilistic skill measurements described above.

Chapter 3

Deterministic Verification

3.1 Introduction

Deterministic verification is a quantitative, non-statistical approach to verifying weather forecasts. This type of verification compares the actual values of the observed and forecast variable, resulting in accuracy scores that are measured in terms of the value and unit of the variable being verified, not the probability of occurrence. This allows for easy interpretation of the analysis (Bowler, 2006; Fawcett, 2008). Deterministic verification can analyse the error, correlation or bias of a set of forecasts and is the basis of the root mean square error. The root mean square error is the verification score that was studied and discussed in this chapter.

3.2 Root Mean Square Error

This chapter focuses on the analysis of the Root Mean Square Error (RMSE) verification of short-range rainfall forecasts made for the time period of January and February, 2009 and 2010. This verification consists of two parts: The average ensemble RMSE in graph form, and spatial distribution maps of RMSE. From Equation 3.2.1 that describes the RMSE, a perfect forecast will have an RMSE of 0, showing zero error in the forecast.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{3.2.1}$$

Weather forecast error increases with increasing lead-time (Kalnay, 2003), a notion that is also shown in the verification of the ensemble mean hindcasts of January 2009 (Figures 3.3.1 and 3.3.2). However, the forecast error initially increases at a higher rate for the low resolution case as opposed to the high resolution case. Notwithstanding the smaller errors associated with forecasts for the first few days, forecast errors made towards the longest lead-time forecasts (i.e., forecast errors for days 5 to 7) are about equally large for both resolutions, thus suggesting that the high-resolution forecasts are mainly advantageous for forecasts up to about day 4. This is consistent with research where it is concluded that forecast error becomes so large at long lead-times that it essentially renders the forecast as inaccurate (Lu *et. al.*, 2007; Bowler *et. al.*, 2008; Grimit, 2004).

3.3 January 2009

The two graphs in Figure 3.3.1 show that the errors for the four ensembles are very similar in both the high resolution and low resolution cases. Figure 3.3.2 shows the spatial distribution of RMSE values for forecasts made for day 1, for each of the 4 ensembles, for the high-resolution case. The maps of this figure show similar values and spatial distribution of error, suggesting that the forecast errors are not much influenced by the size of the ensemble. This notion is further demonstrated in Figure 3.3.3 that shows the error difference between the largest ensemble forecasts and the forecasts from progressively larger ensembles. However, although the differences are small, the largest error differences are obtained between the largest and smallest ensembles, which implies that there may be advantages for short-range weather predictions to use the largest possible ensemble.

From Figures 3.3.1 and 3.3.2 the conclusion can be drawn that the high-resolution forecasts show the smallest errors, especially for forecasts up to day 4, but that the size of the ensemble had a much less dramatic influence on forecast performance (Figure 3.3.3). A maximum of only 8 members were considered here, based on initialization time and the selection of cloud parametrisation scheme. For the latter, two cloud schemes (Rotstayn, 1997; McGregor, 2003) were used, but it was found that the selection of different micro-physical schemes to increase the ensemble size may only have a negligible influence on forecast performance (Liu & Moncrieff, 2007). Hence if this is the case, the ensemble size used here is effectively only 4, and the use of the full ensemble has already alluded to the advantages that can be obtained by improving on the ensemble size.

Future modelling work at this time-scale should therefore devise new schemes or use existing approaches to significantly improve on the size of the ensemble for weather forecasting purposes (Buizza, 2010).

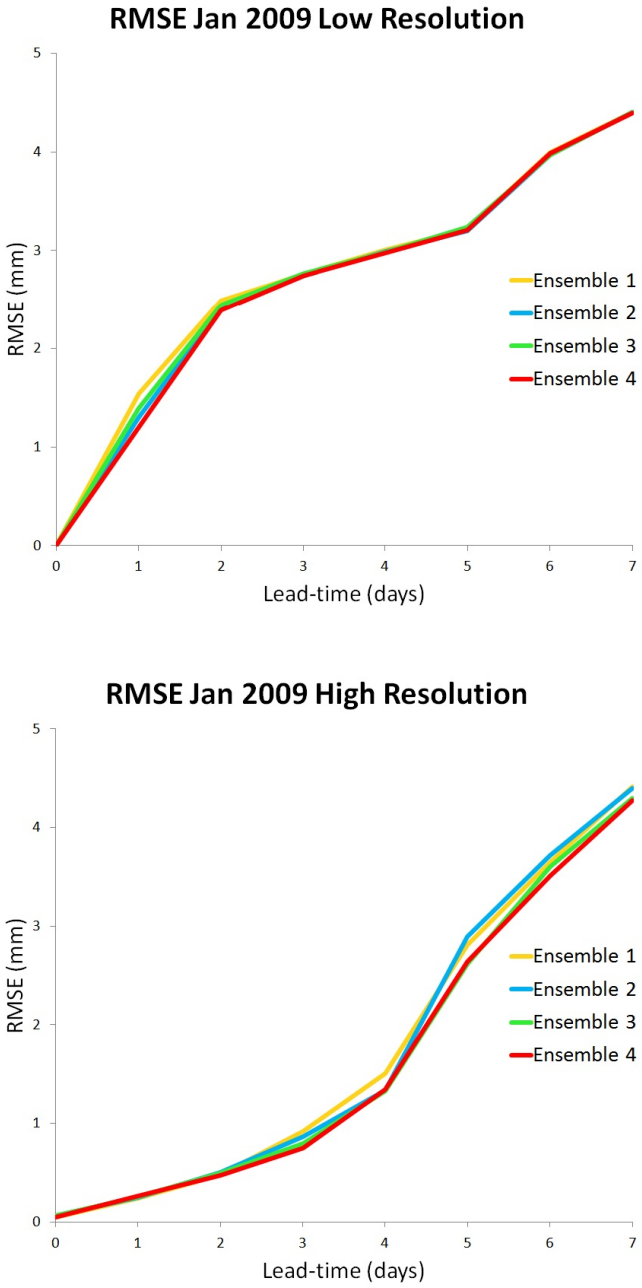

Figure 3.3.1: RMSE graph for January 2009. Above: Low Resolution, Below: High Resolution.

]

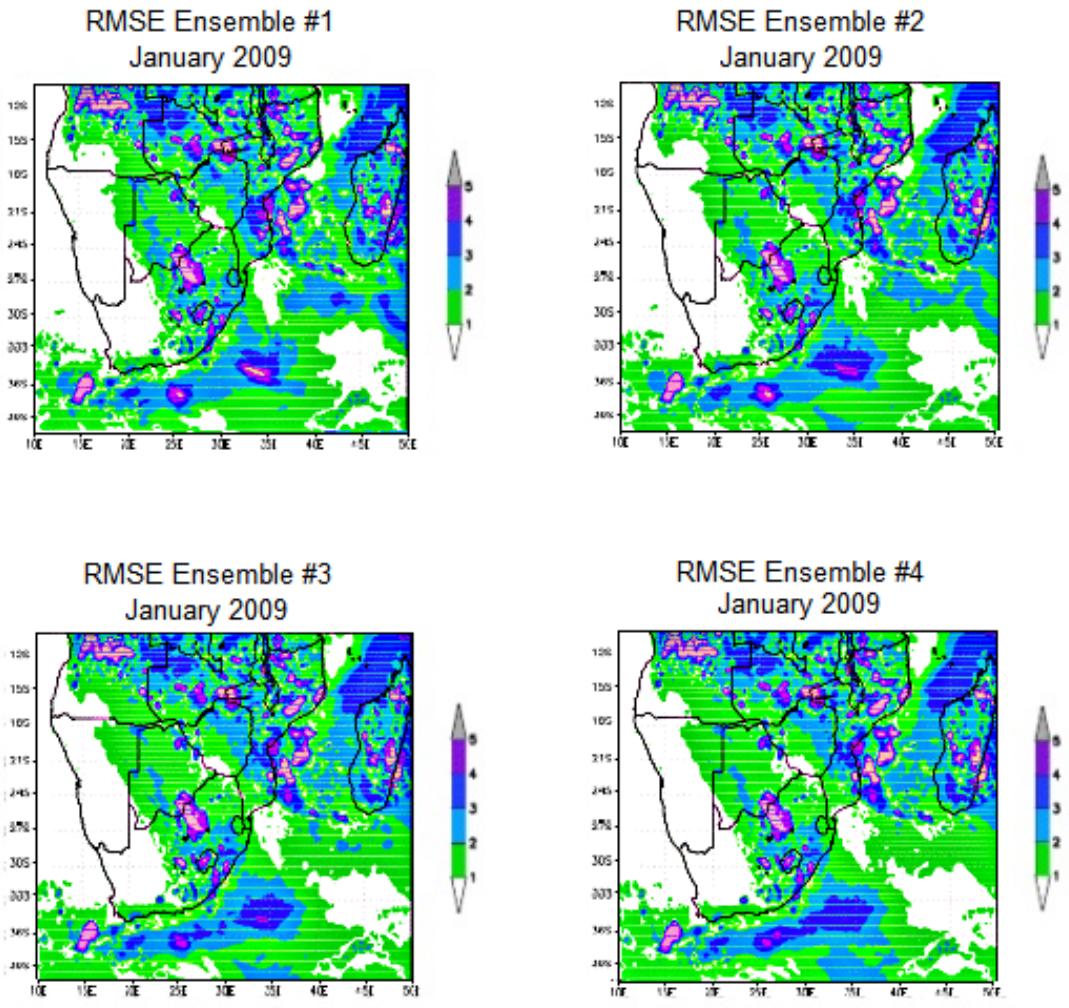

Figure 3.3.2: Spatial distribution of Error at High Resolution

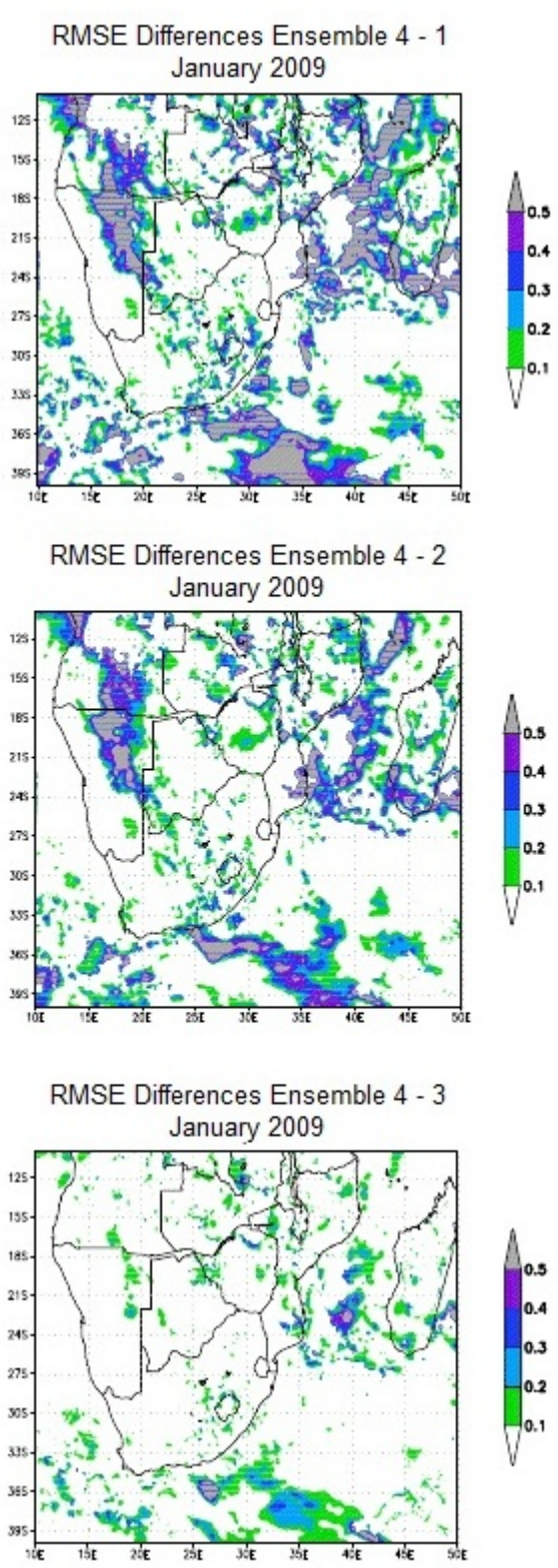

Figure 3.3.3: Difference in Error between Ensemble 4 and Ensemble 1, 2, 3 at High Resolution

3.4 February 2009

As in January 2009, the errors for the four ensembles in February 2009 for both resolutions are very similar to each other: this can be seen in Figure 3.4.1 that shows lower error values for the high resolution forecasts than the low resolution forecasts. All four ensembles at low resolution have almost identical errors, but the high resolution forecasts show a noticeably larger in error in Ensemble 1 compared to Ensemble 2, 3, and 4.

The spatial distribution of RMSE values for the high-resolution forecasts on day 1 shown in Figure 3.4.2 illustrates the similarities between the distributions, implying that the size of the ensemble does not significantly improve the forecast error.

Figure 3.4.3 of the spatial error differences between Ensemble 4 and the smaller ensembles also substantiates the conclusion that, in most cases, the ensemble size has little effect on the forecast error. As an exception, it can be seen in Figure 3.4.3 that the largest difference in error is between Ensemble 4 and Ensemble 1, which suggests that the forecasts do still improve, albeit marginally, with an increase in ensemble size in this case.

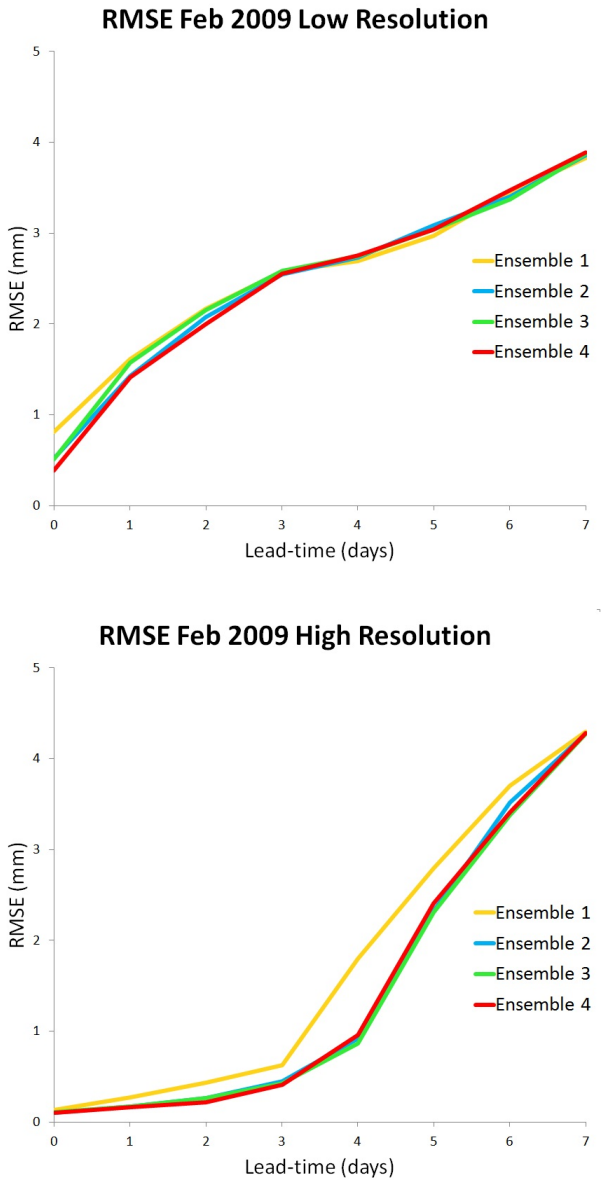

Figure 3.4.1: RMSE graph for February 2009. Above: Low Resolution, Below: High Resolution.

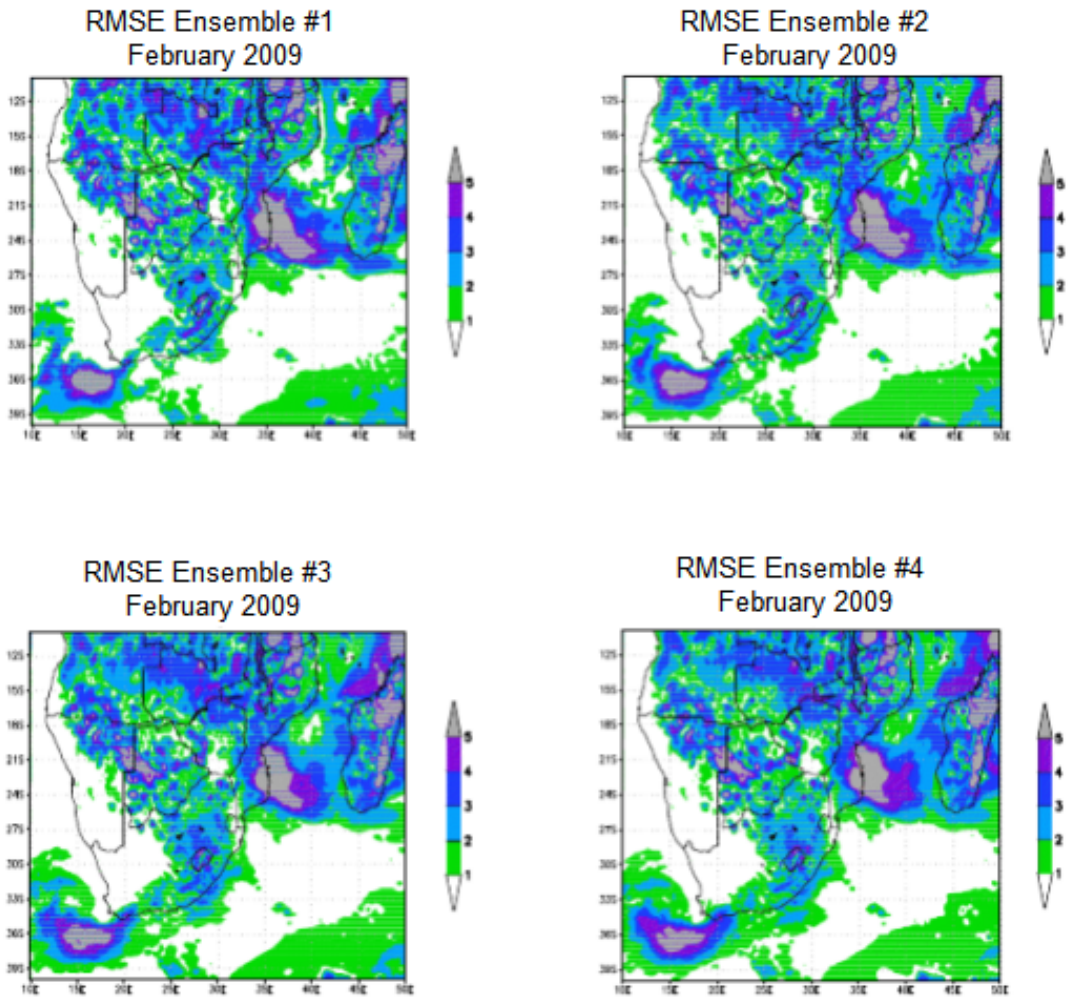

Figure 3.4.2: Spatial distribution of Error at High Resolution

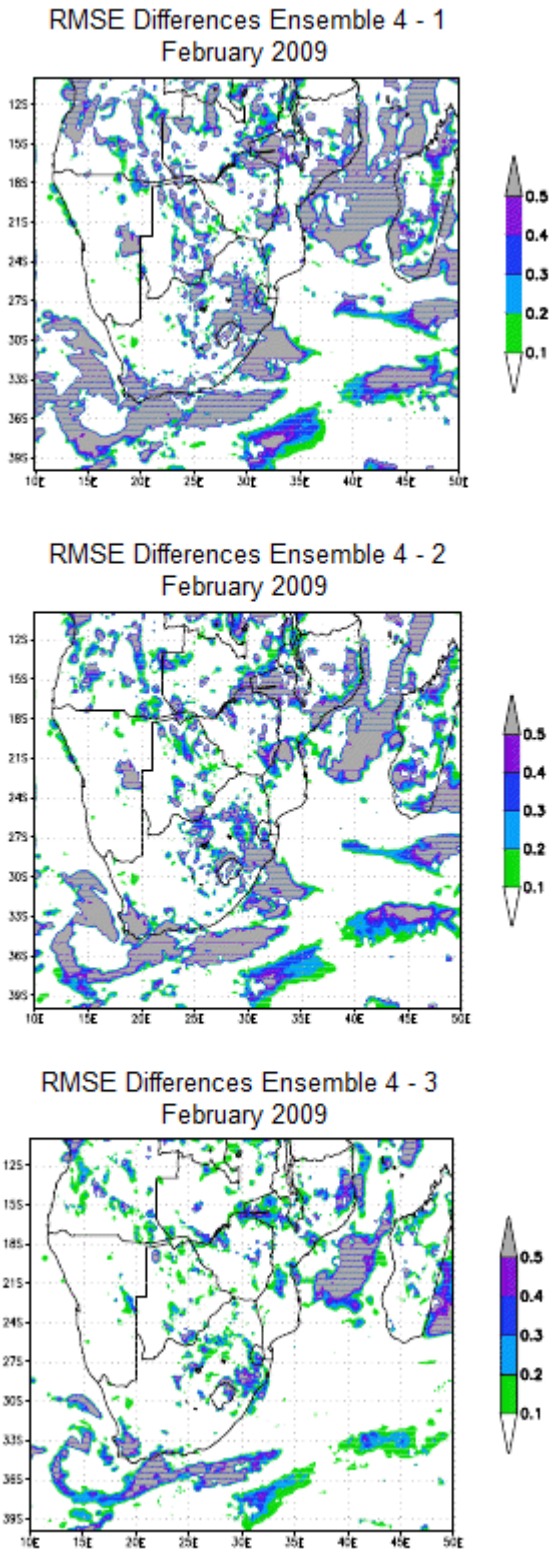

Figure 3.4.3: Difference in Error between Ensemble 4 and Ensemble 1, 2, 3 at High Resolution

3.5 January 2010

In the low resolution forecasts shown in Figure 3.5.1, the errors for all four ensembles are very close except for the forecast of Day 1, where the largest ensemble performs the best. For the high resolution forecasts (Fig. 3.5.1) the four ensembles perform very similarly, with Ensemble 3 and 4 slightly outperforming Ensemble 1 and 2. The errors for the high resolution forecasts are far less than those for the low resolution forecasts, indicating that there is a great improvement in accuracy with an increase in resolution. This implies that a continuing increase of the spatial resolution will render results with even smaller error.

