

Application of the NCEP Ensemble Prediction System to Medium-range
Forecasting in South Africa: New Products, Benefits and Challenges

Warren J. Tennant^{*}, Zoltan Toth[#] and Kevin J Rae^{*}

^{*}South African Weather Service

[#]National Centers for Environmental Prediction

Submitted to Weather and Forecasting: December 20, 2005

Revised: June 9, 2006

Corresponding Author Address:

South African Weather Service

Private Bag X097

Pretoria, 0001

SOUTH AFRICA

email: tennant@weathersa.co.za

Tel: +27-12-367-6004

Fax: +27-12-367-6189

Abstract

The National Centers for Environmental Prediction (NCEP) Ensemble Forecasting System (EFS) is used operationally in South Africa for medium-range forecasts up to 14-days ahead. The use of model-generated probability forecasts has a clear benefit in the skill of the 1-7 day forecasts. This is seen in the forecast probability distribution being more successful in spanning the observed space than a single deterministic forecast, and thus substantially reducing the instances of missed events in the forecast. In addition the probability forecasts generated using the EFS are particularly useful in estimating confidence in forecasts. During the second week of the forecast the EFS is used as a heads-up for possible synoptic-scale events and also for predicting average weather conditions and probability density distributions of some elements such as maximum temperature and wind.

This paper assesses the medium-range forecast process and the application of the NCEP EFS at the South African Weather Service. It includes a description of the various medium-range products, adaptive bias-correction methods applied to the forecasts, verification of the forecast products and a discussion on the various challenges that face the researchers and forecasters alike.

1. Introduction

Weather forecasts have potential use at a variety of space and time-scales. As a public weather forecast service, the South African Weather Service (SAWS) is tasked to provide a comprehensive forecast service from a few hours ahead through all scales up to several seasons ahead. The medium-range (three to 14 days) is particularly popular through a number of sectors and thus considerable effort has been invested in improving forecasts for this time-scale. To this end the National Center for Environmental Prediction (NCEP) Ensemble Forecasting System (EFS) (Toth and Kalnay 1997; Toth et al. 2001; Buizza et al. 2005) is used operationally in South Africa for medium-range forecasts up to 14-days ahead.

Ensemble methods (Leith 1974) are considered to be an effective way to estimate the probability density function of future states of the atmosphere by addressing uncertainties present in initial conditions and in model approximations. Notwithstanding, biases remain in these forecast distributions, especially in user-orientated fields such as rainfall (e.g. Hamill and Colucci 1998) and surface temperature (e.g. Hamill et al. 2004). Various bias correction methods and verification statistics are described in the literature. However, it is important to get a grasp of the practical implementation of a forecast guidance system in a regional setting and to establish the strengths and weaknesses of the EFS in these regions.

The objective of this paper is to introduce some novel medium-range forecast products that have been generated from the NCEP EFS, including bias-correction, and assess the success of these in an operational environment at the South African Weather Service.

2. Ensemble Forecast System and Verification Methodology

a. NCEP Ensemble Forecast System

The SAWS has been downloading subsets of the NCEP EFS on a daily basis since 2000. The operational configuration at NCEP used for this study has been in effect since May 2000 and consists of 23 ensemble members per day out to 16 days ahead. At 00Z the suite consists of a high-resolution (T170L42, T254L64 since April 2003) control run up to 7-days, truncated to T62L28 for the remaining 9 days, a low-resolution (T62L28) control run, plus 5 pairs of perturbed integrations derived from the breeding cycle (Toth and Kalnay 1993; 1997). At 12Z the high-resolution control run extends only to 3.5 days, and there are another 5 pairs of independently bred perturbations. The full ensemble set thus consists of 23 members per day. Since March 2004 the ensembles have been available four times daily at a T126L28 resolution (more detail on the NCEP EFS is available online at <http://wwwt.emc.ncep.noaa.gov/gmb/ens/index.html>). Owing to bandwidth constraints the SAWS has, up to writing of this paper, only downloaded the coarser 2.5° resolution datasets. Although the model output assessed here is at this lower resolution, the model performance has been continuously improving through upgraded model physics, resolution and data assimilation and these effects are automatically manifest in the coarser output sets.

b. Verification Methods

Atmospheric variables such as pressure-level geopotential heights and sea-level pressure are compared against the ensemble control run analysis. These continuous data are verified using root-mean-square error and anomaly correlation coefficient scores. In order to calculate the anomalies used for the correlation calculations, monthly climatological fields, derived from NCEP/NCAR reanalysis data (Kalnay et al. 1996), are converted to daily values by performing a linear interpolation between the two nearest monthly values, and are subtracted from the forecast fields.

Surface variables such as rainfall and maximum/minimum temperature are verified against station data from the SAWS database. There are 96 temperature stations across South Africa and between 1500 and 2000 rainfall stations (Fig. 1). At the 2.5° EFS forecast resolution, roughly 30 to 200 rainfall stations fall into each model box, with the lower values corresponding to the sparsely populated arid areas in the western interior of South Africa. This verification is done using two approaches. The first consists of model forecast values of temperature and rainfall probabilities being interpolated to the 96 forecast station locations and the second (used mainly in rainfall verification) of a value constructed for each model box using an average of the stations within the box. The EFS is designed for box average probabilistic quantitative precipitation forecasts (PQPF) and not point values, and is therefore not expected to perform well at individual points. A further caveat to note here is that convective-type rainfall in the summer rainfall areas of South Africa has a particularly high spatial variability and station measurements are

sometimes not representative of the area – in this case the 2.5° model grid box. This also highlights the reverse problem of deriving a point forecast of rainfall from model grid boxes and will be addressed in future research. In a study over Australia (which has similar conditions to South Africa), Ebert et al. (2003) suggest that the difference between forecast and observed rainfall fields are much greater than errors in verification data (from representativeness), suggesting that we may proceed to verify PQPF with caution. This verification is done using the Equitable Threat Score (ETS) (Ebert 2001) that has the advantage of measuring the fraction of observed and/or forecast events that were correctly predicted but adjusting the score for hits associated with random chance.