The spatial distribution of error for forecasts made at high resolution for Day 1 shows almost identical error for all four ensembles as can be seen in Figure 3.5.2. The largest difference in error is between Ensemble 4 and Ensemble 2, where Ensemble 2 has smaller error than Ensemble 4. This result once again motivates the idea that the reduced effective ensemble size due to minimal difference in the two cumulus parametrisation schemes results in two very small ensembles being compared, and so there is very little difference in error.

The error difference graphs in Figure 3.5.3 show that there is not much consequence to increasing the ensemble size from 2 members to 8 members when forecasting at high resolutions. It is possible that a greater ensemble size is needed in order to see a large reduction in error; although the error values for these diminuative ensembles can be seen to be small (as illustrated in Figure 3.5.3).

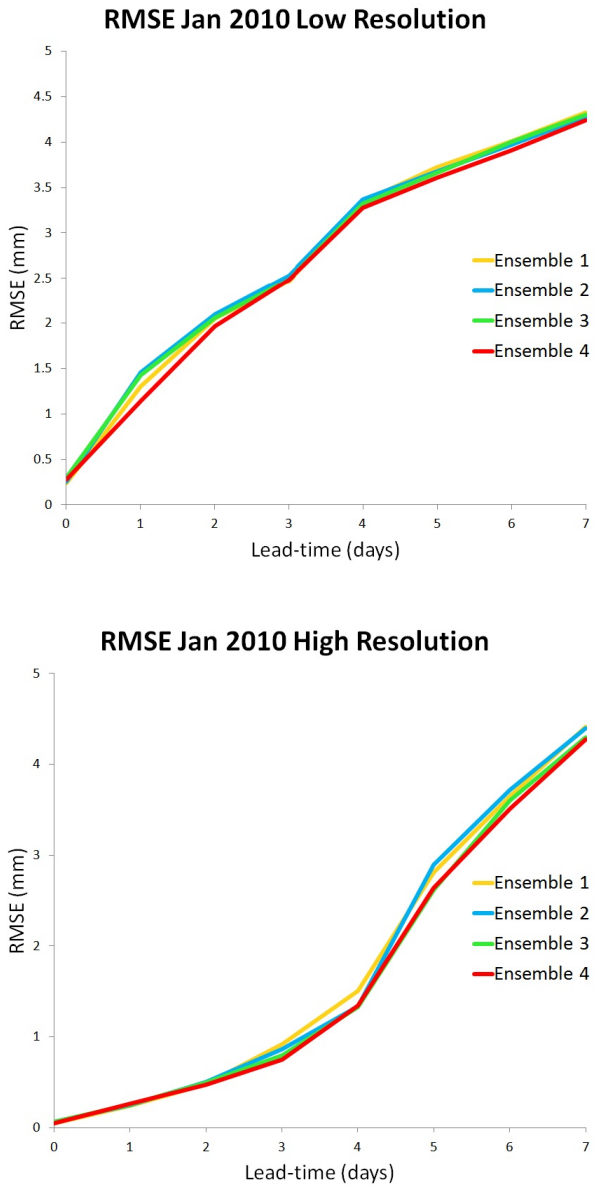

Figure 3.5.1: RMSE Graph for January 2010. Above: Low Resolution, Below: High Resolution.

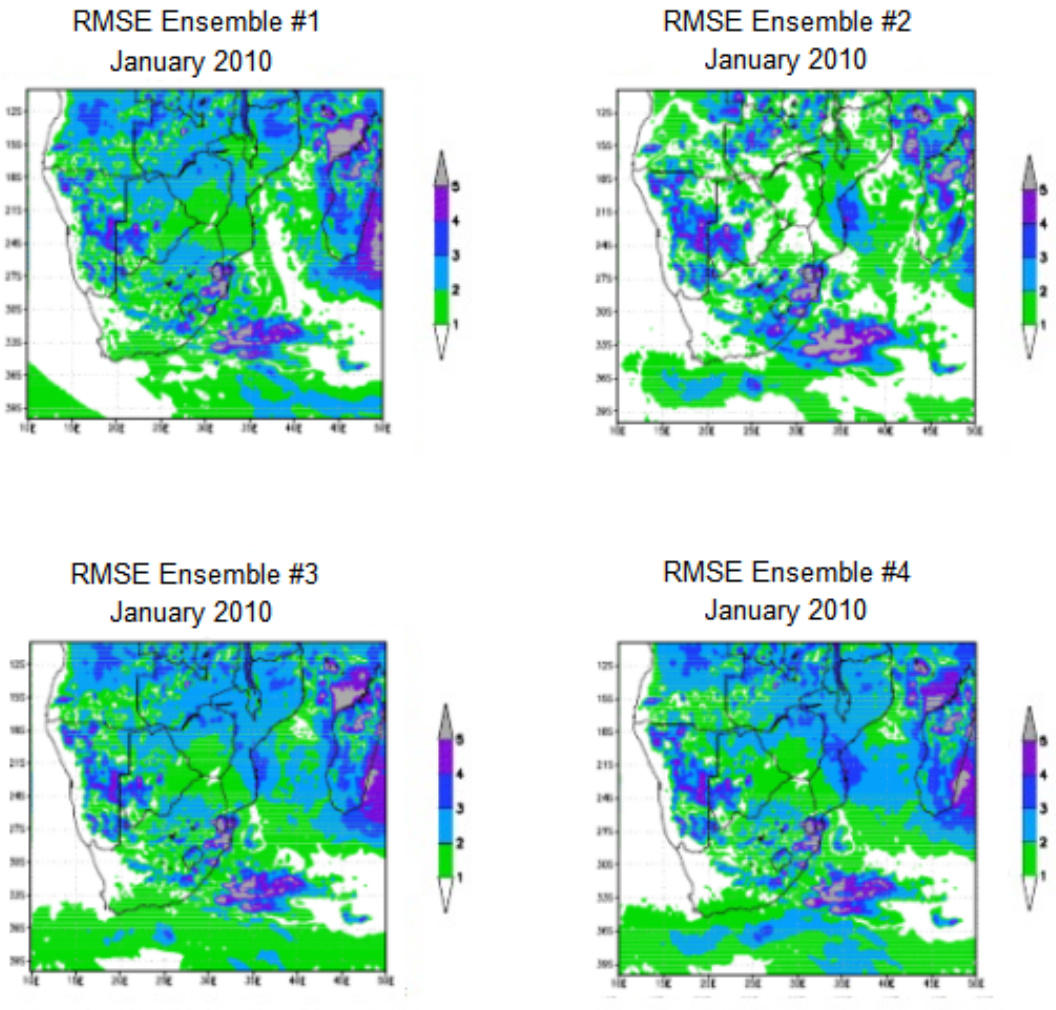

Figure 3.5.2: Spatial distribution of Error at High Resolution

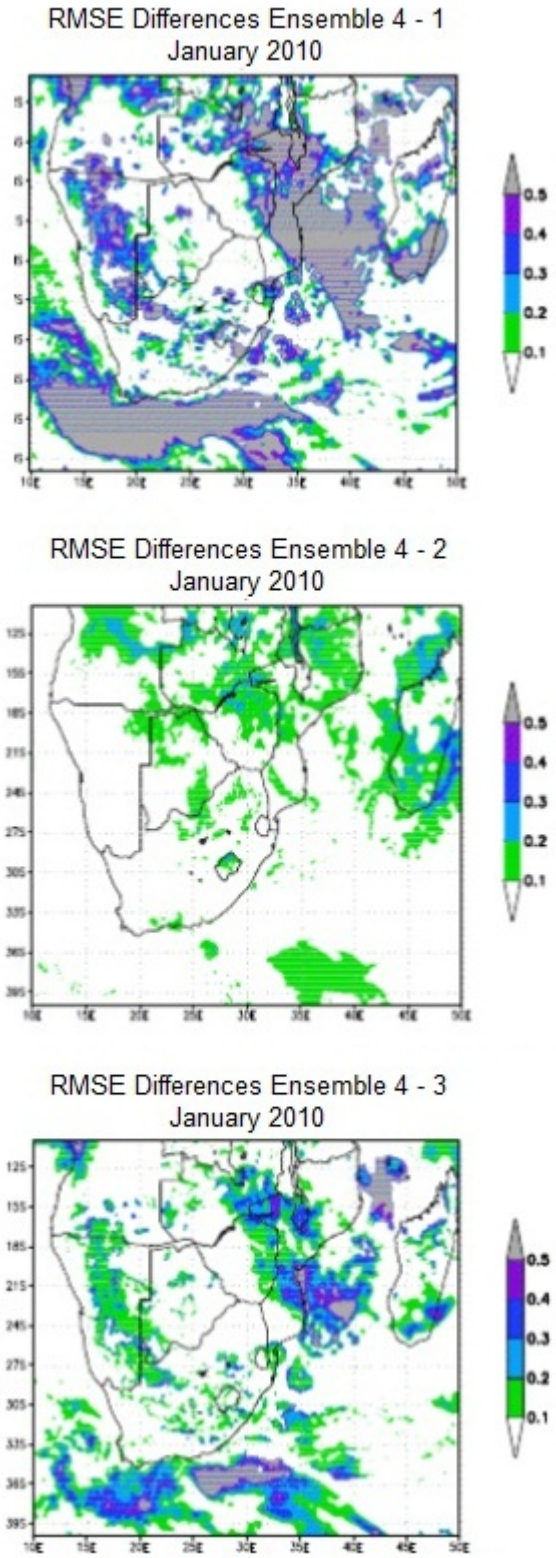

Figure 3.5.3: Difference in Error between Ensemble 4 and Ensemble 1, 2, 3 at High Resolution

3.6 February 2010

The four ensembles have near identical errors, seen in Figure 3.6.1. The RMSE values for the forecasts performed at low resolution are greater than those at high resolution proving that there is a decrease in error at higher spatial resolution, except at longer lead-times (6 and 7 days).

Figure 3.6.2 shows maps of the spatial distribution of RMSE values for the high-resolution forecasts on day 1 for all four ensembles. The four maps have nearly identical features which implies that for small ensembles, a small increase in size of the ensemble does not make a significant difference in the spatial error of the forecast.

Figure 3.6.3 illustrates the spatial error differences between Ensemble 4 and Ensemble 1, 2 and 3 and these maps also conclude that a relatively small increase in ensemble size does not make much of an improvement to the forecasts. However, the largest difference in error is seen between the largest and smallest ensembles (Ensemble 4 and Ensemble 1), and this suggests that even though an increase in the size of small ensembles does not make a large difference in forecast accuracy, the largest ensemble produces forecasts with the least amount of error.

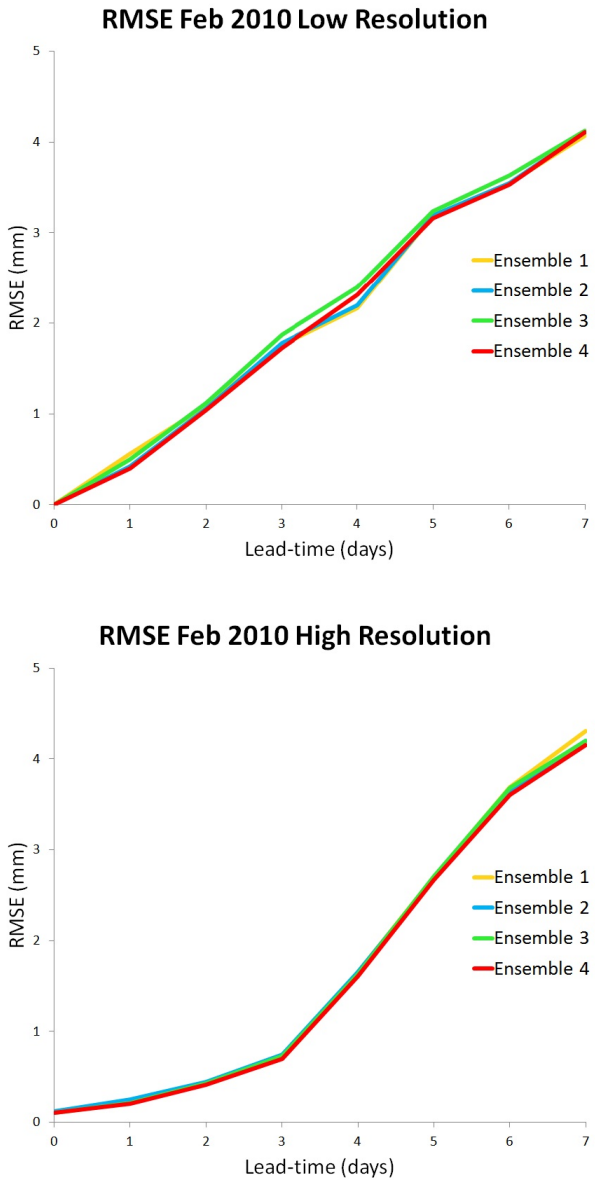

Figure 3.6.1: RMSE Graph for February 2010. Above: Low Resolution, Below: High Resolution.

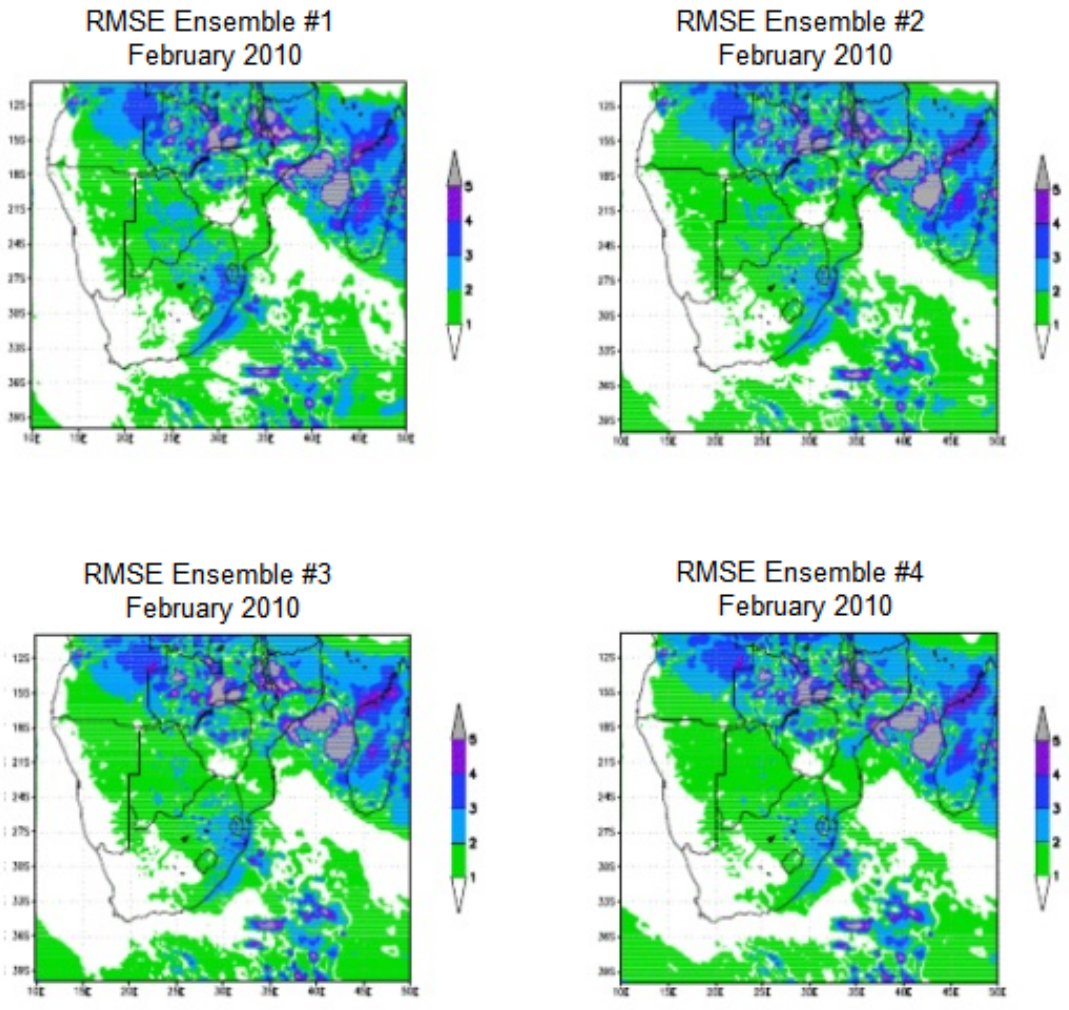

Figure 3.6.2: Spatial distribution of Error at High Resolution

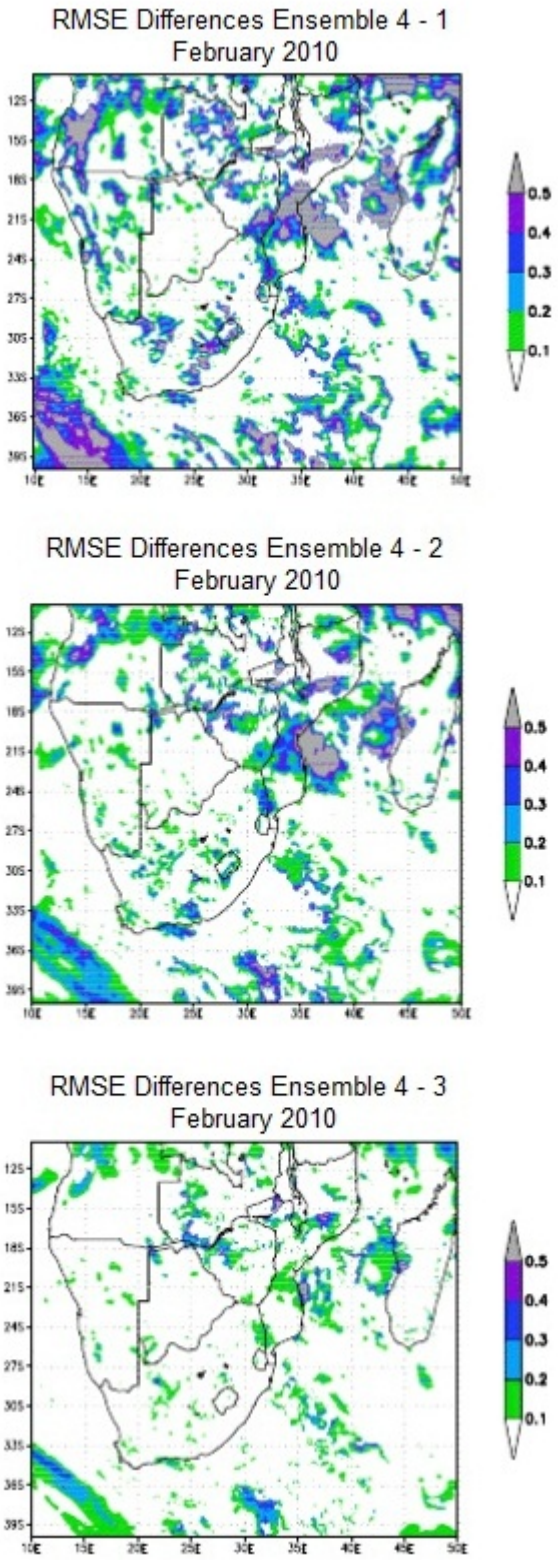

Figure 3.6.3: Difference in Error between Ensemble 4 and Ensemble 1, 2, 3 at High Resolution

3.7 Synopsis

This chapter documented the Root Mean Square Error results for the ensembles 1, 2, 3, and 4 at both high and low resolution, for January and February 2009 and January and February 2010. The figures showing the graphically represented average RMSE for the ensembles (Figures 3.3.1 and 3.3.2, 3.4.1 and 3.4.2, 3.5.1 and 3.5.2, 3.6.1 and 3.6.2) show that the high resolution forecast consistently has lower error than the low resolution forecasts. Up to the three-day lead time the high resolution forecasts have error of less than 1 mm for all ensembles which is a very small margin of error, indicating good accuracy. The graphs of the RMSE curves show that the low resolution forecasts have a constant increase in error for forecasts made for after day 1, showing an immediate drop-off in accuracy for all four ensembles with an increase in lead-time. The high resolution forecasts made up to day 4 have little error, followed by a sharp increase in error from day 4 to day 7 indicating a drop-off in accuracy for forecasts made by all four ensembles after a four-day lead time.

The figures of the spatial representations of the RMSE of the high resolution forecasts (Figures 3.3.3, 3.4.3, 3.5.3, 3.6.3) show that there is very little difference in error between the four ensembles. The area of largest error is in the North-Eastern region of the southern African domain and the area of least error is the Western coastal region of South Africa, Namibia and South-West Angola. These errors can be contributed to the fact that the Northern regions have summer rainfall heavily influenced by activity from the tropics and the Indian Ocean, which can be highly variable convective systems. Very high resolution models would be needed in order to minimise error over these regions. The West coast of South Africa, Namibia and South Angola have very little error because these regions experience extremely little summer rainfall. Therefore the error measured would only be for anomalous rainfall events that are unlikely to happen. This can be seen more clearly

in the RMSE Difference maps (Figures 3.3.4, 3.4.4, 3.5.4, 3.6.4) that show the difference in error between the high resolution forecasts made by Ensemble 4 and 1, Ensemble 4 and 2, and Ensemble 4 and 3. It is seen that in all four cases, the error difference between the ensembles decreases as the size of the ensemble increases.

The results in this chapter demonstrate the increase in accuracy of the forecasts when made at a high resolution as opposed to low resolution. As the spatial resolution is increased, so the error in the forecasts decrease. This result is consistently seen throughout this chapter.

A significant outcome that should be noted is that the cloud parameterisation schemes do not have a great effect on the configuration of the model. Since the ensemble sizes are already so small (Ensemble 1 has two members, Ensemble 4 has eight members), and the two different cloud schemes do not produce two significantly different ensemble members, the effective size of the ensembles is half the actual size. The effective size of Ensemble 1 is one member, and Ensemble 4 is four members. Since the ensembles in this study are so small, an increase in ensemble size by a few members has little effect in the error found in the forecasts produced.

The figures of spatial distribution support this conclusion since the difference in error between the smallest and largest ensemble is slight. However, it can also be seen that there is a small improvement from the smallest to the largest ensemble and therefore a much larger ensemble would most likely result in a great increase in accuracy and reduction of error in the forecasts.

Chapter 4

Probabilistic Verification

4.1 Introduction

Probabilistic verification of weather forecasts is a statistical approach at assessing either the value of a forecast, or the ability of a model to accurately forecast the weather (Wilks, 2006). Ensemble forecasting provides a basis for probabilistic forecasting (Kalnay, 2003). Probabilistic forecasting results in a range of possible outcomes which then need to be verified statistically in order to accurately assess the quality of the forecast, and the ability of the model to produce valuable forecasts. These probabilistic verification methods are used to determine various attributes of an ensemble forecasting systems, including the skill, reliability, resolution and ability to discriminate. By analysing the likelihood of a forecast event (e.g.: rainfall above a certain threshold) compared with the observations, statistical methods can be implemented to determine how successful a model is.

This chapter describes the results of probabilistic verification performed on the forecasts made by Ensemble 1, 2, 3 and 4, at high and low resolution for January and February, 2009 and 2010. The verifications are: the Brier skill score, The relative operating char-

acteristics, and the reliability. These will assess the skill, discrimination and reliability of the forecasts made by each ensemble at low and high resolution and determine which ensemble and which resolution produce the most accurate rainfall forecasts.

4.2 Brier Skill Score

The Brier skill score is a measurement of the accuracy of the model forecast relative to that of a reference forecast, in this case chosen as the persistence forecast. This skill indicates how much better the forecasts made by the model are in comparison to the method of persistence, and is therefore also an indication of how useful the model forecasts are. A perfect BSS is equal to 1, with no skill equal to 0. Short-range weather prediction models have a natural decrease in skill with an increase in lead-time. This is expected since forecasts made for long-lead times are more affected by the chaotic nature of the atmosphere and the uncertainties in the model than short lead-times because the model cannot perfectly predict future changes in atmospheric conditions.

For this study, the BSS has been calculated for all four ensemble systems, at high and low spatial resolution, for the months of January and February, 2009 and 2010. The results are presented here graphically, showing the measure of skill for each case. For each of the two spatial resolutions, the BSS curves of the four ensembles are superimposed onto the graphs in order to easily interpret the effect of ensemble size on forecast skill.

It can be seen from Figures 4.2.1 - 4.2.4 that for all four months, and at both resolutions, all four ensembles demonstrate similar skill. The figures also illustrate the drop-off in skill as the lead-time increases which is expected due to the difficulty in predicting an inherently chaotic atmosphere. It can be seen that day 1 has the highest skill, and day 7

has the lowest skill (except for Fig. 4.2.3.b)), with an anomalous drop in skill at day 3 in Fig. 4.2.1, Fig. 4.2.2.b) and Fig. 4.2.3. This could be due to model errors which affect prediction skill at day 3, or observation errors.

Figure 4.2.1 of January 2009 shows that the drop-off in skill is much the same for the four ensembles in both (a) and (b). When comparing the low resolution forecasts to the high resolution, the shape of the curves are once again much the same indicating that there is little improvement in the skill of the forecast system with an increase in horizontal resolution or an increase in ensemble size.

In Figure 4.2.2 (a), Ensemble 4 has a slow drop-off in skill, whereas the smallest ensemble, Ensemble 1, has a dramatic drop in skill after day 4. In Figure 4.2.2 (b) Ensemble 4 starts with a much higher skill and remains consistently more skilful than the other three ensemble systems for all seven days.

For January 2010, Figure 4.2.3 (a) shows identical skill for all ensembles until day 3, subsequently from day 3 to day 7 Ensemble 4 displays the most skill. The skill curves in Figure 4.2.3 (b) indicate that Ensemble 4 has a marginally higher skill than the other three ensembles for the 7 day lead time.

The final figure, Figure 4.2.4, shows the skill during February 2010. The four ensembles in (a) have a gradual drop-off in skill with lead-time. Ensemble 1 has the most dramatic reduction in skill with time compared to the other three ensembles. In (b) all four ensemble systems have near-identical skill curves.

This analysis of the BSS shows how altering ensemble size and spatial resolution affects the skill of the model systems. The ensemble size can be concluded to have little effect on the skill of the forecast, with the largest EPS having slightly improved skill over the others in some cases. In order to further increase the skill of the EPS, more ensemble members need to be introduced to contribute to the system. The high spatial resolution

forecasts display better skill than the low resolution forecasts in some cases, but they are generally seen to be similarly skilful. The interpretation thereof is that the resolution of the model must be increased further in order to increase the skill of the forecasts.

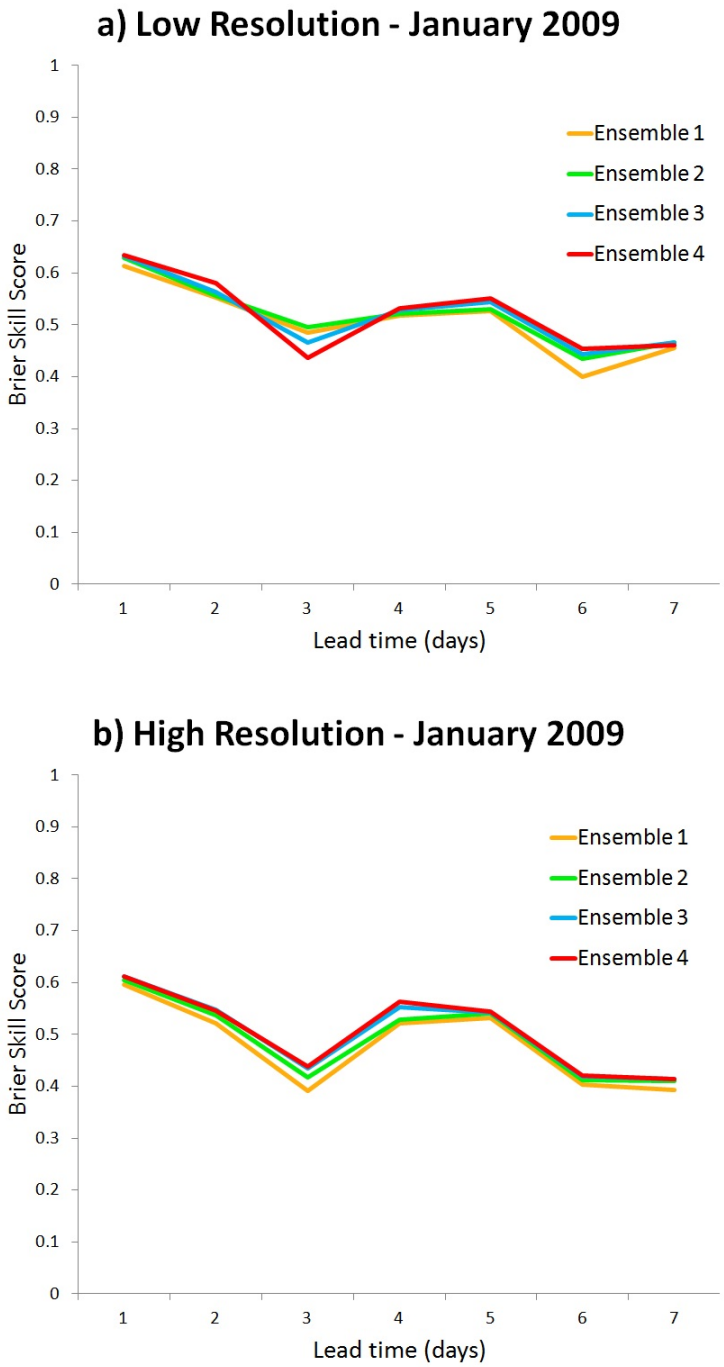

Figure 4.2.1: Brier skill score curves for January 2009: a) Low Resolution and b) High Resolution

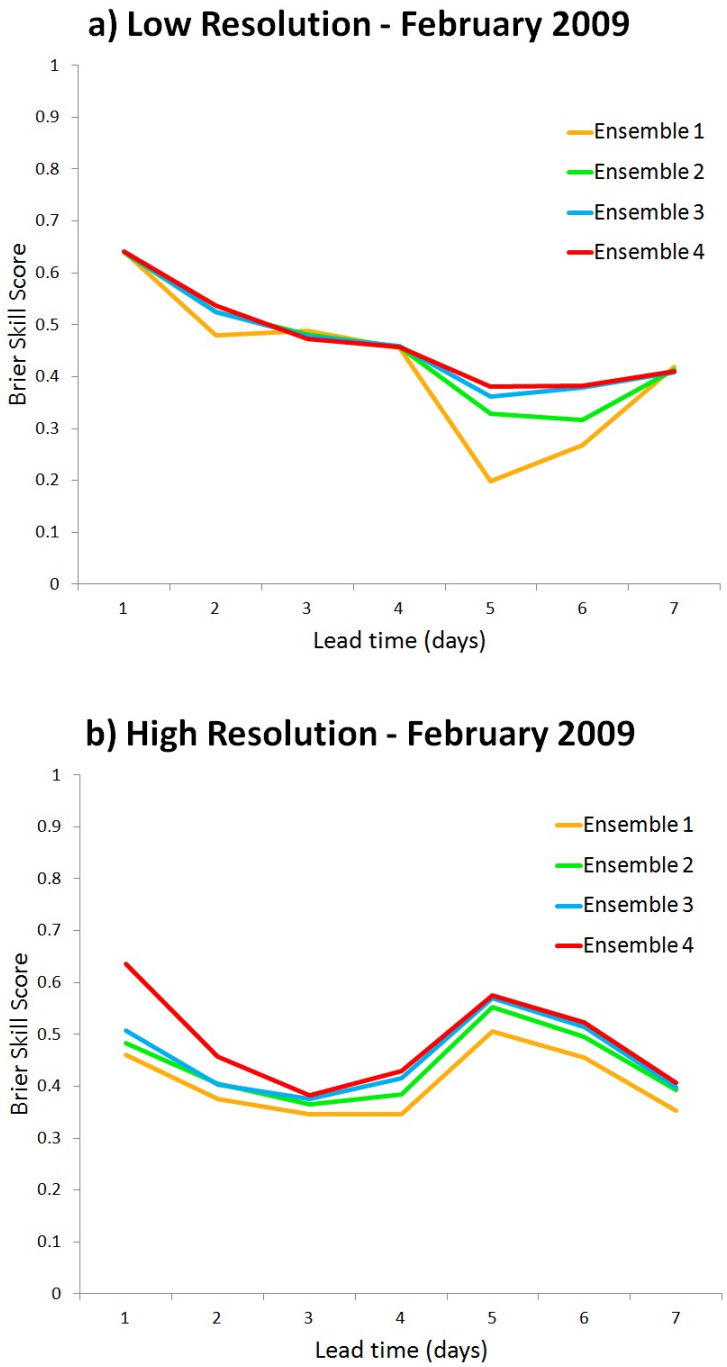

Figure 4.2.2: Brier skill score curves for February 2009: a) Low Resolution and b) High Resolution

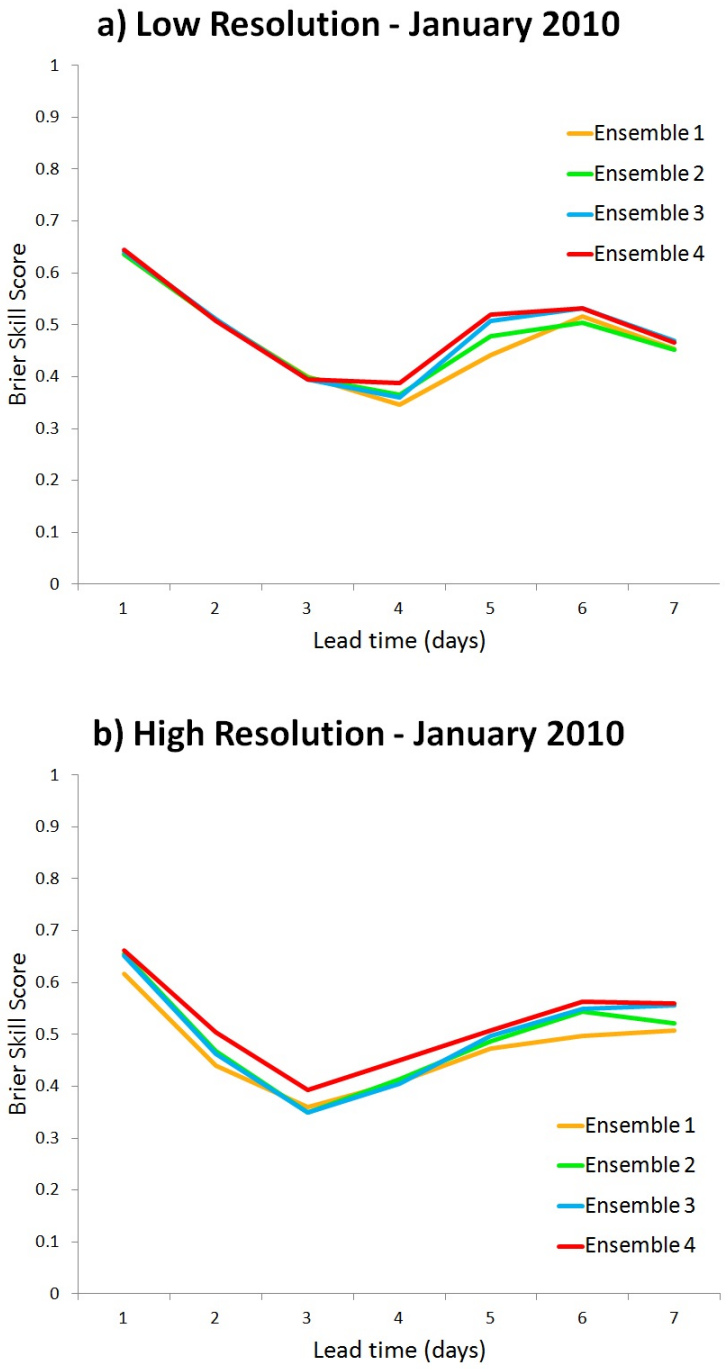

Figure 4.2.3: Brier skill score curves for January 2010: a) Low Resolution and b) High Resolution