The skill of probabilistic forecasts is verified using the Brier Skill Score (BSS) (Brier 1950; Wilks 1995). This skill score measures the squared probability error relative to climatology. Murphy (1973) decomposed the Brier Score into reliability (agreement between forecast probability and observed frequency), resolution (ability of the forecast probabilities to distinguish between events and non-events) and uncertainty (depends on climatology and has a maximum value of 0.25 when the observed frequency is 50%). A further measure of the ability of forecasts to distinguish between events and non-events is shown using the Relative Operating Characteristic (ROC) (Mason 1982). The BSS looks at performance stratified by forecast probabilities and the ROC performance based on the observations. The ability of the spread of the EFS to represent the variability of the real atmosphere is measured using the Rank Histogram (Talagrand et al. 1997; Hamill 2001). This measure is particularly useful in determining whether the EFS has errors in its mean and spread. Again, as pointed out by Hamill (2001), errors in observations (mostly

through representativeness) may introduce false signals in this histogram. More on this will follow in the discussion.

c. Forecast Products

The products from the EFS can be divided into two groups. The first is the set of products provided to the forecasters that is used as guidance in compiling their medium-range forecast. The second group consists of computer-generated products that are disseminated to the public and/or specialized users. These are each described separately below.

- 1) *Forecaster Guidance:* Part of the success in introducing the EFS to the forecast offices in the SAWS has been the development of a user-friendly forecast display system. This is based on HTML WebPages and consists of a homepage in the form of a table with the vertical axis representing the forecast days from 1 to 14. The horizontal axis contains different forecast parameters. Each panel consists of a thumbnail-size spatial map that provides a quick preview of the expected weather in the medium-range. The maps include a date with forecast day, PQPF for 1mm and 8.5mm over the 24-hour period, probability of 24-hour maximum and minimum temperature change (from the previous day) exceeding +2 and -2 degrees Celsius, contours of pressure-level geopotential heights of the high-resolution control run overlying a shaded field that indicates expected forecast uncertainty (based on ensemble spread) and the probability of the 850-500hPa thickness falling below 4200 and 4100gpm. Temperature changes over 24-hours are preferred to actual values because surface air temperature varies significantly

from one location to another (often within a model grid box) based on altitude, terrain and proximity to bodies of water. Each thumbnail image on the webpage table can be expanded to full size through a mouse click and navigation around the expanded images in the table is done using arrow links. The expanded pages also provide links to additional detail such as additional thresholds, bias-corrected fields and spaghetti diagrams. The webpage includes a built-in automatic update instruction to ensure that forecasters are always viewing the latest forecast.

- 2) *Public Forecast Products:* While forecasters have a good understanding of the use of numerical weather prediction models, the public can easily misinterpret such computer-generated information. Therefore, the products issued to the public need to be carefully designed, adequately documented and limited to a few understandable parameters (Ryan 2003). The SAWS issues deterministic station-specific forecasts for the first 7 days, as has been the practice for many years before the introduction of the EFS, but this does include a probabilistic forecast of precipitation (available online at <http://www.weathersa.co.za>). For the second week computer-generated products of probabilities of maximum and minimum temperature categories and wind roses (Fig. 2) are issued for the 96 temperature-forecast stations shown in figure 1. These graphics include climatological distributions of the forecast parameter to indicate the expected departure from the mean. The advantage of these products is that they make full use of all of the ensemble information. Furthermore, the products span a week, so the impacts of errors in the individual ensemble member forecasts, regarding the timing of specific events during the second week, are reduced.

3. Verification of Medium-Range Forecast Products

a. Continuous variables

This section focuses predominantly on the 00Z 500hPa height forecasts for the period January 2001 to December 2004 and over a domain covering the South Atlantic Ocean, southern Africa and the southwestern Indian Ocean. Results for sea-level pressure concur on the whole with the 500hPa heights and are thus not repeated in the results. The ensemble average (10 perturbation members) has a lower RMSE than the low-resolution (high-resolution) control run after five (six) days (Fig. 3). However, the ensemble average RMSE exceeds the RMSE obtained by using climatology (the limit of skill) after day eight, only one day after the control runs reach this threshold. This shows that the benefit of the ensemble technique over the control run to make skillful forecasts of the instantaneous 500hPa fields is realized on average for only one day in this region. Notwithstanding, when taking the best ensemble member at each forecast case (identified a posteriori), the RMSE is smaller than that of climatology up to day twelve. Bourke et al. (2004) also note the advantage of using the Bureau of Meteorology EFS, with a 33% relative gain of the best ensemble member to the high-resolution control over Australia at day 5. A tally of the ensemble members shows that the high-resolution control run is the most accurate more often than any of the other ensemble members for the first three days of the forecast. The low-resolution control is usually most accurate about half of the amount of time that the high-resolution control is most accurate, illustrating the level of improvement in skill that can be realized by increasing model resolution. From forecast

day four, one of the perturbed ensemble members is usually the most accurate. By the end of the forecast period (day 16) any ensemble member has an equal chance of being the best. In essence the ensemble technique is successful in producing at least one forecast member that can be used to make a skillful forecast for this region up to twelve days ahead. This suggests that the breeding method is able to capture the uncertainty in the initial conditions such that the probability distribution of the ensemble forecast envelope does cover the observed events out to this lead-time.