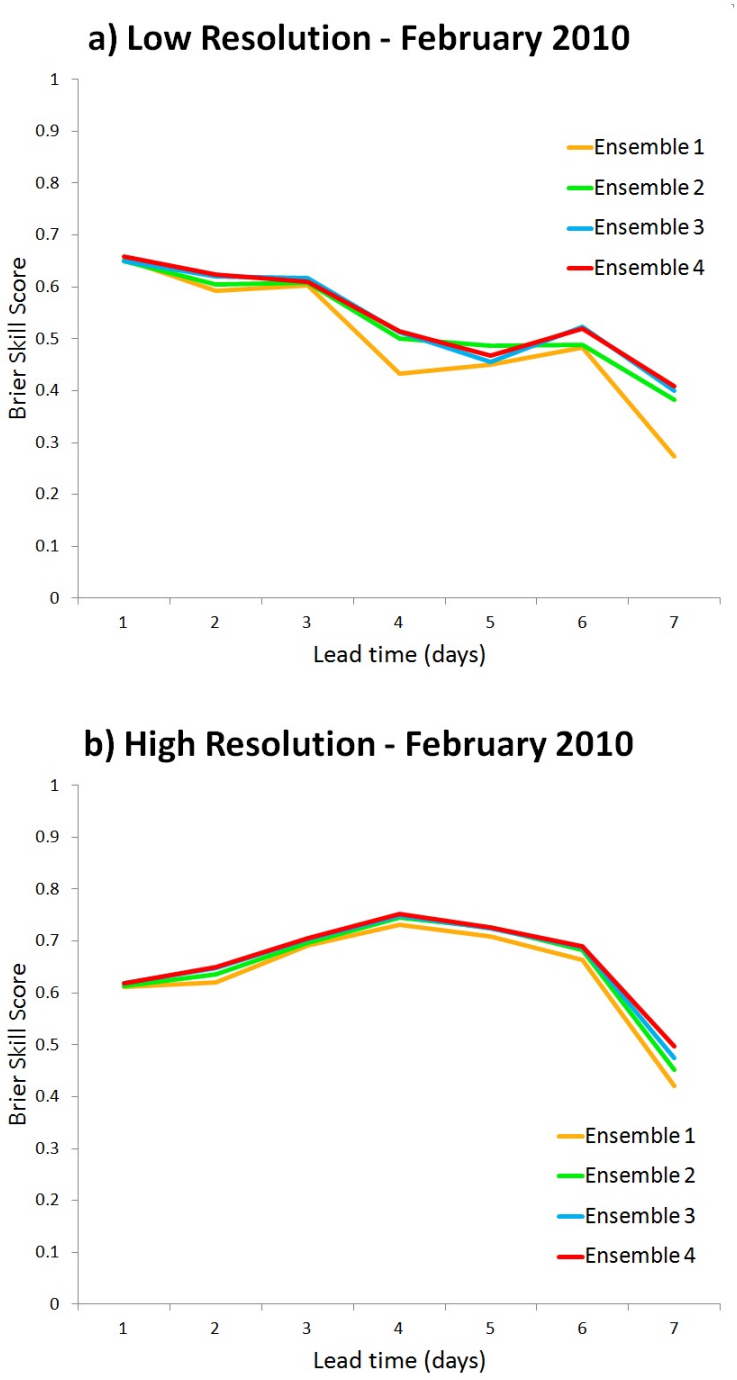

Figure 4.2.4: Brier skill score curves for February 2010: a) Low Resolution and b) High Resolution

4.3 Relative Operating Characteristics

As discussed in Chapter 2 (Methodology), the relative operating characteristics (ROC) verification score measures the ability of a forecast system to discriminate between events ("yes") and non-events ("no"), usually at a certain threshold (e.g.: 5 mm of rainfall). The skill of a forecast when using the ROC is calculated using Table 2.1, the contingency table. In this chapter the ROC is used to measure the ability of the EPSs to discriminate two thresholds:

1. Discrimination of 1 mm rainfall.
2. Discrimination of 5 mm rainfall.

The first threshold of 1 mm was chosen since in this study, any recorded precipitation of less than 1 mm was considered to be no rainfall. Therefore a threshold of 1 mm assesses how well the EPSs discriminate between rainfall and no rainfall. The second threshold of 5 mm was chosen as the threshold of heavier rainfall. This limit assesses the ability of the EPSs to discriminate between light and more heavy rainfall events.

4.3.1 Discrimination of 1 mm of rainfall

This section discusses the ability of the EPSs to discriminate precipitation events: rainfall or no rainfall. The area under the ROC curves were calculated using the trapezium method and the results compiled in Table 4.1. From Table 4.1 it is clear that the high resolution forecasts far out-performed the low resolution forecasts in every case. For all EPSs, Table 4.1 documents an increase in area with an increase in ensemble size: Ensemble 4 has the largest area under the ROC curve in all cases.

Table 4.3.1: Areas under the ROC curves for the 1 mm rainfall discrimination

January 2009				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.525	0.512	0.513	0.529
High Resolution	0.701	0.691	0.705	0.718
February 2009				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.548	0.548	0.554	0.548
High Resolution	0.763	0.849	0.849	0.854
January 2010				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.527	0.540	0.546	0.550
High Resolution	0.810	0.852	0.858	0.862
February 2010				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.555	0.553	0.565	0.563
High Resolution	0.798	0.806	0.811	0.815

Figures 4.3.1 - 4.3.4 show the ROC graphs of each EPS for each month, with the high and low resolution forecasts superimposed. Figure 4.3.1 - 4.3.4, (a) - (d), all show a large difference between the high and low resolution systems. The high resolution curves are more rounded and approaching the top left corner, whereas the low resolution curves are close to the diagonal line of no skill. The conclusion that can be drawn from this is that an increase in spatial resolution will most likely result in an increase in ROC score, irrespective of ensemble size.

In Figure 4.3.1 it can be seen that the high resolution curves in (a) - (c) do not change noticeably with an increase in ensemble size, and neither do the low resolution curves. Moving to Figure 4.3.1 (d), the high and low resolution curves both show a slight improvement from the previous graphs.

In Figure 4.3.2 the graphs from (a) - (d) illustrate a steady improvement in skill for the high resolution curves with an increase in ensemble size, specifically there is a jump in skill between the (a) and (b) high resolution curve. In all four graphs the low resolution curves to not change noticeably. There is a marked difference in skill between the high and low resolution curves in all four graphs, with the high resolution curves largely more skilful than the low resolution.

Figure 4.3.3 (a) - (d) show a gradual increase in skill for the low resolution curves. The high resolution curves improve more pronouncedly with an increasing ensemble size from graphs (a) to (d). In all four graphs there is a vast difference in skill between the high and low resolution curves, with the high resolution forecasts having much better skill.

The last figure, Figure 4.3.4 (a) - (d), both the high and low resolution curves improve gradually with an increase in EPS size.

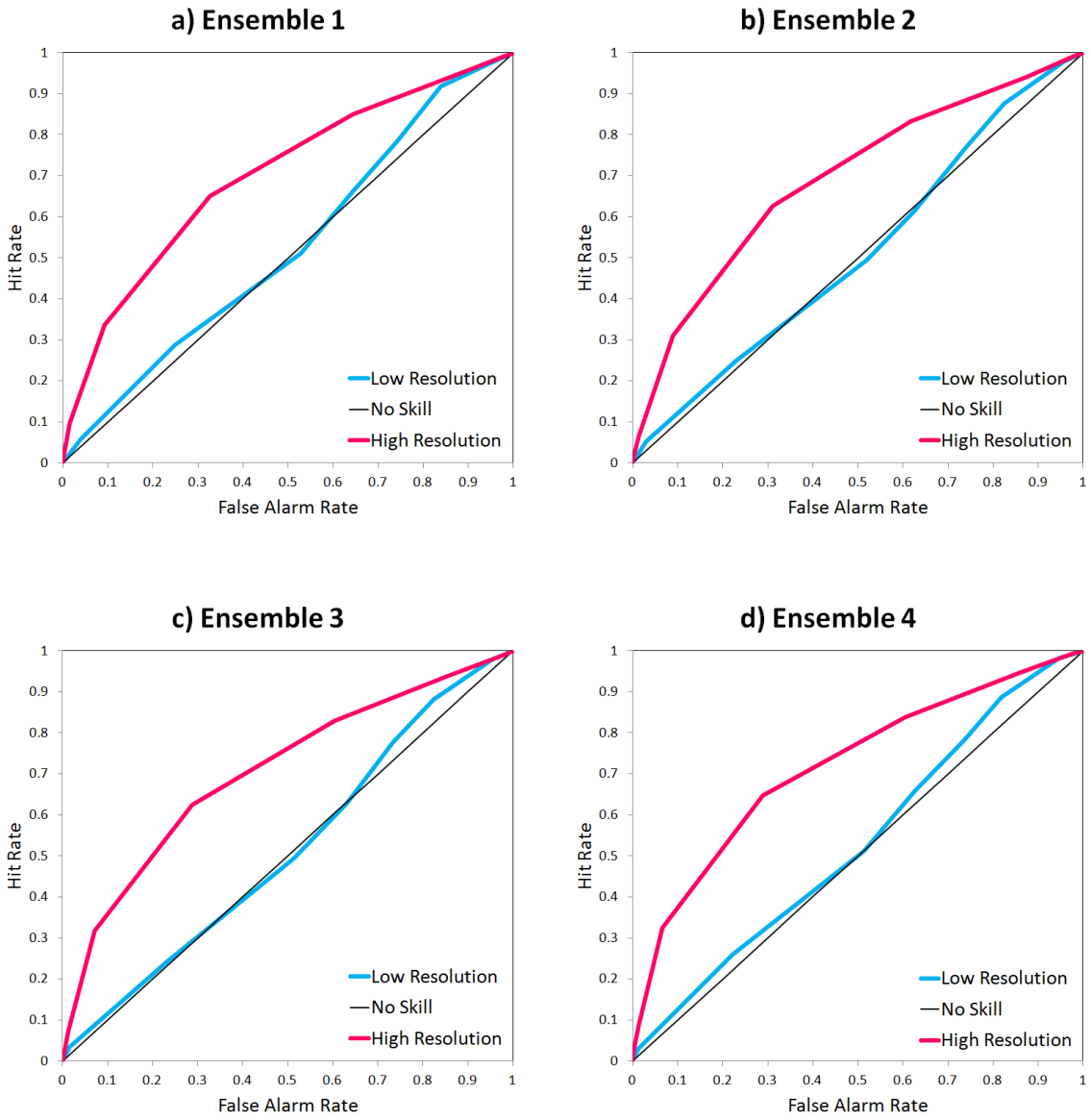

Figure 4.3.1: 1 mm Threshold ROC Diagrams for January 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

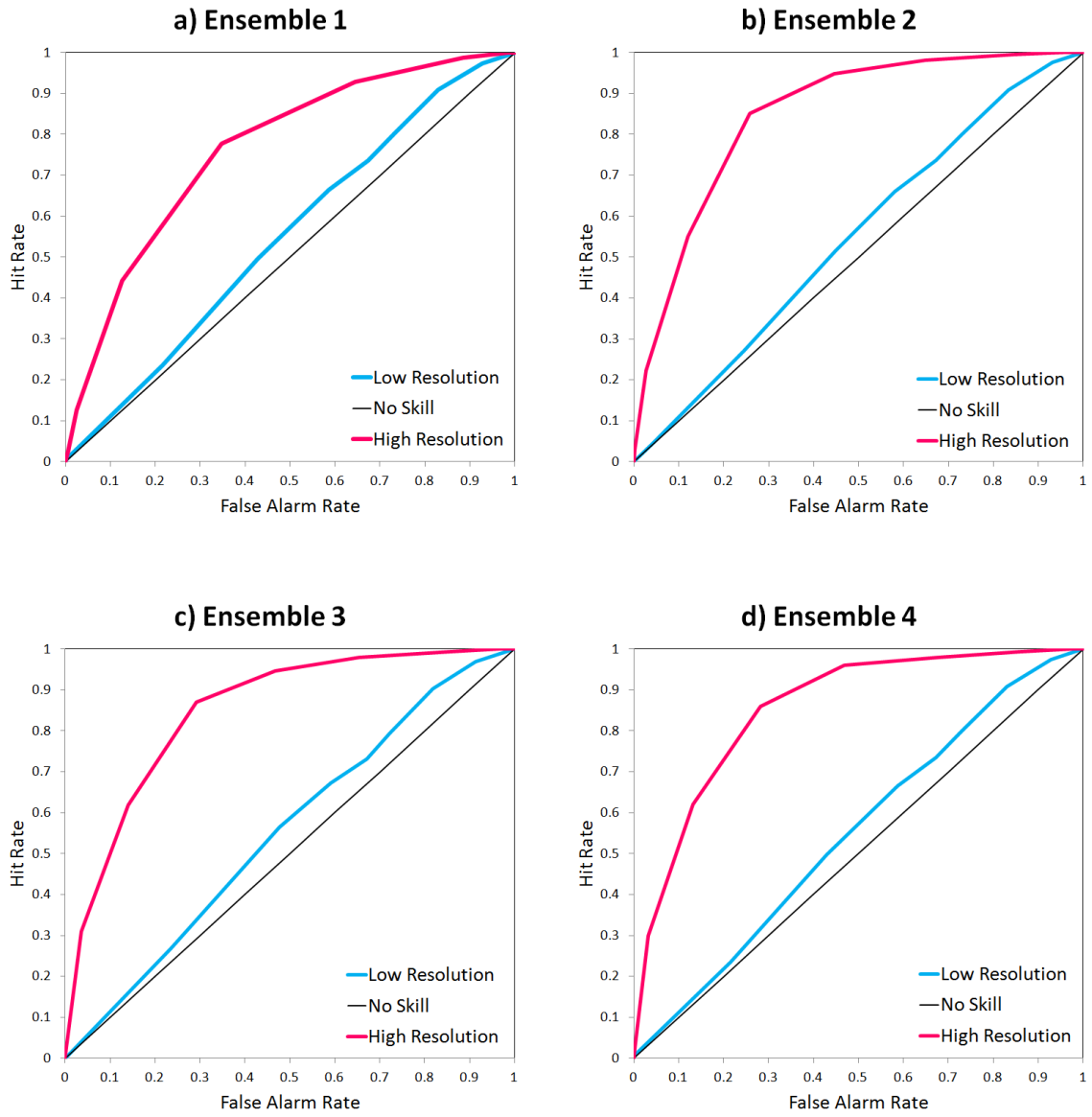

Figure 4.3.2: 1 mm Threshold ROC Diagrams for February 2009: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

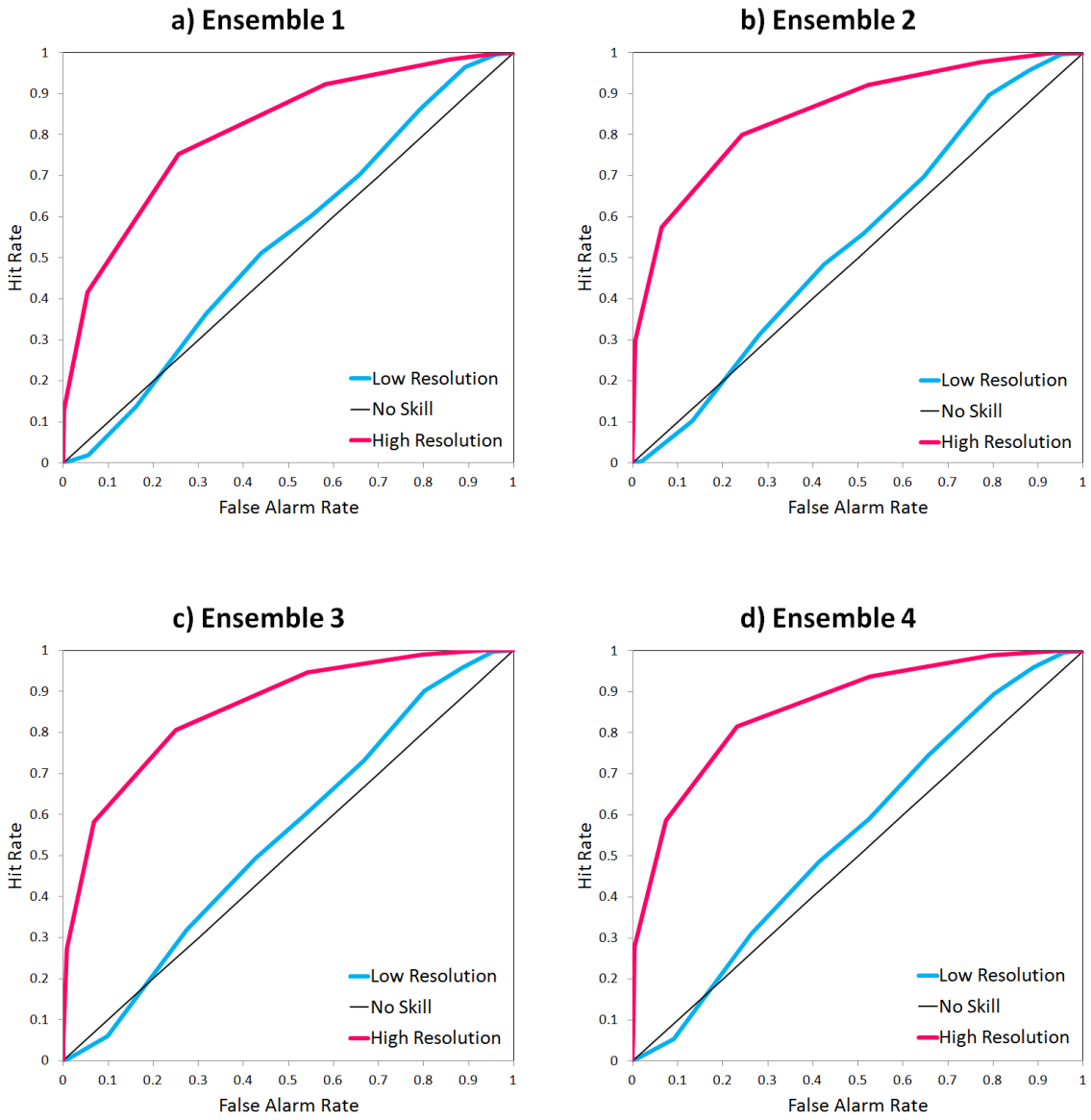

Figure 4.3.3: 1 mm Threshold ROC Diagrams for January 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

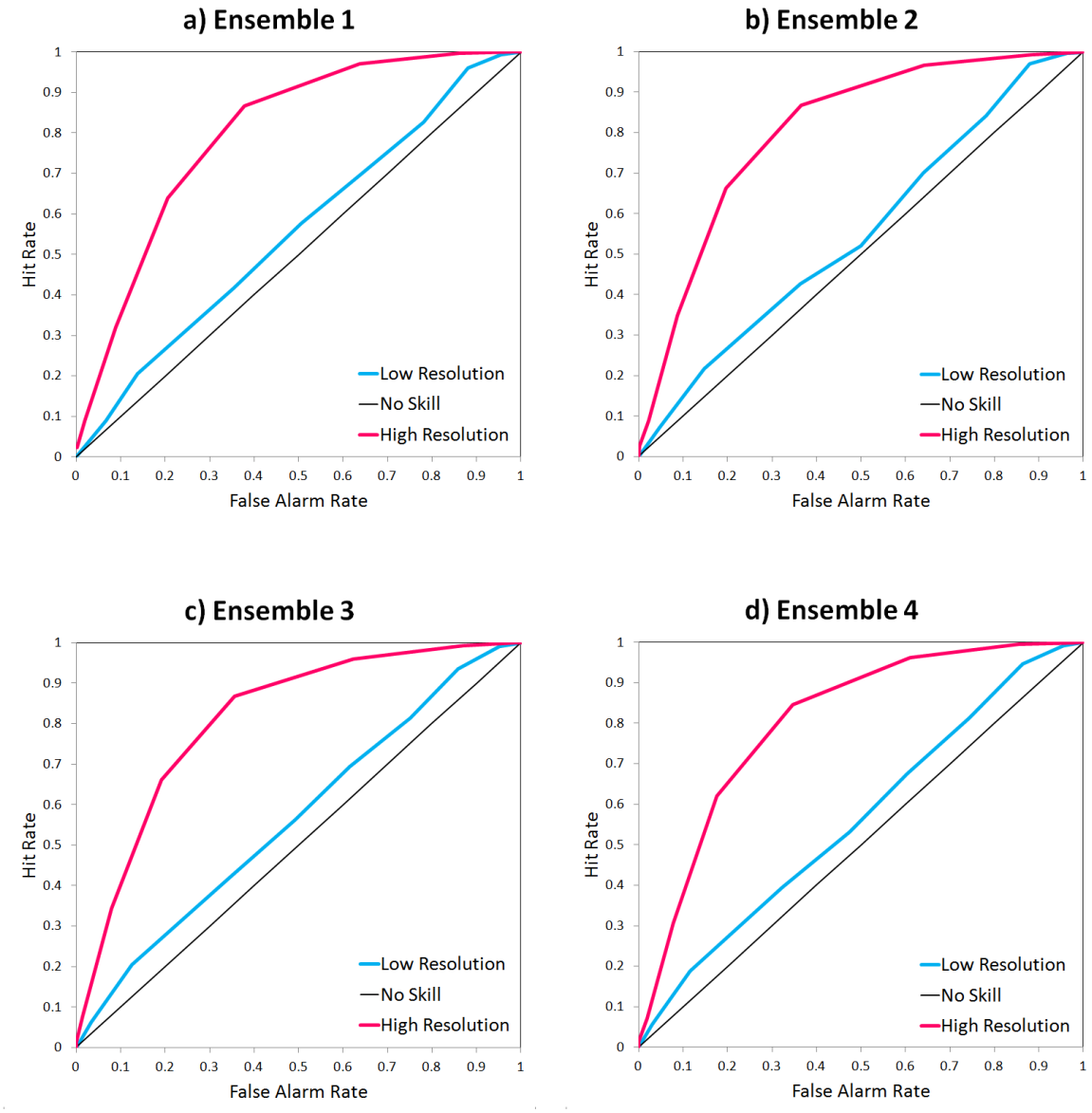

Figure 4.3.4: 1 mm Threshold ROC Diagrams for February 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

4.3.2 Discrimination of 5 mm of rainfall

This section analyses how well the EPSs perform in discriminating precipitation events: rainfall above 5 mm or rainfall below 5 mm. The area under the ROC curves were calculated using the trapezium method and the results compiled in Table 4.2.

Table 4.3.2: Areas under the ROC curves for the 5 mm rainfall discrimination

January 2009				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.587	0.587	0.593	0.717
High Resolution	0.869	0.876	0.870	0.871
February 2009				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.511	0.5	0.5	0.5
High Resolution	0.701	0.734	0.729	0.732
January 2010				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.556	0.554	0.556	0.562
High Resolution	0.862	0.886	0.877	0.882
February 2010				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.528	0.524	0.535	0.565
High Resolution	0.677	0.680	0.682	0.701

It is clearly recorded in Table 4.2 that there is a large discrepancy between the high and low resolution forecasts, and the high resolution forecasts out-perform the low resolution forecasts in every case. Also evident in Table 4.2 is a trend of increase in ROC area from ensemble 1 to ensemble 4. The highest values for area are seen either for ensemble 2

or ensemble 4, but the results are extremely close. The increase in area with ensemble size is clearer in the low resolution ensemble systems. The change is not large, but even this small growth of the area indicates that an increase in ensemble size results in an improvement in the discrimination ability of the ensemble systems.

Figures 4.3.5 - 4.3.8 show the ROC graphs of each EPS at high and low resolution, for each month. All figures, Figure 4.3.5 - 4.3.8, (a) - (d), illustrate a significantly large difference between the high and low resolution systems, where the high resolution curves capture far more area than the low resolution. This difference in area indicates that there is a considerable increase in skill with an increase in ensemble size.

In Figure 4.3.5, the low resolution curves remain steady for graphs (a) - (c) and suddenly there is a big increase in area in graph (d). This demonstrates the increase in skill with an increase in ensemble size, for the low resolution EPSs. All four of the high resolution forecasts (a) - (d) have similar areas suggesting that the increase in ensemble size has no effect on the skill. However the area under the high resolution curves is much greater than the low resolution, showing improvement with increase in spatial resolution.

In Figure 4.3.6 the graphs from (a) - (d) show no skill for the low resolution forecasts, since their curves lie exactly on, or just above, the diagonal. The high resolution curves all ((a) - (d)) have almost consistent area, showing that there is little improvement in forecast skill with an increase in spatial resolution.

Figure 4.3.7 (a) - (d) illustrates excellent skill for the high resolution forecasts. The areas below the curves are more than 0.8 but remain consistent for all four EPSs. There is therefore no improvement in skill with increase in ensemble size. There is, however, a vast difference in skill between the high and low resolution systems, with the low resolution forecasts being close to the line of no skill. There is thusly a large improvement in skill with spatial resolution.

The last figure, Figure 4.3.8 (a) - (d) has the high resolution forecasts performing

consistently for the first three ensemble systems, and then exhibiting a noticeable improvement in skill with ensemble size in graph (d), where the highest skill corresponds with the largest ensemble. The low resolution curves are little above the line of no skill, suggesting little or no improvement with an increase of ensemble size. The high resolution curves out-perform the low resolution curves in all four cases.

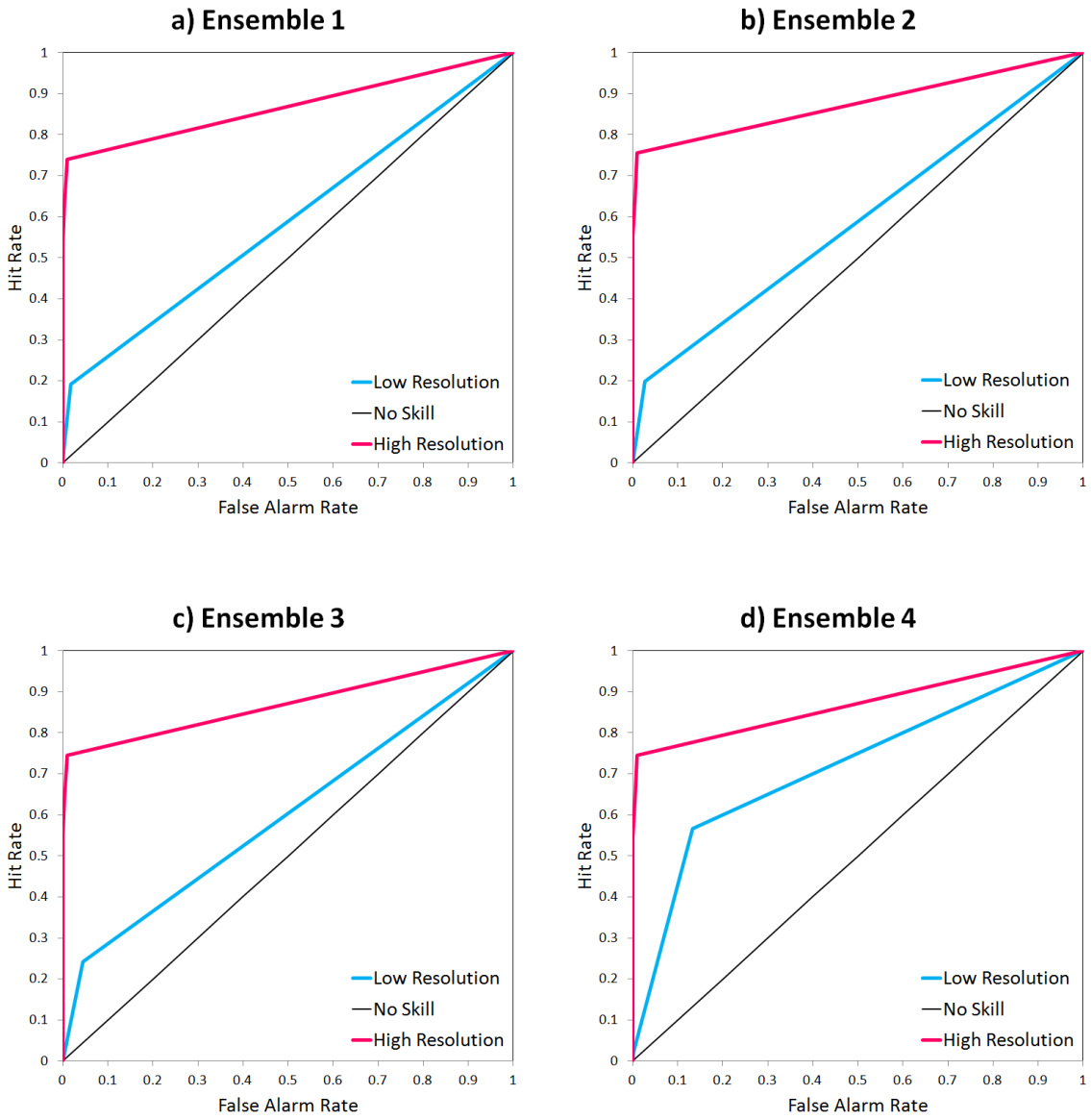

Figure 4.3.5: 5 mm Threshold ROC Diagrams for January 2009: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

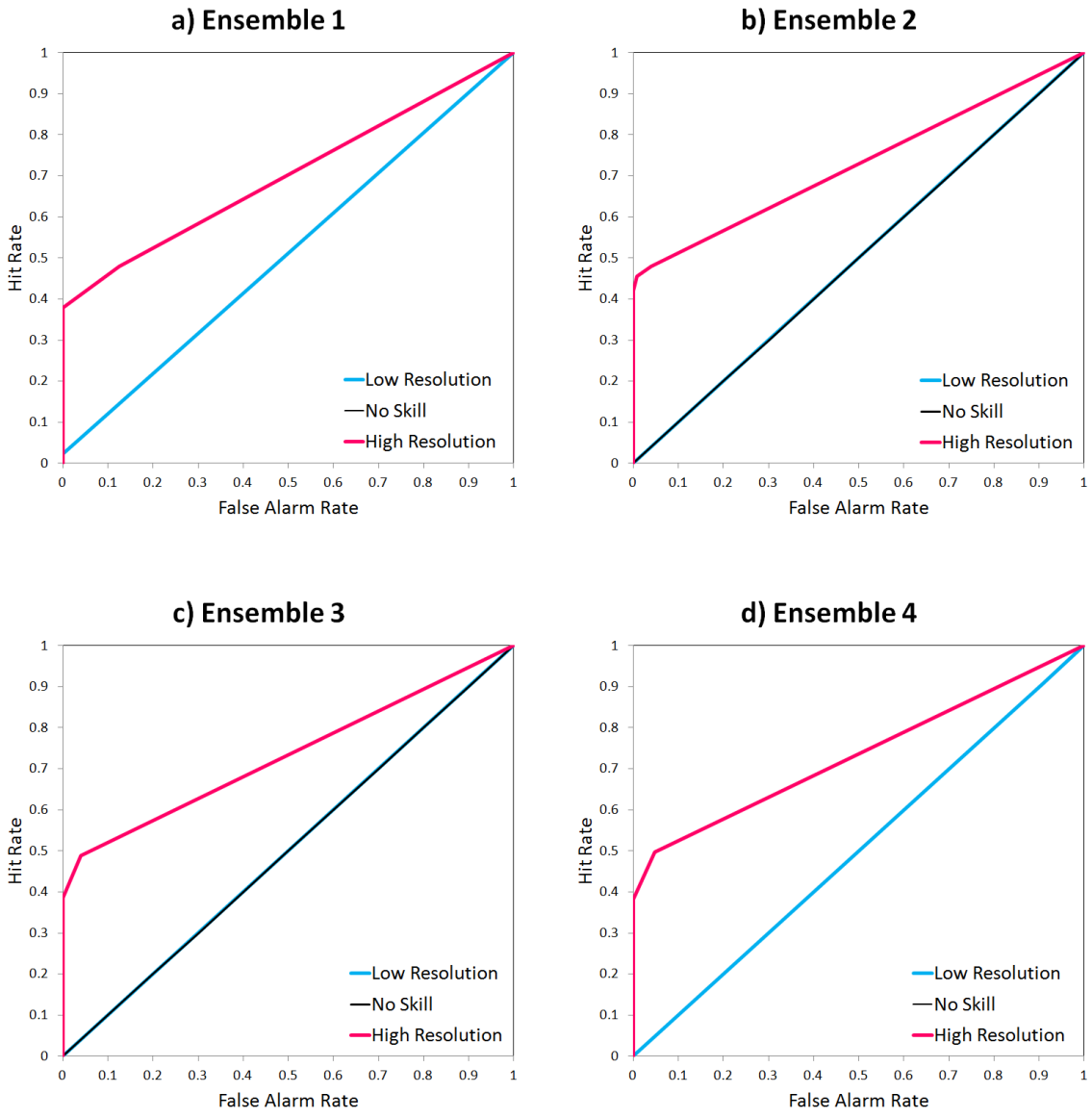

Figure 4.3.6: 5 mm Threshold ROC Diagrams for February 2009: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

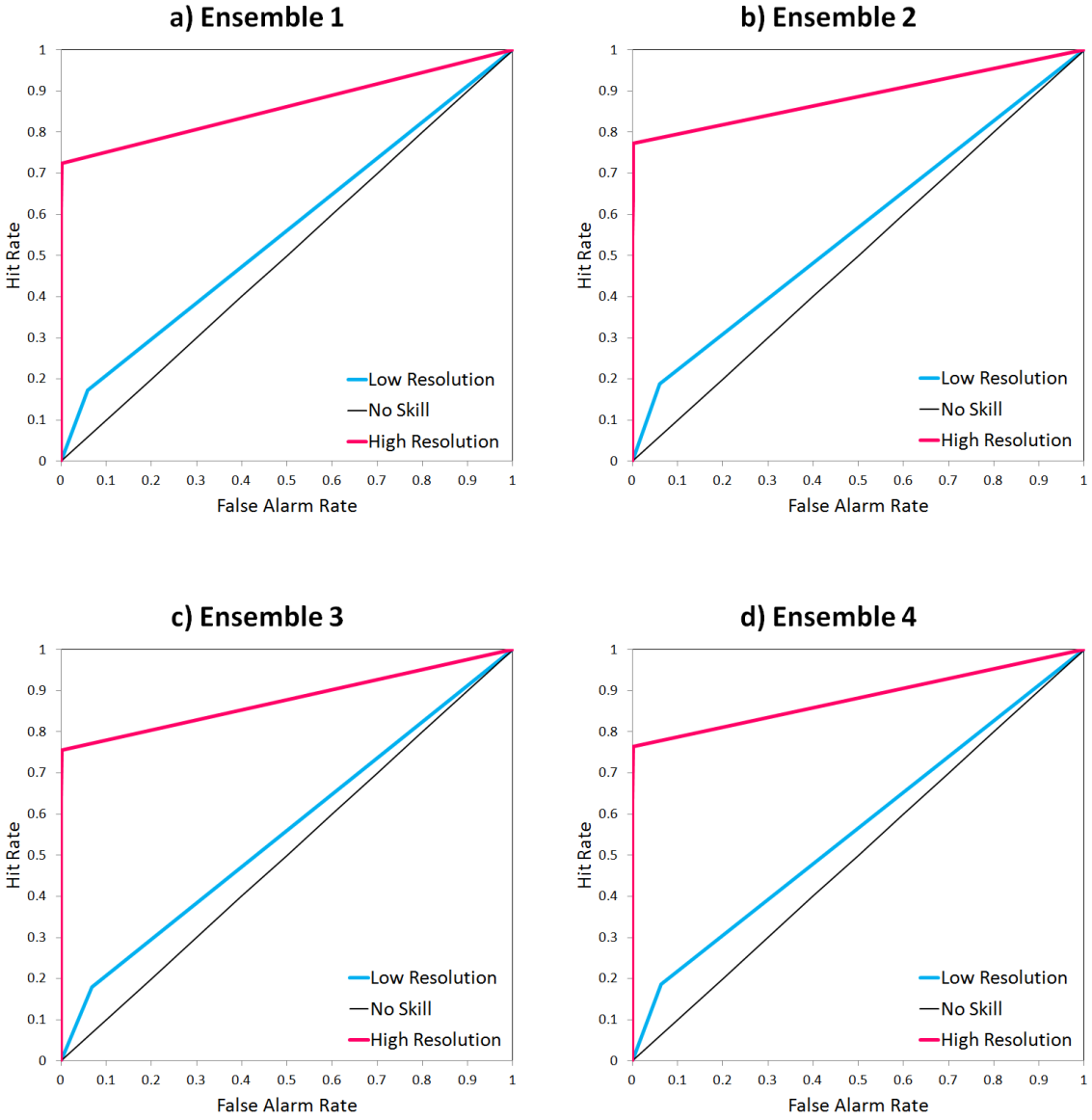

Figure 4.3.7: 5 mm Threshold ROC Diagrams for January 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

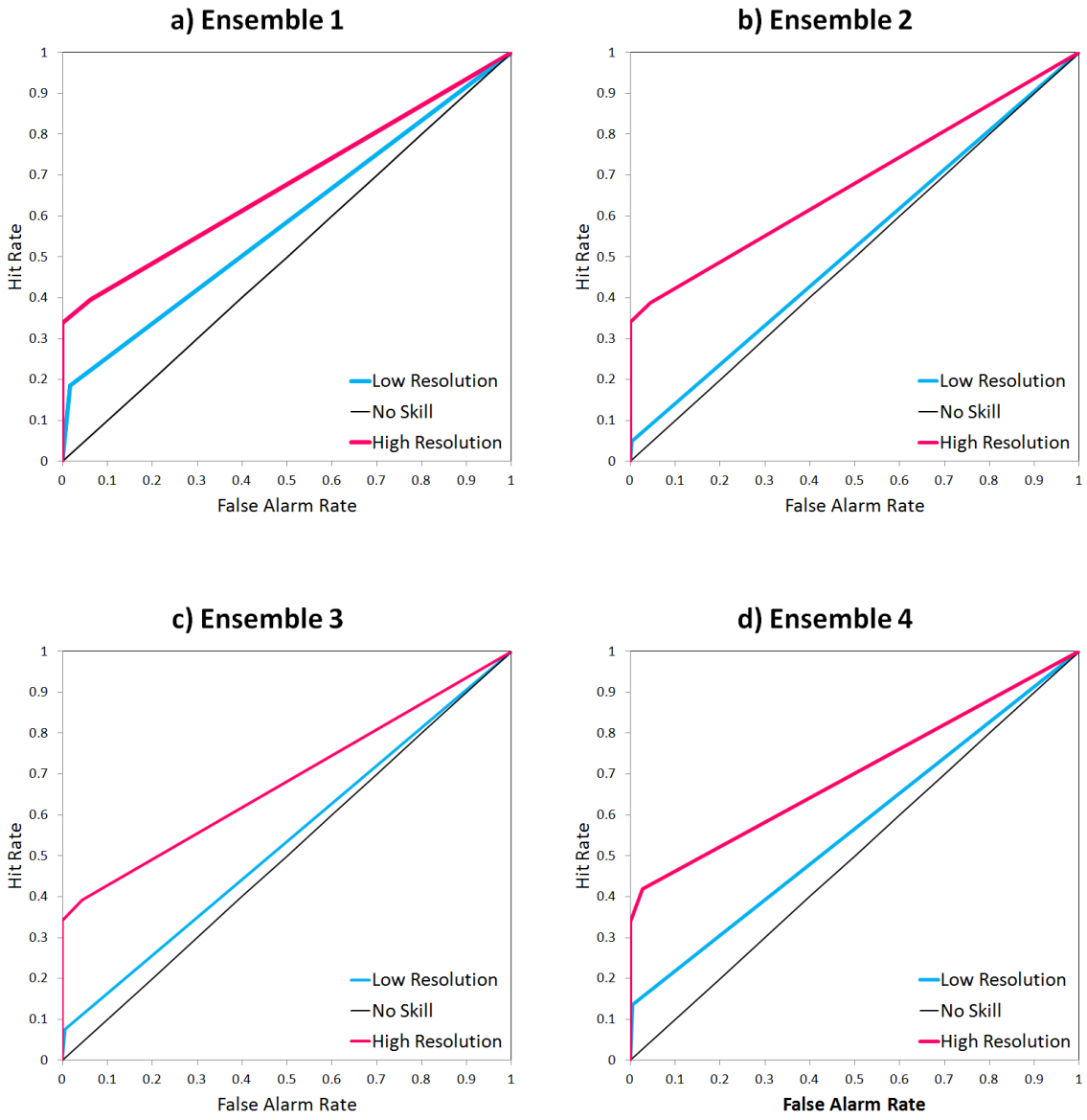

Figure 4.3.8: 5 mm Threshold ROC Diagrams for February 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

4.4 Reliability

The reliability diagram is an indication of the correspondence between the Forecast Probability and the Observed Relative Frequency of the ensemble system. Perfect reliability is represented as the diagonal line from (0, 0) to (1, 1) with a gradient of 1. This analysis explores the reliability of each of the four ensemble systems, for each month, for low and high resolution. The individual reliability points are averaged into a straight line: the closer the gradient of this straight line is to 1 (perfect reliability), the better the reliability. The gradients of each reliability line are compiled into Table 4.4 below.