As already mentioned, upgrades to the GFS model at NCEP have led to improvements in the model performance. This is evident in the ensemble average, best member and control runs of the EFS. Generally, skill of the best member has improved by almost 2 days to 14 days and the bias has been reduced by 60% when comparing the scores for 2001 against those of 2004 (not shown). It is also worth noting that the ensemble mean is slightly worse than the control run of the same resolution, for the first few days (also noted in Szunyogh and Toth 2002). This is because the ensemble suite is generated using the breeding method (Toth and Kalnay 1993; 1997) that adds perturbations to the control analysis, in the form of stochastic noise, with the aim of reducing model systematic error in the ensuing forecasts. The uncertainty estimate mask that determines the size of these perturbations has only recently been upgraded at NCEP and was known to produce inflated perturbations for the southern hemisphere. This would partially explain the problem with the ensemble mean. Additionally, the perturbed ensemble member forecasts would be initially disadvantaged relative to the control forecasts if verified against the control analysis. However, the ensemble generating technique is designed to capture the

uncertainty in the initial conditions and hopefully provide a more useful forecast probability distribution. The benefit of this becomes clearer later in this paper, particularly with the 850-500hPa thickness probability forecasts.

The anomaly correlation coefficient (ACC) scores are similar to the RMSE scores (Fig 4). One notable difference is that over the smaller domain (37.5-17.5°S and 10-40°E) the ensemble average ACC scores remain poorer than the high-resolution control throughout the forecast period, whereas the ensemble average beats the high-resolution control from day 6 in the large domain. Furthermore, the score of the best ensemble member at each case is higher over the small domain, than the large domain. These indicate that locally the correct type of weather system is being simulated by the model by (at least) some members and the poorer score for the ensemble average could be caused by the smoothing effect of combining forecasts of the same weather system, but located at different positions.

A useful forecast system must be able to capture as much of the range of natural variability as possible. One way to measure this is to plot the standard deviation of the ensemble member fields from their own time-averaged value for the full verification period at each forecast lead-time (Fig. 5). During the first week of the forecast the variance decreases, but then begins to increase during the second week. The control runs have a lower variance than that observed, with the low-resolution control considerably lower than the high-resolution control. This is consistent with the constrained atmospheric variability, in this region in particular, caused by the finite resolution of

models (Stratton 1999; Tennant 2003). The perturbation breeding method introduces additional variability and results in the first three days of the perturbed ensemble members having a higher variance than observed. The ensemble average obviously underestimates the variance significantly, making this field unsuitable for event forecasting in this region despite its apparent advantage over individual ensemble members in terms of RMSE and ACC between forecast days five and eight.

b. Bias Correction

The most basic way to correct model systematic biases is by subtracting the long-term mean error of the forecasts (Richardson 2001). This is usually done independently for each grid point and forecast lead-time. However, this omits two other rather important dependencies, namely those related to the season and to the circulation regime. To address this issue Atger (2003) proposed a spatially and temporally dependent bias correction. In order to introduce this sort of bias correction in an operational environment Eckel and Mass (2005) adopted a 14-day running mean bias calculation. Cui et al. (2005) also discuss these methods. This study follows such a bias-correction technique with the following additions. The running mean was tested for 30, 14 and 7 days, and the 14-day running mean was found to be optimal. Bias correction was done independently for each control ensemble member but as a group for the perturbed ensemble members. This was necessitated by the different resolution (and hence bias) of the control runs (Szunyogh and Toth 2002) and possible differences introduced by the initial perturbations on the perturbed members. There is no reason to expect the bias of any particular perturbed

ensemble member to be different from the others, so the same average bias of all the perturbed members was subtracted from each member.

The bias correction procedure was performed as follows. Starting at 1 January 2001, a bias for each forecast lead-time for forecasts valid for the 14-day window period (ending 31 December 2000) was calculated. This was done by calculating the mean difference between the forecasts and the observed fields multiplied by a factor *alpha* (estimated empirically at 0.33, i.e. 33% of the forecast error could be attributed to the forecast bias). All the forecasts valid over the window period were corrected by subtracting the bias factor. The process was then repeated for 2 January 2001 (with the 14-day window extending from 17 December 2000 to 1 January 2001) and each day in turn until 1 January 2005. During this process each particular forecast case was bias-corrected 14 times as the window moved across the time. This iterative process provided more stability to the bias correction process. The latest forecast (simulating an operational environment) was bias-corrected using the mean difference between the standard forecasts and the latest set of iteratively bias-corrected forecasts over the 14-day window. In this way the most up-to-date bias information was used to correct the current forecast. These forecasts (with only one bias-correction step) were saved separately and used in the verification process. The main advantage of this bias-correction method is that the bias can be calculated from a relatively short period and thus be implemented easily in an operational environment.

Over the southern Africa domain the GFS model exhibits an increasing area-average negative bias with forecast lead-time (Fig. 6). The magnitude of the bias is dependent on model resolution, with a larger bias associated with the lower resolution. As expected from results of other studies (Atger 2003; Cui et al. 2005) the bias correction method is successful in reducing the magnitude of the bias considerably throughout the forecast period. Talagrand diagrams also confirm this bias reduction with a more even distribution (not shown). The forecast bias cannot be totally removed using these sort of correction methods in a real-time forecasting sense, as it is not possible to fully anticipate the systematic-error component of future forecast errors based on past errors. Furthermore, efficient bias-correction does not always necessarily lead to an improvement in forecast skill, but the method discussed here does also improve skill somewhat (Fig. 4). This improvement is also evident in the increase in variability after bias-correction (Fig. 5). This probably occurs when the adaptive bias-correction shifts the model forecast away from the model climate (with constrained variability) towards reality with more variability.