Table 4.4.1: Gradient of the Reliability Diagrams

January 2009				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.838	0.852	0.826	0.844
High Resolution	1.263	1.285	1.215	1.205
February 2009				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.050	0.136	0.149	0.341
High Resolution	1.417	1.275	1.270	1.238
January 2010				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.713	0.809	0.839	0.902
High Resolution	1.243	1.376	1.329	1.269
February 2010				
	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Low Resolution	0.050	0.074	0.091	0.240
High Resolution	1.298	1.275	1.270	1.238

The four diagrams, Figures 4.4.1, 4.4.3, 4.4.5, 4.4.7, (a) - (d), and Table 4.4 show that for most cases the high resolution forecasts have a gradient closer to the line of perfect reliability than the low resolution forecasts. The months of February 2009 and 2010 display a huge improvement from the low to the high resolution forecasts. This indicates that, generally, there is an improvement of reliability with an increase in spatial resolution. There is also a general trend of improvement of reliability with an increase in ensemble size: in almost every case for both the low and high resolution systems the gradients of the reliability lines approach 1 as the ensembles become bigger. In all of the sharpness diagrams, Figures 4.4.2, 4.4.4, 4.4.6, 4.4.8, (a) - (b), none of the EPSs express good sharpness. The High resolution systems slightly out-perform those at low resolution since their values are closest to 1, but the ensemble size seems to have no impact on sharpness.

Figure 4.4.1 shows a steady trend from (a) - (d) of the high resolution reliability lines to approach the diagonal of perfect reliability, suggesting that an increase in ensemble size has a slight impact on reliability. The low resolution forecasts remain mostly steady in gradient, consistently close to the diagonal showing good reliability but no discernible improvement with a larger ensemble. The high resolution forecasts show consistently slight underconfidence (good resolution), and the low resolution forecasts show consistently slight overconfidence (poor resolution).

Figure 4.4.2 (a) - (b) is a bar graph of the sharpness of the graphs in Figure 4.4.1. The sharpness of the low resolution forecasts has peaks at the low forecast probability bins which indicates sharpness, but the rest of the bins are mostly constant in height indicating a lack of sharpness. The high resolution system does not display good sharpness although there is a peak in the second low forecast probability bin.

In Figure 4.4.3 (a) - (d) the low resolution forecasts display very little or no reliability. There is a trend of improvement in reliability with an increase in ensemble size, as can be seen from graph (a) to graph (d). However, the biggest improvement occurs with

an increase in resolution. The high resolution forecasts far out-score the low resolution forecasts, and also show a trend of improvement in reliability with ensemble size. These results suggest that there is a positive correlation between reliability and ensemble size, and a higher spatial resolution results in a much better reliability of the forecasts. The high resolution forecasts show consistent underconfidence (good resolution), and the low resolution forecasts show consistently very large overconfidence (poor resolution).

Figure 4.4.4 illustrates the sharpness of the systems for February 2009. Both the high and low resolution graphs show a peak at the first forecast probability bins and another peak near the last bins. This shows a general trend of sharpness, but the height of the bars in the central bins still results in these systems having low sharpness.

Figure 4.4.5 shows the low resolution forecasts approaching the diagonal line steadily from graph (a) to (d), with an increase in ensemble size. The high resolution graphs are closest to the diagonal in graph (a) and graph (d) which shows that an increase in ensemble size has no effect on the reliability. The low resolution actually out-performs the high resolution ensemble in graph (d). The high resolution forecasts show consistently slight underconfidence (good resolution), and the low resolution forecasts show consistently slight overconfidence (poor resolution).

The sharpness diagrams in Figure 4.4.6 are quite flat in both (a) and (b) indicating little or no sharpness. The high resolution graph (b) has higher peaks than the low resolution forecasts (b), which shows that it has better sharpness.

Figure 4.4.7 (a) - (d) show little or no reliability for the low resolution forecast systems. The gradients for these lines improve with ensemble size (from (a) to (d)), but they are still far from the line of perfect reliability. The high resolution forecasts have lines with gradients close to the diagonal, which indicate good reliability. There is also a slight improvement of reliability as the size of the ensembles increase, showing that ensemble size does positively affect the reliability. The high resolution forecasts show consistently slight

underconfidence (good resolution), and the low resolution forecasts show consistently very large overconfidence (poor resolution).

The sharpness diagrams in Figure 4.4.8 illustrates low sharpness for both (a) and (b) - the low and high resolution systems.

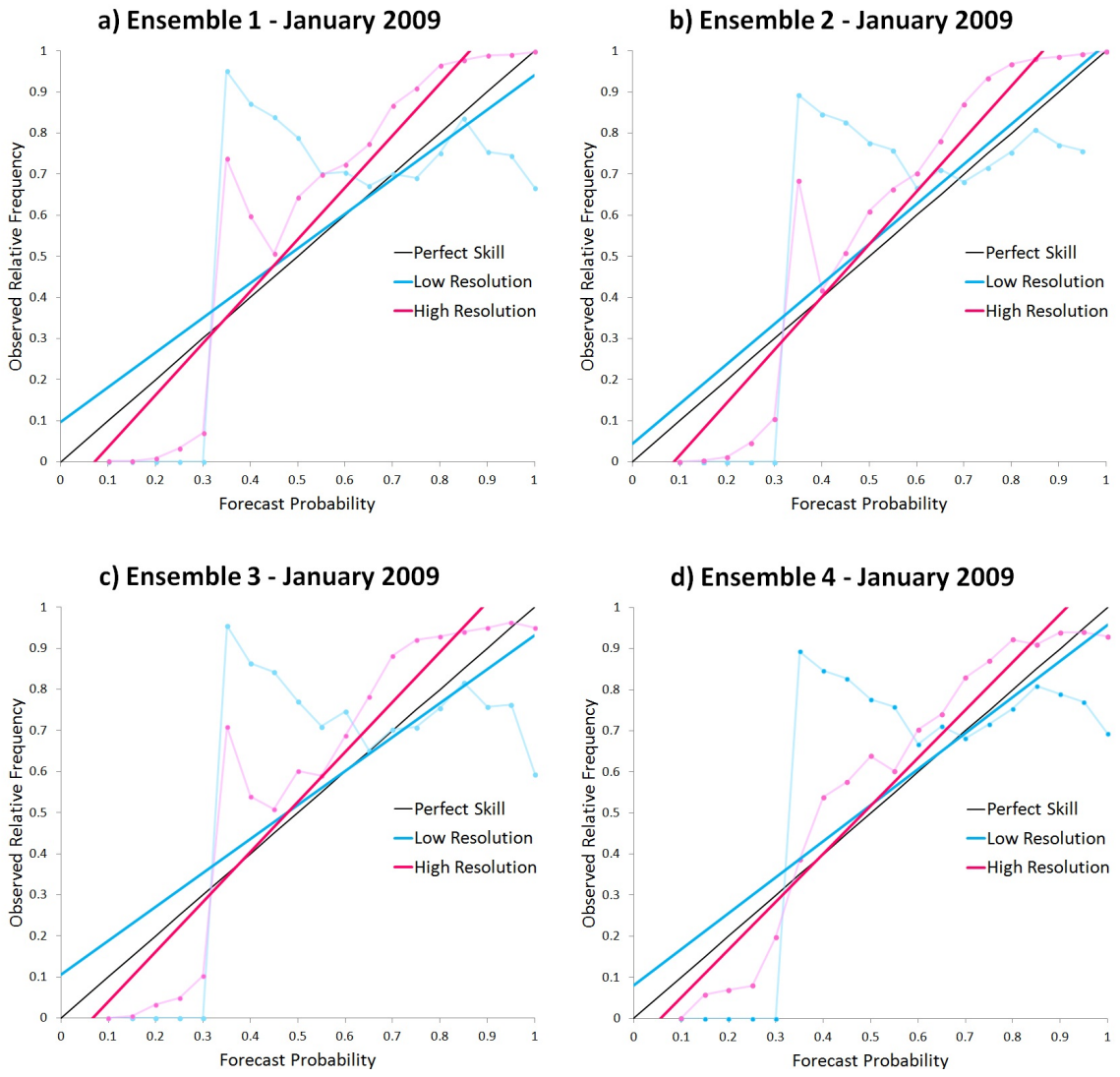

Figure 4.4.1: Reliability Diagrams for January 2009: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

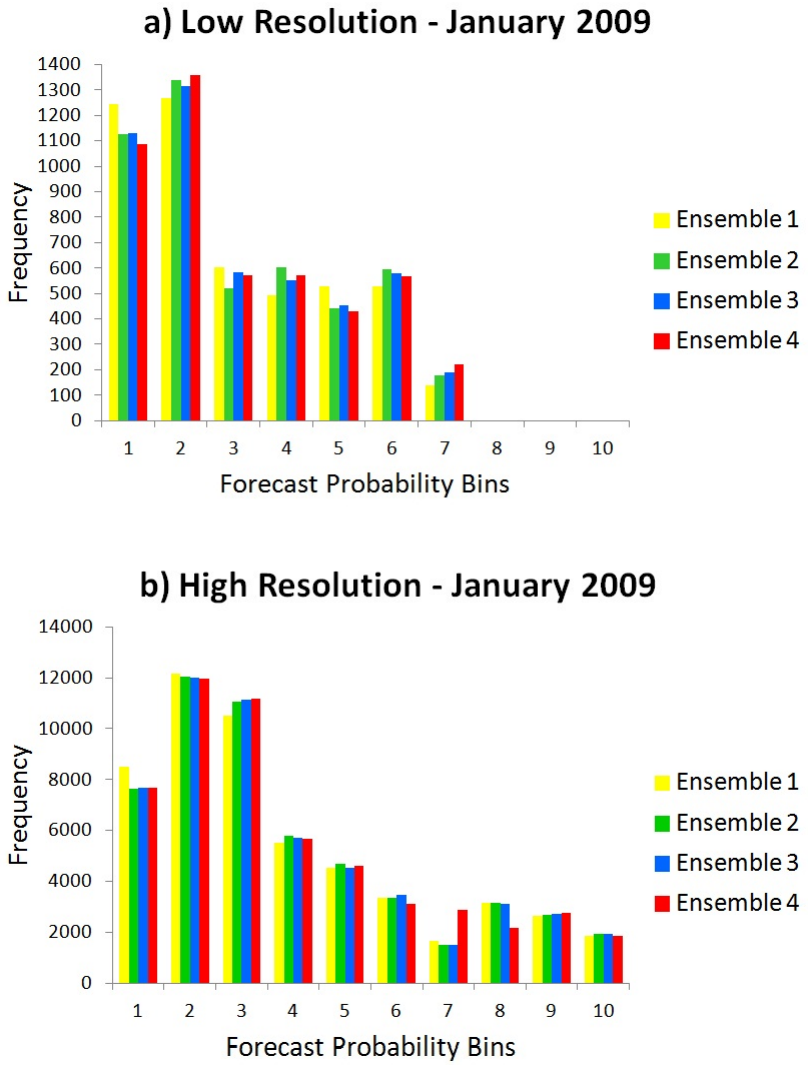

Figure 4.4.2: Sharpness Diagram for January 2009: a) Ensembles 1, 2, 3, and 4 at Low Resolution. b) Ensembles 1, 2, 3, and 4 at High Resolution.

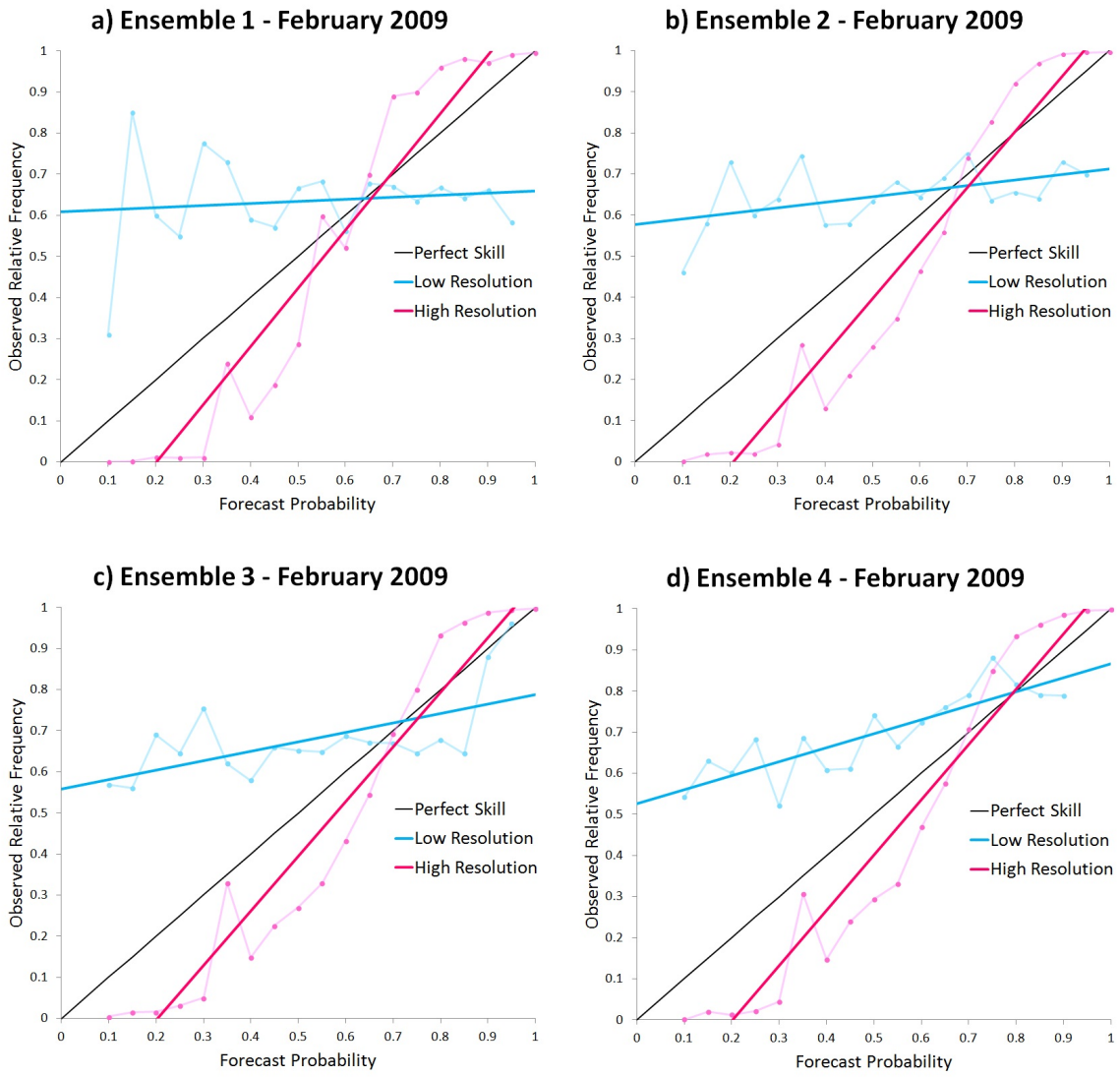

Figure 4.4.3: Reliability Diagrams for February 2009: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

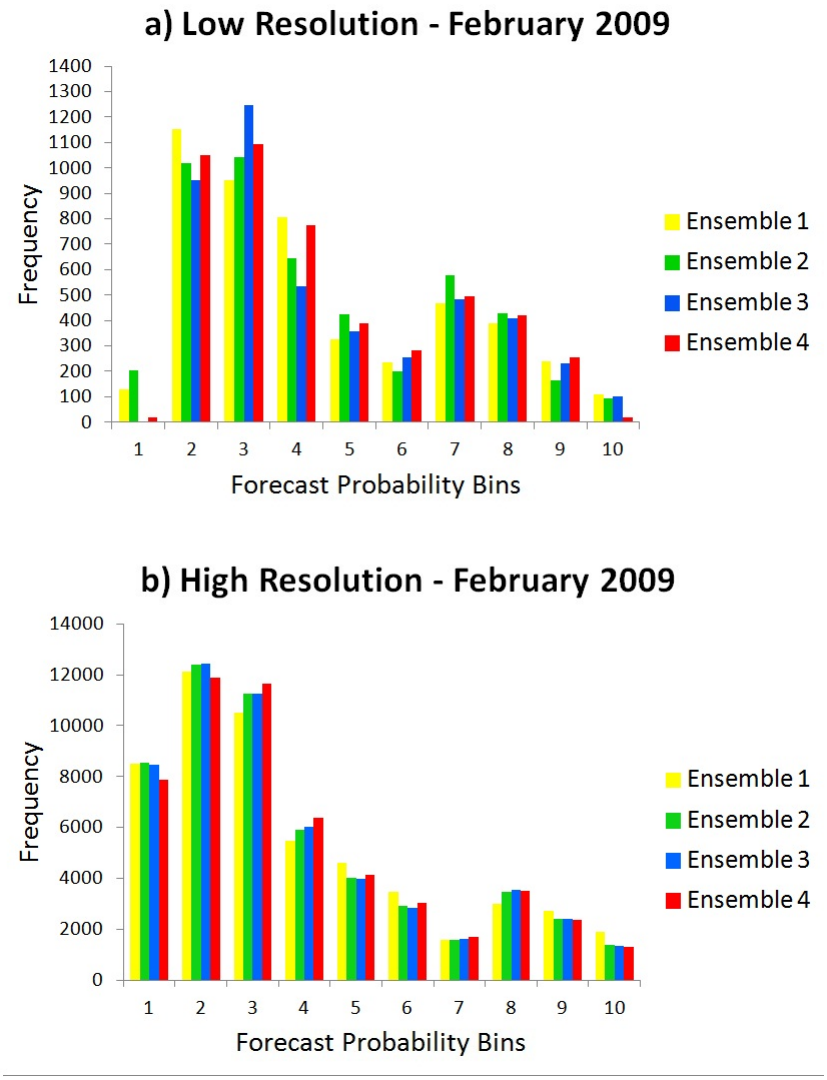

Figure 4.4.4: Sharpness Diagram for February 2009: a) Ensembles 1, 2, 3, and 4 at Low Resolution. b) Ensembles 1, 2, 3, and 4 at High Resolution.

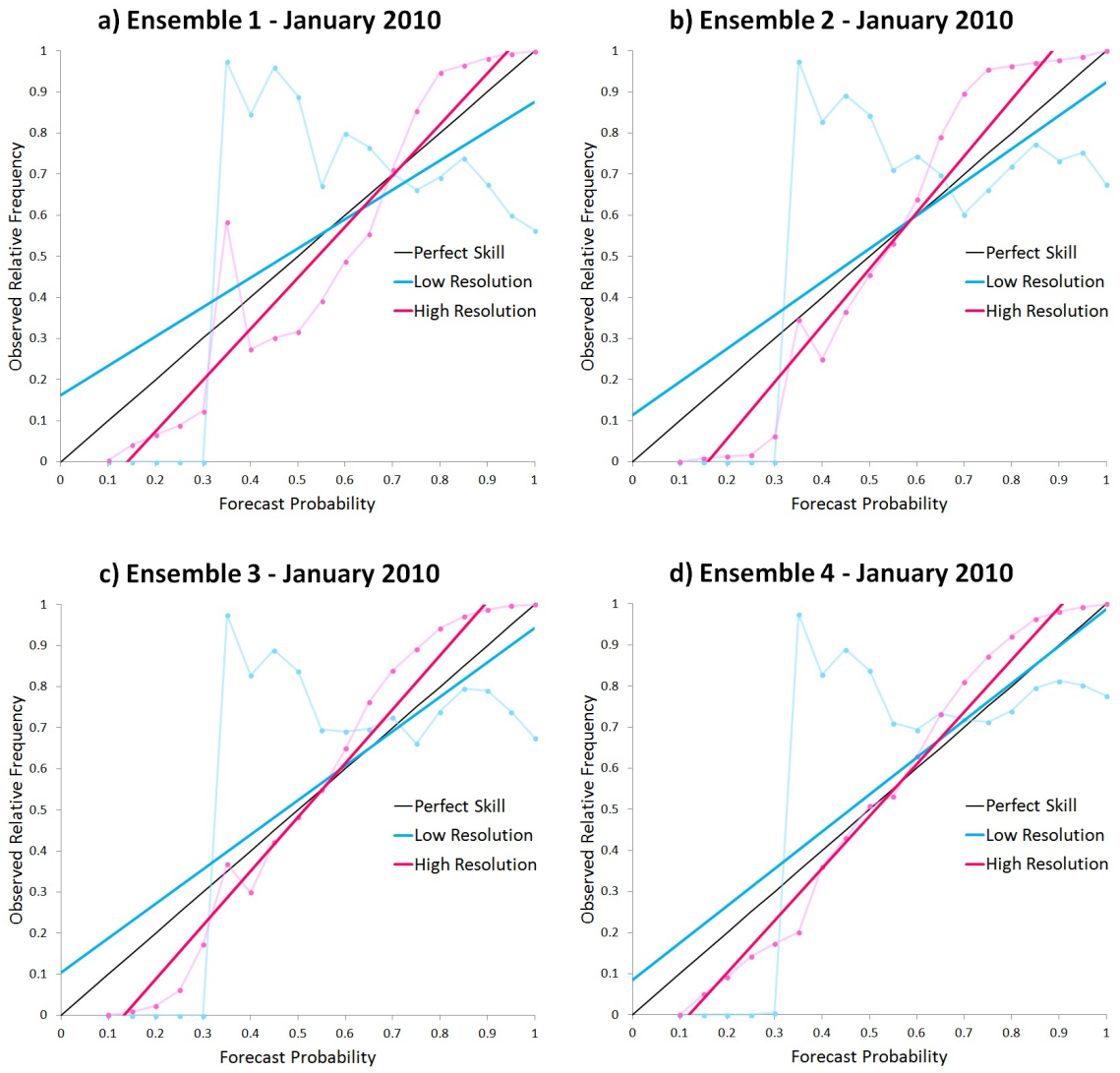

Figure 4.4.5: Reliability Diagrams for January 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

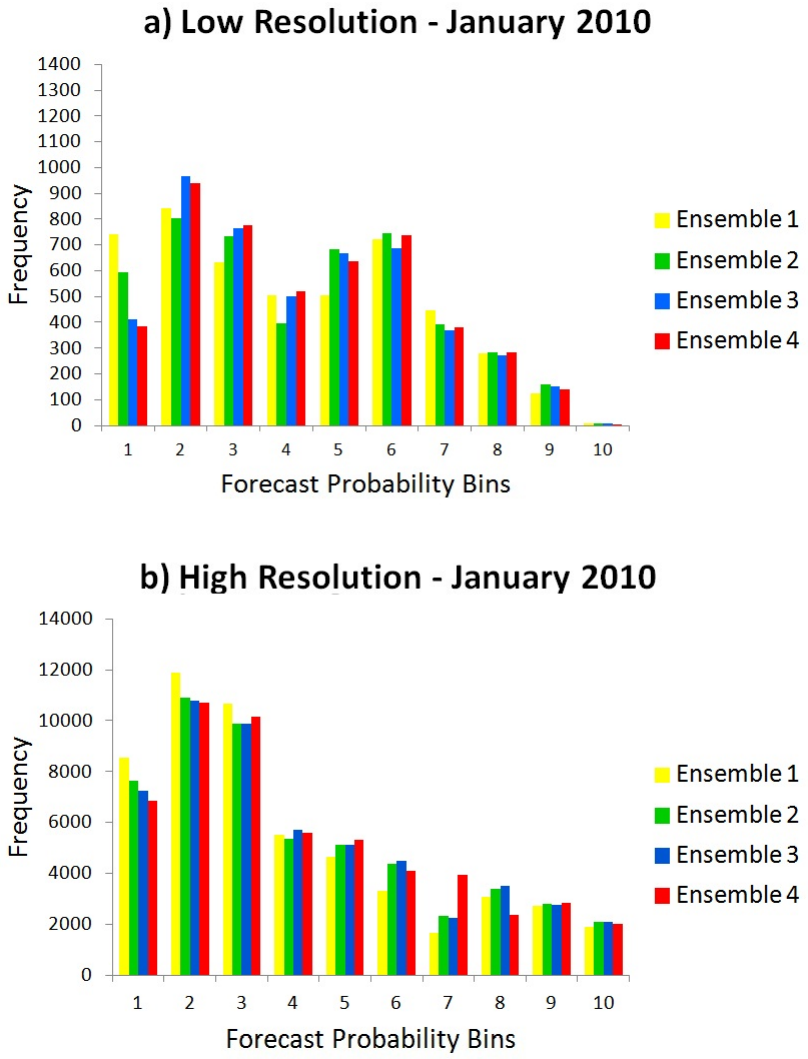

Figure 4.4.6: Sharpness Diagram for January 2010: a) Ensembles 1, 2, 3, and 4 at Low Resolution. b) Ensembles 1, 2, 3, and 4 at High Resolution.

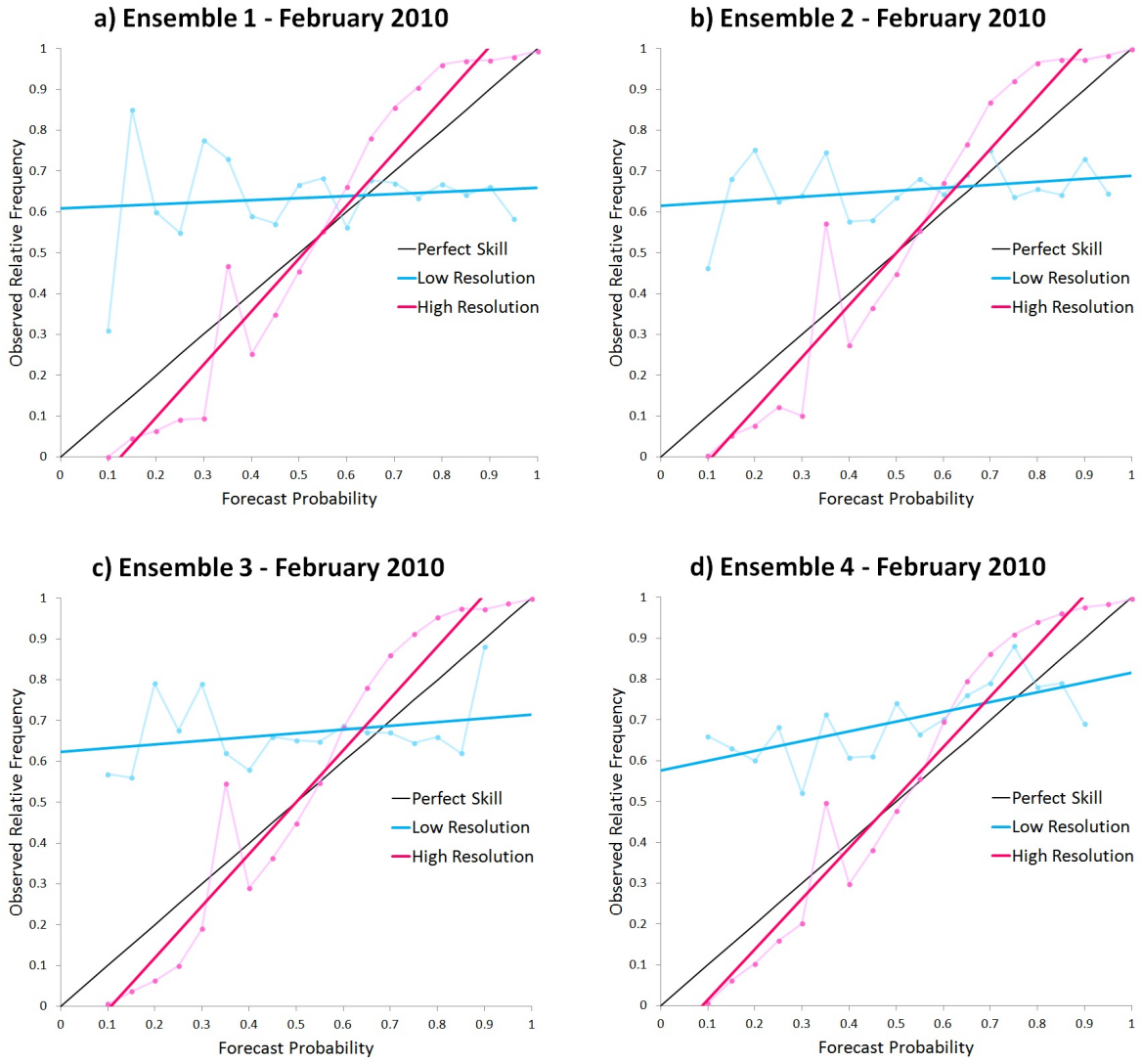

Figure 4.4.7: Reliability Diagrams for February 2010: a) Ensemble 1 at High (Pink) and Low (Blue) Resolution. b) Ensemble 2 at High and Low Resolution. c) Ensemble 3 at High and Low Resolution. d) Ensemble 4 at High and Low Resolution.

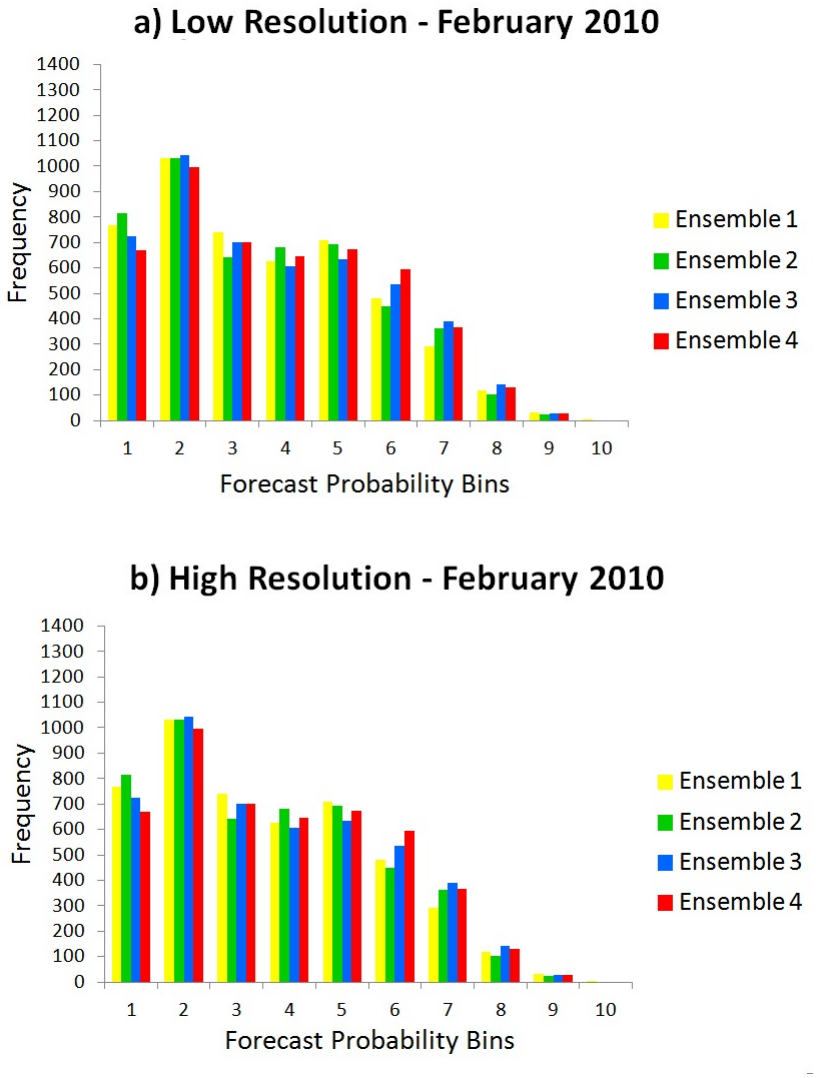

Figure 4.4.8: Sharpness Diagram for February 2010: a) Ensembles 1, 2, 3, and 4 at Low Resolution. b) Ensembles 1, 2, 3, and 4 at High Resolution.

4.5 Synopsis

The largest ensemble system (Ensemble 4) run at the high spatial resolution outscores in discrimination and reliability, and is more skilful than all the other systems. Although the skill scores for each month are similar for all eight EPSs, the largest ensembles at each resolution display greater skill than the others; and Ensemble 4 at high resolution is the most skilful in all but one case. The ROC diagrams show that all of the EPSs at high resolution far outscore those at the low resolution, and that the Ensemble 4 run at high resolution performs the best of all systems when discriminating both 1 mm and 5 mm of rainfall. Finally, the reliability was generally the best for the high resolution Ensemble 4 system, although the sharpness of all the EPSs was seen to be low.

This chapter demonstrates the increase in skill, reliability and discrimination with an increase in ensemble size, and particularly spatial resolution. Although in some cases the improvement in the these scores was slight, the general trend of improvement with an increase in these two attributes is clear.

Chapter 5

Conclusions

Short-range weather forecasting has evolved from single deterministic models to large ensemble systems producing probabilistic forecasts. The advantage of such large ensemble prediction systems is that they can produce a range of possible forecast outcomes from the initial conditions; and this helps compensate for errors that arise due to uncertainty in the initial conditions and the physics of the model. Deterministic forecasting is limiting in that it cannot reliably forecast at lead-times of more than one or two days. Alternatively, probabilistic ensemble forecasting systems can predict the weather further than two days ahead. In fact, in this dissertation skilfulness to seven days ahead were demonstrated. Therefore, it is more beneficial for weather forecasting to implement a probabilistic ensemble system than to have a single deterministic system.

In this study a probabilistic, short-range, single-model ensemble prediction system (EPS) has been developed and verified over southern Africa. This was achieved by configuring the CCAM such that four EPSs could be constructed and then evaluated in order to determine the most successful EPS. The CCAM has two different cumulus parametrisation schemes, and initial conditions are updated 6-hourly (four times per day). Therefore, the CCAM could be run eight times a day: four times a day for each cumulus parametrisation scheme. This resulted in eight possible ensemble members, and four different combina-

tions of these members made up the four EPSs that were studied. The EPSs ranged in size from smallest to largest with Ensemble 1, 2, 3, and 4 being two, four, six and eight members in size respectively. One hypothesis that was to be tested was whether an increase in ensemble size has an effect on the performance of the EPSs. In addition to this, another hypothesis that was to be tested was whether a change in the spatial resolution would have any effect on the performance of the EPSs. This was accomplished by performing the forecast simulation of each of the four EPSs at two different spatial resolutions: a low (50 km) and high (15 km) resolution. Ultimately, it was determined which ensemble prediction system produced the most successful forecasts.

The results of the hypotheses were arrived at by verifying all the EPSs with the corresponding observations. Then firstly comparing the performance of each ensemble to the other at the same resolution in order to assess the effect of size on performance. Secondly, comparing the performance of each ensemble to its equivalent at the different resolution so as to determine whether a change in spatial resolution had any effect on the performance of the EPSs. Finally, to conclude which EPS performed the best in all or most cases.

Deterministic and probabilistic verification scores were used to analyse the success of each EPS for the months of January and February, 2009 and 2010. The deterministic verification score that was used was the root mean square error (RMSE). The probabilistic verification scores that were used were the Brier skill score (BSS) and the relative operating characteristic (ROC). These scores assessed the accuracy, skill and discrimination of each EPS. Reliability diagrams were also used as a graphical representation of reliability of each EPS.

For the first hypothesis, the RMSE results show a very similar error curve for all four EPSs. However, the difference in error spatial maps display areas of error difference that illustrate the error decreasing with an increase in ensemble size. The skill of all four EPSs are very similar, however, the BSS curves show an improvement, albeit small, in skill with an increase in ensemble size. It can therefore be concluded that in order for an EPS to

begin demonstrating an improvement in skill, the EPS needs to be of a size much larger than the eight members considered here. The ROC graphs for the 1 mm discrimination of rainfall show the area under the curve increasing slightly as the ensemble size increases, establishing that the largest ensemble has a greater ability to discriminate 1 mm of rainfall. The ROC graphs for discriminating 5 mm of rainfall display a small increase in area with an increase in ensemble size, suggesting that a further increase in ensemble size would increase the ability of the model to discriminate 5 mm of rainfall. The reliability of the EPSs was tested with reliability diagrams, and these showed that in most cases there was a trend of increasing reliability with an increase in EPS size. Notwithstanding the small difference in skill, the general conclusion that can be drawn from the testing of this hypothesis is that the largest EPS performed best in all verification scores.

The effect of spatial resolution on the performance of the EPSs was tested for the second hypothesis. The RMSE graphs show that the errors in the high resolution forecasts are less than in the low resolution forecasts. This result is particularly pronounced up to the four-day lead time, where the high resolution EPSs have error far smaller than the EPSs at low resolution. The ROC curves for 1 mm exhibit a considerable improvement from the low and high resolution forecasts, with the high resolution forecasts showing an excellent ability to discriminate 1 mm of rainfall. The 5 mm threshold ROC curves show this same result, with the high resolution EPSs far out-performing the low resolution EPSs. The reliability diagrams display a general trend of increase in reliability with increase in spatial resolution, especially in the months of February 2009 and 2010, when the high resolution EPSs have substantially better reliability than the low resolution EPSs. The final result from this hypothesis is that the EPSs at high resolution outscored the low resolution EPSs in every case.

The two hypotheses above provide the results that the largest sized EPS performs the best when being assessed with the four chosen verification scores, and that the EPSs perform best at high resolution. It can therefore be concluded that the most successful EPS was found to be the largest ensemble: Ensemble 4, configured at a high spatial resolution.