It is interesting that the ensemble perturbation members have a smaller bias than the low-resolution control run at forecast days two and three (shown by the ensemble average curve in figure 6). The only differences between these runs are the perturbations added to the initial conditions, suggesting that this bias is influenced by the initial conditions. Buizza et al. (2005) state that a successful EFS should capture the effect of both initial condition and model uncertainties on forecast errors. To investigate this, the spatial variation of the difference between the perturbed ensemble members and the control run

is shown in figure 7. Here we see that the ensemble breeding method has an overall spatial pattern where the perturbed ensemble members (on average at time zero) tend to have higher values over the land areas and South Atlantic anticyclone, and negative in the mid-latitude westerlies for both the 500hPa height and sea level pressure fields relative to the control analysis. These are clearly very small values but certainly spatially coherent. Although the cause of this is not clear, some further investigation revealed that these patterns seem to be related to the interpolation of the high-resolution analysis (from NCEP's global data assimilation system) to create the analysis for the lower model resolution of the ensemble set. This lower-resolution analysis is used for the breeding of the ensemble perturbations. After 24 hours these same patterns amplify to amounts that neatly offset the traditional model bias in the region of a negative bias in the subtropics and a positive bias in the mid-latitudes of the region. Such a pattern where the southern hemisphere jet stream is displaced toward the equator by many GCMs is familiar in the region (Tennant 2003). Furthermore, the bias correction method, although successful in reducing the bias spatially, does leave some of the spatial pattern of the bias behind (Fig. 8). These highlight the need for regime-dependent bias correction, since model performance can be linked to correctly simulating circulation regimes (Chessa and Lalaurette 2001).

c. PoP and Temperature Change Forecasts

A clear bias in the EFS is evident over southern Africa in terms of inflated quantitative precipitation probabilities. The root of this problem appears to be that the NCEP GFS model overestimates rainfall amounts, climatologically speaking, by up to 300% over the

summer rainfall areas of South Africa, especially along the eastern escarpment at 30°E (Fig. 9). Surprisingly, there is little improvement in the high-resolution forecast over the low-resolution forecast, as usually such biases tend to be related to model resolution. Of further note, is that this bias becomes greater for the higher rainfall amounts greater than 5mm, especially for the longer forecast lead-times (Fig. 10). The bias score used here, calculated as the number of forecast cases divided by the number of observed cases, is thus very large (up to 50), especially given the small number of observed cases of >20mm over the 2.5°x2.5° model grid-box. This would partially explain why the high-resolution model has such a strong bias because large rainfall amounts tend to be more easily generated by higher-resolution models (Mullen and Buizza 2002). Furthermore, forecast lead-time appears to have little bearing on the magnitude of the bias (except high-resolution control at 5 days), suggesting a problem with model physics (possibly the precipitation parameterization schemes) over this region. In contrast to the summer rainfall areas, rainfall is underestimated in the winter rainfall region of the southwestern Cape (Fig. 9 and 11). In this region we have large-scale rain-bearing frontal systems that are enhanced by local topography, a feature probably not adequately captured by the current resolution of the NCEP global model.

Talagrand diagrams (Hamill 2001) confirm the findings above (Fig. 12). The summer rainfall region has a clear bias of forecasting too much rain too often (left-skewed diagram), where the observation is less than the driest ensemble member a third of the time. The winter rainfall region on the other hand has a U-shaped diagram indicating a

lack of variability in this region, although the diagram does exhibit a slight left-sided bias. Again these patterns do not change significantly with forecast lead-time.

Quantitative precipitation forecast (QPF) fields were calibrated by adjusting the event threshold (i.e. the precipitation forecast amount defining a “yes” or “no” forecast event), so that the forecast frequency matched the observed frequency for the verification period (July 2000 to June 2005). For this study, a cross-validation technique was used where each year was withheld from the calculations while the other four years were used to calculate the forecast frequencies. Roebber et al. (2004) suggest that an increase in model resolution, post-processing of model data and combining high resolution with ensemble techniques are practical ways to improve rainfall prediction. These approaches concur with the findings in this study in the following way.

Over the summer rainfall area, Equitable Threat Scores (ETS) of the calibrated high-resolution control forecasts are improved for light rainfall amounts (Fig. 10). Over the winter rainfall area the calibration of the high-resolution control was not as successful, except for five-day forecasts of larger rainfall amounts over the southwestern Cape (Fig. 11). This is probably because the model does not capture the orographic augmentation of rainfall in this region adequately, resulting in a largely systematic bias that can easily be corrected. It is noteworthy that the equitable threat score for highly simplified ensemble probability forecasts (assuming a deterministic “yes” forecast when the ensemble probability exceeds 50%) generally beats the control run scores for rainfall amounts less than 20mm, more so at the longer lead times (Fig. 10). This demonstrates the utility of

the EFS to do QPFs for light rainfall events. Unfortunately, as found by Legg and Mylne (2004), the calibration of the ensemble QPF, particularly in the summer rainfall region, causes a significant deterioration of the ETS score for heavier rainfall events by reducing the probabilities too far as seen from the negative bias in figure 10. Perhaps this requires a different interpretation of the probabilities for the more extreme events, as these would hardly ever exceed 50% in practice. The negative bias here is dependent on a deterministic decision of what probability should be considered to be a “yes” forecast.

Equivalent results were obtained from frequency adjustment of the ensemble PQPF values. An increase in the BSS and a marginal increase in the ROC were evident for the first few days of light rain (< 2mm) forecasts over the summer rainfall area, but for heavier events (>10mm) the ROC score deteriorated. Over the winter rainfall region, fairly good improvements were made to the BSS and ROC score during the first week. This is attributed to the successful correction of the systematic bias in model-simulated rainfall in this region as mentioned above.

Probability forecasts of temperature changes of 2°C and 5°C are skillful for week 1 over most of South Africa and for week 2 over parts of the interior (not shown). Coastal temperatures are probably not resolved sufficiently by the model resolution and are less skillful than those over the interior. Maximum temperatures are more skillful than minimum temperatures, pointing again to lack of resolution of sub-grid scale processes at night.

d. Forecaster Guidance – probability of events

A useful EFS product for event forecasting is the probability of the 850-500hPa-thickness field falling below 4200 and 4100gpm. These synoptic situations are good indicators of extreme cold weather and possible snowfall in South Africa, both of which are considered high-impact events in the region. Over the northern parts of the country the thickness fields almost never fall below 4100gpm, but do fall below 4200 around ten times per winter season. Given the high altitude of the escarpment (2000-3000m) and interior plateau (1500m), surface temperatures are usually below freezing over night over large areas. In the southern parts of the country thickness fields below 4200gpm occur almost half the time (corresponding to a maximum uncertainty value in the Brier Score decomposition), but occurrences of values less than 4100gpm match that of 4200gpm over the northern parts. The results discussed below for the 4100gpm events in the southwestern Cape correspond roughly to the 4200gpm events in the northern regions.

The EFS is able to capture these events in South Africa adequately for the first seven days of the forecast, as shown by positive Brier Skill Scores (BSS) and ROC scores in excess of 0.5 (Fig. 13 and 14). Although reliability and resolution deteriorate during the second week, the forecasts still retain reasonable resolution throughout the forecast period. This suggests that calibration may be able to improve the skill of these forecasts.

The first calibration method tested here is the bias correction described above, where the 850-500hPa thickness fields were adjusted at each grid point and new thickness probability fields calculated. This method was successful in improving the BSS,

resolution and reliability considerably during the second week (Fig. 13). Rare events show a significant improvement in the BSS, but this is not reflected in the ROC scores, which are related to forecast resolution (Fig. 14).

The second calibration method, where the event threshold is adjusted to match the forecast probability with the observed frequency, is more successful in improving both the BSS and ROC score for the 4200gpm events (Fig 13). The reliability is much improved (as the calibration is intended to do) and there is also a small increase in resolution during the second week. For the more rare 4100gpm event, the BSS is again improved and the ROC score is worse than the raw output (Fig. 14). Overall the first calibration method is more successful as it addresses the bias in the physical patterns more directly. However, neither calibration method does much for the resolution of the more extreme (rarer) events and is consistent with findings in Legg and Mylne (2004) where the skill in predicting extreme events often deteriorates after bias-correction.

Forecast uncertainty is another useful forecaster guidance tool. This is indicated by shading ensemble spread (categorized into a strong, medium, or weak signal and no predictability) under the control prognostic 500Z and sea-level pressure fields. It is determined using a basic definition of calculating the ensemble standard deviation around the ensemble mean and dividing by the observed field standard deviation for that time of year at each grid point (Scherrer et al. 2004). Values from 0-33%, 33-66%, 66-100% and > 100% correspond to strong, medium, weak and no signal respectively. Periods of strong atmospheric instability may also lead to a large ensemble spread but not

necessarily increased forecast uncertainty. Toth et al. (2001) proposed a relative measure of predictability based on the position of the forecast value in terms of climatological distribution to provide a quantitative probability of forecast uncertainty. The intention at the SAWS for now however, is that forecasters use these uncertainty fields as a qualitative rather than a quantitative measure of forecast confidence. Notwithstanding, further development and refinement of this process is underway at the SAWS.

4. Discussion

The EFS was introduced to the SAWS National Forecast Centre (NFC) in the beginning of 2004. Although this forecast guidance system was received apprehensively at first, consistent benefits in using this system have been experienced during this time and now the EFS is used widely in the NFC and regional forecast offices. This section now relates the application of the scientific aspects covered in section 3 to local forecaster experiences.

The EFS 1-14 day thumbnail webpage display system is particularly useful to a forecaster, as one quickly gains an overview of each of the different parameters, extending from 1 through to 14 days. With a glance, one can rapidly assess the broad trend in one or more parameters, e.g. is rain generally increasing (decreasing) over time and is the rain indicated to be “mostly in the west” at the beginning of the sequence – possibly migrating eastwards as one advances through the forecast period? Speed of assessment (of a weather pattern, whether the pattern at hand pertains to one playing out in the next few hours or in the next few days) within the forecast office environment, is a

theme we cannot escape from. Effective time-management is critical in an operational forecast setting – so any visualization scheme (such as the EFS html pages) that can cut time and allow a forecaster to reach an informed decision, is a very welcome operational tool. The product is easily accessible and the auto-update function ensures one always views the latest data.