In almost all cases, the largest ensemble at the highest resolution was found to outscore all the others, and increasing the horizontal resolution provided greater improvement of the forecasts than increasing the size of the ensemble prediction system.

The results from this work provide a great practical improvement in short-range numerical weather prediction. Implementing large ensemble and high resolution EPSs will result in a reduction in error, an increase in the ability to discriminate rainfall thresholds, and an increase in reliability. In order to improve these scores even further, future research should concentrate on creating a very large ensemble using perturbation techniques, and increasing the spatial resolution to a very fine grid. Other verification scores could also be studied, such as bias, in order see where the EPS could be further improved. A way to try to improve the configuration of the CCAM in the future would be to introduce different cumulus parametrisation schemes into the model and test whether they are more skilful at resolving small-scale cloud systems. A different way to proceed would be to include this EPS into a multi-model ensemble and contribute to a larger system of weather models.

This dissertation has contributed to the development of numerical weather prediction systems in southern Africa, since the conclusions drawn on the size of the forecast ensemble, and on the horizontal resolution required for improved weather prediction quality, have paved the way in which future numerical weather prediction systems will be developed in the region. Before, a small number of integrations at fairly coarse resolution were accepted as the standard, but now for numerical weather predictions to be improved in South Africa, new initialisation strategies and best practices to increase the model's resolution should become a more prominent focus for future development in the field.

References

Arribas A., Robertson KB, Mylne KR. 2005. Test of a Poor Man's Ensemble Prediction System for short-range probability forecasting. *Monthly Weather Review*, **133**, 1825–1839.

Atger, F., 1999: The skill of ensemble prediction systems. *Monthly Weather Review*, **127**, 1941-1953.

Ban, R., 2007: Completing the Forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts (2006). In *Sixth Communications Workshop.*

Batisani, N., 2011: The Spatio-temporal-severity dynamics of drought in Botswana. *Journal of Environmental Protection*, **2(6)**, 803-816.

Bhalotra, Y. P. R., 1987: Climate of Botswana. Part II: elements of climate. *Meteorological Services, MWTC, Gaborone, Botswana.*

Blamey, R.C., Reason, C.J.C., 2012: Mesoscale convective complexes over southern Africa. *Journal of Climate,* **25(2)**, 753-766.

Bowler, N. E., 2006: Explicitly accounting for observation error in categorical verification of forecasts. *Monthly Weather Review*, **13**4, 1600-1606.

Bowler, N.E., Arribas A., Mylne K.R., Robertson, K.B., Beare, S.E., 2008: The MOGREPS Short-range Ensemble Prediction System. *Q. J. R. Meteorol. Soc.,* **134**, 703–722.

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

Bradley, A. A., Schwartz, S. S., Hashino, T., 2008: Sampling Uncertainty and Confidence Intervals for the Brier Score and Brier Skill Score. *Weather and Forecasting.*, **23**, 992-1006.

Brancovic, C., Palmer, T. N., 1997: Atmospheric Seasonal Predictability and Estimates of Ensemble Size. *American Meteorological Society, Monthly Weather Review,* **125**, 859-874.

Brier, G. W., 1950: Verification Of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, **78**, 1–3.

Bright D.R., Mullen S.L., 2002. Short-range Ensemble Forecasts of Precipitation During the South-west Monsoon. *Weather and Forecasting*, **17**, 1080–1100.

Buizza, R., Petroliagis, T., Palmer, T.N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., Wedi, N., 1998. Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society,* **124(550)**, 1935-1960.

Buizza, R., Palmer, T.N., 1997. Impact of Ensemble Size on Ensemble Prediction. *American Meteorological Society*, **126**, 2503-2518.

Buizza, R., 1997: Potential Forecast Skill of Ensemble Prediction, and Spread and Skill Distributions of the ECMWF Ensemble Prediction System. *Monthly Weather Review*, **125**, 99–119.

Buizza, R., 2008: Comparison of a 51-Member Low-Resolution (TL399L62) Ensemble with a 6-Member High-Resolution (TL799L91) Lagged-Forecast Ensemble. *American Meteorological Society*, **136**, 3343-3362.

Buizza, R., Hollingworth, A., Lalaurette, F., Ghelli, A., 1998: Probabilistic Predictions of Precipitation Using the ECMWF Ensemble Prediction System. *American Meteorological Society*, **14**, 168-189.

Buizza, R., M. Miller, T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorological Society*, **125**, 2887– 2908.

Casati, B., Wilson, L.J., Stephenson, D.B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E.E., Brown, B.G., Mason, S., 2008: Review: Forecast Verification: Current Status and Future Directions. *Meteorological Applications*, **15**, 3-18.

Colle, B.A., Mass, C.F., 2000: The 5-9 February 1996 flooding event over the Pacific Northwest: Sensitivity studies and evaluation of the MM5 precipitation forecasts. *Monthly Weather Review*, **128(3)**, 593-617.

Collins, C.B., Hawkins, S.W., 1973: "Why is the Universe isotropic?". *The Astrophysical Journal*, **180**, 317–334.

Cook, C., Reason, C. J. C., Hewitson, B. C., 2004: Wet and Dry Spells Within Particularly Wet and Dry Summers in the South African Summer Rainfall Region. *Climate Research*, **26**, 17-31.

Cook, K.,H., 2001: A Southern Hemisphere response to ENSO with implications for southern Africa precipitation. *Journal of the Atmospheric Sciences*, **58(15)**, 2146-2162.

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

Davis, C. L., 2010: A climate change handbook for north-eastern South Africa.

Davis, C. L., 2011: Climate risk and vulnerability: A handbook for southern Africa. *Council for Scientific and Industrial Research, Pretoria, South Africa*, pp 92.

Du, J., SREF Development Team, 2007: Uncertainty and ensemble forecast. *Science and technology infusion lecture series, National Weather Service.*

Du, J., Mullen, S.L., 1997: Short-range ensemble forecasting of quantitative precipitation. *American Meteorological Society*, **125**, 2427-2459.

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.

European Centre for Mid-Range Weather Forecasts, 2012, *Sharpness,* accessed 11 December 2012, http://www.ecmwf.int/products/forecasts/guide/Sharpness.html.

Eckel F.A., Mass C.F., 2005. Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, **20**, 328–350.

Engelbrecht, F., (2005). Simulations of climate and climate change over southern and tropical Africa with the conformal-cubic atmospheric model. *Climate Change and Water Resources in southern Africa: Studies on scenarios, impacts, vulnerabilities and adaptation*, 57-74.

Engelbrecht, F. A., McGregor, J. L., & Rautenbach, C. D. W. (2007). On the development of a new nonhydrostatic atmospheric model in South Africa.

Engelbrecht, F.A., Landman, W.A., Engelbrecht, C.J., Landman, S., Bopape, M.M., Roux, B., McGregor, J.L., Thatcher, M., 2011: Multi-scale climate modelling over Southern Africa using a variable-resolution global model. *Water SA,* **37**. 647-658

Engelbrecht, C.J., Engelbrecht, F.A., Dyson, L.L., 2012: High-resolution model-projected changes in mid-tropospheric closed-lows and extreme Multi-scale climate modelling over Southern Africa using a variable-resolution global modelrainfall events over southern Africa. *International Journal of Climatology.*

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

Ferro, C.A.T., 2006: Comparing Probabilistic Forecasting Systems with the Brier Score. *Weather and Forecasting*, **22**, 1076-1088.

Fox-Rabinovitz, M., Côté, J., Dugas, B., Déqué, M., & McGregor, J. L., 2006: Variable Resolution General Circulation Models: Stretched-grid Model Intercomparison Project (SGMIP). *Journal Of*

Geophysical Research, **111**, D16104.

Fritsch, J.M., 1998: Quantitative Precipitation Forecasting: Report of the eighth prospectus development team, U.S. Weather Research Program, *Bull. American Meteorological Society*, **79**, 285-299.

Garstang, M., Tyson, P.D., Swap, R., Edwards, M., Kallberg, P., Lindesay, J.A., 1996: Horizontal and vertical transport of air over southern Africa. *Journal of Geophysical Research*, **101(D19)**, 23721-23723.

Grimit, E.P., 2004: *Probabilistic mesoscal forecast error prediction using short-range ensembles,* PhD Dissertation, University of Washington.

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

Hamill T.M,. Colucci S.J., 1997: Verification of eta–RSM short-range ensemble forecasts. *Monthly Weather Review*, **125**, 1312–1327.

Hamill T.M. 1997: Reliability diagrams for multicategory probabilistic forecasts. *Weather and Forecasting,* **12**, 736–741.

Hardiker, V., 1997: A Global Numerical Weather Prediction Model with Variable Resolution. *American Meteorological Society:* **125**. 59-73.

Harrison, M.S.J., 1984: A Generalised Classification of South African Rain-bearing Synoptic Systems, *Journal of Climatology*, **4**, 547-560.

Hastenrath, S., Greischar, L., Van Heerden, J., 1995: Prediction of the Summer Rainfall Over South Africa. *Journal of Climate*, **8**, 1511-1518.

Hoffman, R. N., & Kalnay, E. (1983). Lagged Average Forecasting, an Alternative to Monte Carlo Forecasting. *Tellus A*, **35**, 100-118.

Houtekamer, P. L., Derome, J., 1995: Methods for Ensemble Prediction. *Monthly Weather Review*, **123**, 2181-2196.

Houtekamer, P.L., Derome, J., 1994: Prediction experiments with two-member ensembles. *Monthly Weather Review,* **122**, 2179-2191.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* John Wiley & Sons.

Kalnay, E., 2003: *Atmospheric Modelling, Data Assimilation and Predictability.* Cambridge University Press.

Landman, S., Engelbrecht, F.A., Engelbrecht, C.J., Dyson, L.L., Landman, W.A., 2012: A short-range weather prediction system for South Africa based on a multi-model approach. *Water S.A.*.

Landman, W.A., Beraki, A., 2012: Multi-model forecast skill for midsummer rainfall over southern Africa. *International Journal of Climatology*, **32**, 303-314.

Lawrence, A.R., Hansen, J.A., 2007: A Transformed Lagged Ensemble Forecasting Technique for Increasing Ensemble Size. *Monthly Weather Review*, **135**, 1424-1438.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts (stochastic atmospheric processes). *Monthly Weather Review*, **102**, 409–418.

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

Lin, C., Vasic´, S., Kilambi, A., Turner, B., Zawadzki, I., 2005: Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophysical Research Letters*, **32**, L14801.

Liu, C., Moncrieff, 2007: Sensitivity of Cloud-Resolving Simulations of Warm-Season Convection to Cloud Microphysics Parametrizations. *American Meteorological Society*, **135**, 2854-2868.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences,* **20(2)**, 130-141.

Lu, C., Yuan, H., Schwartz, B.E., Benjamin, S.G., 2007: Short-Range Numerical Weather Prediction Using Time-Lagged Ensembles. *Weather and Forecasting*, **22**, 580-595.

Makarau, A., Jusy, M.R., 1997: Predictability of Zimbabwe Summer Rainfall. *International Journal of Climatology*, **17**, 1421-1432.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B.A., 2002: Does increasing horizontal resolution produce more skillful forecasts?. *Bulletin of American Meteorological Society*, **83**, 407–430.

McGregor, J.L., Dix, M.R., 2001: The CSIRO conformal-cubic atmospheric GCM. *IUTAM Symposium on Advantages in Mathematical Modelling of Atmosphere and Ocean* Dynamics, pp. 197-202, Springer Netherlands.

McGregor, J.L., Nguyen, K., 2010: An ensemble of high-resolution climate simulations for the MTSRF region. *Marine and Tropical Sciences Research Facility (MTSRF) Annual Conference.*

McGregor, J.L., Nguyen, K., Dix, M., Kowalczyk, E., Thatcher, M., 2006: Regional climate modelling using CCAM. *BMRC Workshop November 2006, CSIRO Marine and Atmospheric Research.*

McGregor, J.L., Dix, M., 2005: The Conformal-Cubic Atmospheric Model: progress and plans. *Workshop on High Resolution Atmospheric Simulations and Cooperative Output Data Analysis, Yokohama, Japan (JSPS International Meeting Series. Yokohama: JSPS and JAMSTEC. 2 p. Available: http://www. es. jamstec. go. jp/esc/research/AtmOcn/hires2005/abstract/4 2_ mcgregor. pdf.*

McGregor, J.L., Dix, M.R., 2008: An Updated description of the conformal-cubic atmospheric model. *High Resolution Numerical Modelling of the Atmosphere and Ocean*, pp. 51-75, Springer New York.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B.A.,2002: Does increasing horizontal resolution produce more skillful forecasts?. *Bulletin of American Meteorological Society*, **83**, 407–430.

McGregor, J. L., 2003. A new convection scheme using a simple closure. In *Current issues in the parameterization of convection': Extended abstracts of presentations at the fifteenth annual BMRC modelling workshop* (pp. 13-16).

Mullen, S. L., and R. Buizza, 2002: The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Weather Forecasting*, **17**, 173–191.

Nicholson, S.E., Some, B., McCollum, J., Nelkin, E., Klotter, D.,Berte, Y., Traore, A.K., 2003: Validation of TRMM and other rainfall estimates with a high-density guage dataset for West Africa. Part II: Validation of TRMM rainfall products. *Journal of Applied Meteorology,* **42(10)**, 1355-1368.

Olson, D.A., Junker, N.W., Korty, B., 1995: Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Weather and Forecasting*, **10(3)**, 498-511.

Palmer, T. N., Hardaker, P. J., 2012: Handling uncertainty in science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **369**: 4681-4684.

Potgieter, C.J., 2007: Short-range Weather Forecasting Over Southern Africa with the Conformal-Cubic Atmospheric Model. *MSc dissertation, University of Pretoria, South Africa.*

Preston-Whyte, R. A., Tyson, P. D., 1988. The atmosphere and weather of southern Africa. Cape Town: Oxford University Press.

Rautenbach, C. D. W., Engelbrecht, F. A., Ndarana, T., Engelbrecht, C. J., McGregor, J. L. (2005). Regional Model Development for Simulating Atmospheric Behaviour and Rainfall Over Southern Africa: Final Report.

Reason, C.J.C., Allan, R., Lindesay, J.A., Ansell, T.J., 2000: ENSO and climatic signals across the Indian Ocean basin in the global context: Part I, Interannual composite patterns. *International Journal of Climatology,* **20(11)**, 1285-1327.

Reason, C. J. C., Keibel, A. (2004). Tropical cyclone Eline and its unusual penetration and impacts over the southern African mainland. *Weather And Forecasting*, **19**, 789-805.

Reason, C. J., Engelbrecht, F., Landman, W. A., Lutjeharms, J. R. E., Piketh, S., Rautenbach, C. J., Hewitson, B. C. (2006). A review of South African research in atmospheric science and physical oceanography during 2000-2005. *South African Journal Of Science*, **102**, 35.

Riphagen, H.A., 1985: *The implementation of a splitexplicit weather prediction model for the Southern Hemisphere*, Navorsingsinst.

Riphagen, H.A., Burger, A.P., 1986: Comments on the computational stability of Gadd's adjustment and advection schemes for a split explicit model. *Quarterly Journal of the Royal Meteorological Society*, **112(471)**, 276-282.

Riphagen, H.A., de Coning, E., van Hemert, L.R., Tennant, W.J., Gertenbach, J.D., Hunter, I.T.,1996: Experience with an upgraded Eta assimilation and prediction system at the South African Weather Bureau. *Preprints, 11th Conf. on Numerical Weather Prediction, Norfolk, VA, Americal Meteorological Society*, 88-90.

Riphagen, H.A., Bruyere, C.L., Jordaan, W., Poolman, E.R., Gertenbach, J.D., 2002: Experiments with the NCEP regional Eta Model at the South African Weather Bureau, with emphasis on terrain representation and its effect on precipitation predictions. *Monthly Weather Review*, **130(5)**, 1246-1263.

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. *Q. J. R. Meteorological Society*, **123**, 1227-1282.

Schultz, P., 1995: An explicit cloud physics parametrization for operational numerical weather prediction. *Monthly Weather Review,* **123(11)**, 3331-3343.

Stanski, H. R., Wilson, L. J., Burrows, W. R., 1989: Survey of common verification methods in meteorology. *World Meteorological Organisation.*

Stensrud, D.J., Brooks, H. E., Du, J., Tracton, M. S., Rogers, E., 1999: Using ensembles for short-range forecasting. *Monthly Weather Rev*iew, **127**, 433–446.

Stensrud, D.J., Bao, J.W., Warner, T. T., 1998: Ensemble forecasting of mesoscale convective systems.

In *Preprints, 12th Conf. on Numerical Weather Prediction, Phoenix, AZ, Amer. Meteor. Soc* (pp. 265-268).

Stensrud, D.J., Yussouf, N., 2007: Reliable Probabilistic Quantitative Precipitation Forecasts from a Short-Range Ensemble Forecasting System. *Weather and Forecasting*, **22**, 3-17.

Tennant, W.J., Toth, Z., Rae, K.J., 2006: Application of the NCEP Ensemble Prediction System to Medium-Range Forecasting in South Africa: New Products, Benefits, and Challenges. *Weather and Forecasting*, **22**, 18-35.

Thatcher, M., McGregor, J. L., 2009: Using a scale-selective filter for dynamical downscaling with the Conformal-Cubic Atmospheric Model. *Monthly Weather Review*, **137**, 1742-1752.

Thatcher, M., & McGregor, J. L. (2011). A technique for dynamically downscaling daily-averaged GCM datasets using the conformal cubic atmospheric model. *Monthly Weather Review*, **139**, 79-95.

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

Todd, M., Washington, R. (1999). Circulation anomalies associated with tropical-temperate troughs in southern Africa and the south west Indian Ocean. *Climate Dynamics*, **15**, 937-951.

Todd, M. C., Washington, R., Palmer, P. I. (2004). Water vapour transport associated with tropical–temperate trough systems over southern Africa and the southwest Indian Ocean. *International journal of climatology*, **24**, 555-568.

Toth Z, Zhu Y, Marchok T., 2001: On the economic value of ensemble based weather forecasts. *Bulletin of American Meteorological Society,* **83**, 73–83.

Trenberth, K.E., Dai, A., Rasmussen, R.M., Parsons, D.B., 2003: The changing character of precipitation. *Monthly Weather Review,* **84(9)**, 1205-1217.

Triegaardt, D.O., 1965: The application of dynamic methods of weather prediction to the South African forecasting problem: a progress report. *South African Weather Bureau Newsletter,* **201**, 187-188.

Tyson P.D., Preston-Whyte R.A., 2000. The Weather and Climate of Southern Africa. Oxford University Press: Cape Town, South Africa.

Vitart, F., Anderson, D., Stockdale, T., 2003: Seasonal forecasting of tropical cyclone landfall over Mozambique. *Journal of Climate*, **16**, 3932-3945.

Wandishin, M. S., Mullen, S. L., Stensrud, D. J., and Brooks, H. E., 2001: Evaluation of a short-range multimodel ensemble system. *Monthly Weather Review,* **129**, 729–747.

Weisman, M.L., Skamarock, W.C., Klemp, J.B., 1997: The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review,* **125(4)**, 527-548.

Wiin-Nielsen, A., 1991: The birth of numerical weather prediction. *Tellus A,* **43AB**, 36-52.

Wilks DS. 2006. Statistical Methods in the Atmospheric Sciences, 2nd Edition. Elsevier Academic Press, California, United States of America: 627 pp.

World Meteorological Organisation, 2012. Guidelines on Ensemble Prediction Systems and Forecasting, WMO-No. 1091. Geneva, Switzerland: 29 pp.