NWP models have always been relatively good at anticipating sudden temperature falls (rises) from one day to the next – especially frontally-induced cooling of 10 deg C or more (per 24hrs). The ensemble approach sustains this ability but at lead times generally surpassing that of deterministic NWP products. Notwithstanding poor predictability at extreme lead times, a forecaster is at least able to give the public a “heads-up” (heightened state of alertness), with a fair degree of confidence, in respect of possible extreme or inclement weather, at lead times commonly 3 to 10 days hence (and even 11 to 14 days at a push). Such outlooks or weather-related guidance do not pertain only to general curiosity from the general public but from a commercial and public-safety aspect, such guidance can be crucial. Two local examples are Fire Protection Agencies (FPA), who need to make tactical decisions to hire helicopters / fixed-wing aircraft / pilots on short-term contracts (days or weeks), ahead of expected breakouts of hot dry Berg winds (air cascading from the interior of South Africa down the escarpment toward the coast). A second example is that of small stock farmers, who need to know, at least 2 to 4 days in advance, of impending cold, wet and windy conditions, which may or may not include snow concurrently. The ensemble approach greatly assists the forecaster in assessing the likelihood of such events, which have the potential to affect life and/or property; at

timescales often exceeding those of regular NWP methods (as alluded to earlier). In addition, shaded uncertainty fields give the forecaster a quick eyeball overview as to relative uncertainty as to control run predisposition towards a particular feature (cut off low etc) or event (such as ridging surface high) – both in space and time. The spaghetti diagram link can then assist the forecaster by providing more information on the individual ensemble members handling of the weather system in question.

Probabilities of the 850-500hPa thickness fields dropping below 4200 and 4100gpm is an additional useful tool in assessing relative threat from impending snow events. Situational awareness has improved markedly within the NFC environment – nurturing active discussion and debate well ahead of such events – overcoming negative aspects such as “forecaster inertia” where (perhaps due to excessive non-weather related workload, or other factors) a forecaster may be unaware of an unfolding severe weather pattern and only build up an awareness too late. It is essential (especially for high impact weather) to raise awareness (amongst forecasters and farmers alike) with a sufficiently long lead-time as to allow appropriate mitigative measures to be taken (such as bringing young, vulnerable animals down, out of the hills, a day or two ahead of the onset of such an extreme change). Naturally one needs a suitable balance between early warnings versus an inflated alarm rate. Generally, the ensemble products are utilized to identify possible/probable extreme events at lead times approaching (or often exceeding) a week hence. Deterministic high-resolution (regional model) NWP guidance is then utilized, closer to the time, to closely monitor (in space and time) the unfolding weather event and refine the official forecast scenario ahead of the perceived extreme weather event.

Usually forecasters would only be interested in a single best forecast tool (e.g. the high-resolution control run). However, outlier values and more specifically groups or clusters of outliers are also of interest and value to a forecaster, particularly during the developmental phase of assembling a prognosis. These may well be indicative of an alternative outcome/scenario, from a weather perspective. Experience has shown that most popular scenario is not necessarily the outcome that is finally realized. A secondary (or even tertiary) clustering of members away from the control run might well be indicative of an alternate, and possibly significant outcome. Such alternative scenarios could be incorporated in a forecast prognosis or sometimes it might be more appropriate to adhere to the prognosis implied by the control run but to take cognizance of the herald of a possible deviation from the expected pattern – this would thus imply that the forecaster should closely monitor real-time observations / developments in the area of interest to be aware of a possible deviation or shift towards an alternative scenario and amend/update forecast products accordingly.

A good example of such a case in South Africa that illustrates this point is the 7-day forecast for 12Z 6 November 2005. High-resolution control runs from ECMWF and NCEP differed drastically as to the position and intensity of a cut-off low event over the southwestern Cape (Fig. 15). Synoptically these two scenarios are very different and would have a huge impact on QPF and warnings/advisories. The spaghetti diagrams interestingly showed a similar dichotomy in the prognosis, one cluster similar to ECMWF and one similar to NCEP control. Although the previous NCEP control run

(00Z – bold dashed line in figure 15) resembled the ECMWF control at 12Z, the perturbed ensemble members around the 00Z and 12Z analyses were spread fairly uniformly across both scenarios showing that this situation was particularly sensitive to perturbations in the initial conditions at the start of this 7-day forecast. Furthermore uncertainty fields warned of especially low levels of forecast confidence in this particular event. Consequently the forecasts were done adopting a more cautious approach until greater certainty was evident. It turns out the scenario with the system weaker and displaced to the northeast was closer from a forecaster point of view. However, the system was indeed intense with late spring snow (up 17cm) over the mountains of the Eastern Cape and heavy rainfall along the coast (201mm in East London over the weekend). Several severe storms with medium-sized hail and damaging winds were reported from many parts of the northeast interior of South Africa as well. The suggestion of an intense system, such as forecast by the 12Z ECMWF run on 30 October and confirmed by a significant number of ensemble members, was sufficient to place forecasters on alert to monitor changes very closely up until the event played out. Without the ensemble support the ECMWF forecast scenario might well have been ignored, as the following update was vastly different, or premature warnings could have been issued for the wrong areas.

The EFS approach to forecasting in the medium-term thus frees the forecaster to a certain degree, from being excessively constrained by a single official NWP prognosis. The forecaster is thus empowered to explore (to a limited degree) a spectrum of possible outcomes and juxtapositions of systems in a spatial-temporal context, allowing more

creativity and flexibility on the part of the forecaster / analyst – but at the same time, retaining the strength and support of solid NWP guidance principles and products. Until the advent of the NCEP EFS at SAWS, the forecaster had to really go out on a limb to attempt to visualize possible weather scenarios at lead times of beyond a few days – this in itself was difficult enough. Furthermore, even to attempt variations on those scenarios was near impossible. The medium term Ensemble NWP goes a long way to fulfilling this need.

Bias correction methods discussed in this paper have proved quite successful in improving the EFS forecast skill statistics. There still remain some additional avenues to explore for South Africa, including regime-dependent correction and perhaps ensemble model output statistics (EMOS) (Gneiting et al. 2005). However, forecasters when armed with EFS guidance, even if not calibrated, can make useful forecasts by using their analytical powers to sort through the model forecasts (Bosart 2003). Still one area of concern is the tendency for calibration to adversely affect forecasts of extreme events (Legg and Mylne 2004; Kharin and Zwiers 2003). Thus it is advisable to provide forecasters with both the calibrated and non-calibrated EFS guidance. Accurately forecasting high-impact weather is one of the primary responsibilities of the forecasters, and so they need to be able to fully utilize the EFS guidance to fulfill this need. The success of this lies clearly in training and experience.

5. Summary and Conclusions

The NCEP EFS has been successfully implemented as an integral part of the SAWS forecasting service. Consistent benefits of using the EFS have been noted in the forecast offices and this has strengthened the position of this forecasting tool into the operational environment. Foremost among these benefits is an improved hit rate of forecast high-impact weather events up to and beyond a week ahead. False alarms in these forecasts have also been kept in check by the ability of the EFS to provide useful information regarding the uncertainty in the forecast scenarios.

The EFS is particularly useful to generate objective forecaster guidance products and public forecast products. Probability distribution functions can be calculated objectively using EFS data. These can be used directly as automated end-user products or as tailored forecaster guidance to suit local conditions.

Bias correction of atmospheric fields (e.g. 500hPa geopotential heights) and probability forecasts (e.g. quantitative precipitation forecasts) has proved quite successful at improving forecast reliability. However, the bias correction methods tested in this paper do not lead to much improvement in the anomaly correlation coefficient scores of 500hPa heights (spatial pattern skill). Similarly for the probability forecasts, bias correction really only corrects part of the problem omitting the extreme events. Therefore, in order to capture extreme events more sophisticated post-processing methods are needed.

Another issue currently not properly handled by the EFS systems is QPF at station scale in the summer rainfall areas of South Africa. Although the EFS does verify better at grid-

scale than at station scale, the skill of these forecasts is still relatively poor and are generally only useful for the first few days of the forecast. Part of the problem with summer convective rainfall is the poor correlation between station rainfall and the synoptic situation. The suggested approach to this problem is through a combination of high-resolution modeling and MOS-type post-processing.

Experience of the introduction of an EFS into the operational forecasting environment in South Africa is that forecasters are willing to use EFS provided that the benefit of using lower resolution EFS instead of a single high-resolution control is properly demonstrated and that there is adequate training to assist in understanding and using this forecast tool. Finally, it is imperative that dynamic interaction between forecasters and researchers takes place in order to facilitate the timely implementation of desired products.

Acknowledgments. The authors wish to thank two anonymous reviewers for their helpful and thorough comments that have resulted in significant improvements to this paper.

References

Atger, F., 2003: Spatial and Interannual Variability of the Reliability of Ensemble-Based Probabilistic Forecasts: Consequences for Calibration. *Mon. Wea. Rev.*, **131**, 1509-1523.

Bosart, L. F., 2003: Whither the Weather Analysis and Forecasting Process? *Wea. Forecasting*, **18**, 520-529.

Bourke, W., R. Buizza, and M. Naughton, 2004: Performance of the ECMWF and the BoM Ensemble Prediction Systems in the Southern Hemisphere. *Mon. Wea. Rev.*, **132**, 2338-2357.

Brier, G. W., 1950: Verification of Forecasts Expressed in terms of Probability. *Mon. Wea. Rev.*, **78**, 1-3.

Buizza, R., P.L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076-1097.

Chessa, P.A., and F. Lalaurette, 2001: Verification of the ECMWF Ensemble Prediction System Forecasts: A Study of Large-Scale Patterns. *Wea. Forecasting*, **16**, 611-619.

Cui, B., Z. Toth, Y. Zhu, D. Hou, and S. Beauregard, 2005: Statistical post-processing of operational and CDC hindcast ensembles. *Proc. 21st Conference on Weather Analysis and Forecasting/ 17th Conference on Numerical Weather Prediction*. Washington, DC, Amer. Meteor. Soc., 12B.2.

Ebert, E. E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480.

Ebert, E. E., U. Damrath, W. Wergen, and M. E. Baldwin, 2003: The WGNE Assessment of Short-Term Quantitative Precipitation Forecasts. *Bull. Amer. Meteor. Soc.*, **84**, 481-492.

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

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Mon. Wea. Rev.*, **133**, 1098-1118.

Hamill, T. M., and S. J. Colucci, 1998: Evaluation of Eta-RSM Ensemble Probabilistic Precipitation Forecasts. *Mon. Wea. Rev.*, **126**, 711-724.

Hamill, T. M., 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, **129**, 550-560.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. *Mon. Wea. Rev.*, **132**, 1434-1447.

Kalnay, E., and co-authors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

Kharin, V. V., and F. W. Zwiers, 2003: On the ROC Score of Probability Forecasts. *J. Climate*, **16**, 4145-4150.

Legg, T. P., and Mylne, K. R., 2004: Early Warnings of Severe Weather from Ensemble Forecast Information. *Wea. Forecasting*, **19**, 891-906.

Leith, C. E., 1974: Theoretical Skill of Monte Carlo Forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291-303.

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. *Wea. Forecasting*, **17**, 173-191.

Murphy, A. H., 1973: A New Vector Partition of the Probability Score. *J. Appl. Meteor.*, **12**, 595-600.

Richardson, D.S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473-2489.

Roebber, P. J., D. M. Schultz, B. A. Colle, and D. J. Stensrud, 2004: Toward Improved Prediction: High-Resolution and Ensemble Modeling Systems in Operations. *Wea. Forecasting*, **19**, 936-949.

Ryan, R. T., 2003: Digital Forecasts - Communication, Public Understanding, and Decision Making. *Bull. Amer. Meteor. Soc.*, **84**, 1001-1003.

Scherrer, S. C., C. Appenzeller, P. Eckert, and D. Cattani, 2004: Analysis of the Spread-Skill Relations Using the ECMWF Ensemble Prediction System over Europe. *Wea. Forecasting*, **19**, 552-565.

Stratton, R. A., 1999: A high resolution AMIP integration using the Hadley Centre model HadAM2b. *Climate Dyn.*, **15**, 9-28.

Szunyogh, I., and Z. Toth, 2002: The Effect of Increased Horizontal Resolution on the NCEP Global Ensemble Mean Forecasts. *Mon. Wea. Rev.*, **130**, 1125-1143.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability, 20-22 October 1997*, ECMWF, Shinfield Park, Reading, United Kingdom, 1-26.

Tennant, W. J., 2003: An assessment of intra-seasonal variability from 13-yr GCM simulations. *Mon. Wea. Rev.*, **131**, 1975-1991.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Toth, Z., Y. Zhu, and T. Marchok, 2001: The use of ensembles to identify forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 436-477.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*.
Academic Press, San Diego, CA, 467 pp.

Figure Captions

Figure 1: Location of temperature (large dot) and rainfall (small dot) stations in South Africa, in relation to the NCEP EFS model grid boxes.

Figure 2: Examples of week-2 temperature and wind probabilistic forecasts generated directly from the NCEP EFS data. Forecasts of warmer than average maximum temperatures and normal minimum temperatures for Johannesburg, and a higher incidence of easterly winds with overall lower wind speeds relative to climate for East London, are indicated in these products.

Figure 3: Root Mean Square Error (RMSE) of 500hPa geopotential height forecasts against forecast lead-time averaged over the period January 2001 to December 2004 for the domain 30°W - 60°E and 0 - 60°S . Graphs for high-resolution control, low-resolution control, best ensemble member for each forecast case, ensemble mean, climatology and the ensemble spread defined in legend.

Figure 4: Anomaly Correlation Coefficients (ACC) of 500hPa geopotential height forecasts against forecast lead-time averaged over the period January 2001 to December 2004 for the domain 60°S 30°W - 10°S 60°E (left panel) and 37.5°S 10°E - 17.5°S 40°E (right panel). Graphs for high-resolution control, ensemble average and best ensemble member for each forecast case defined in legend. Bias-corrected scores are indicated by **BC**.

Figure 5: Standard deviation of 500hPa geopotential height forecast fields with the time-mean removed for the period January 2001 to December 2004, showing high-resolution control, low-resolution control, ensemble average and mean of perturbed ensemble members indicated in legend. Bias-corrected scores are indicated by **BC**.

Figure 6: Average bias of 500hPa geopotential height forecast fields for the period January 2001 to December 2004, showing high-resolution control, low-resolution control and ensemble average before and after bias correction (indicated by **BC** in the legend).

Figure 7: Spatial difference of the average of the perturbed ensemble members minus the low-resolution control analysis (left) and 24-hour forecast (right) for 500hPa geopotential height (top) and sea level pressure (bottom). Negative contours are stippled.

Figure 8: Spatial bias of 500hPa height forecasts before (left) and after (right) bias correction for week 1 (forecast days 1 to 7) (top) and week 2 (forecast days 8-14) (bottom) lead-time. Negative contours are stippled. Bias (un)corrected maps have a contour interval of (2) 0.5 gpm.

Figure 9: Annual average rainfall for the period July 2000 to June 2005 for observed gridded rainfall (top left), ensemble average 5-day forecast rainfall as a percentage of observed (top right), high-resolution 5-day control forecast (bottom left) and low resolution 5-day control forecast (bottom right).

Figure 10: Summer rainfall area (27.5°S 30°E) bias score (top row) and equitable threat score (bottom row) as a function of rain threshold for forecast lead-times of 5 days (left column), 10 days (middle column) and 15 (right column) days for high and low-resolution control, ensemble probability (>50% taken as a categorical yes forecast) and frequency adjusted high-resolution control and ensemble probability.

Figure 11: As for figure 10 but for the winter rainfall area (35°S 17.5°E).

Figure 12: Talagrand diagram for NCEP 5-day forecasts (all ensemble members) for the summer rainfall area (left) and winter rainfall area (right) area of South Africa for the period July 2000 to June 2005. The gray line shows a perfect distribution.

Figure 13: Brier Skill Score decomposition and Relative Operating Characteristic (ROC) of forecasts of the probability of the 850-500hPa thickness field dropping below 4200gpm at the grid point 35°S 17.5°E for the May-September months of 2004/5. Solid lines represent standard output, dashed line calibrated 850 and 500hPa height fields and dotted line frequency adjusted probabilities.

Figure 14: As for figure 13 but for 4100gpm thickness.

Figure 15: 7-Day 500hPa geopotential height forecast for 12Z06 November 2005 issued by ECMWF high-resolution control (top left), NCEP GFS high-resolution control (top right), spaghetti diagram of 5700gpm contour of NCEP 23-member ensemble suite

initialized on 2005103100 and 2005103112 (bottom left) and NCEP GFS analysis for 2005110612 (bottom right). Solid (dashed) bold line denotes the 00Z (12Z) high-resolution control forecast on 31Oct2005.

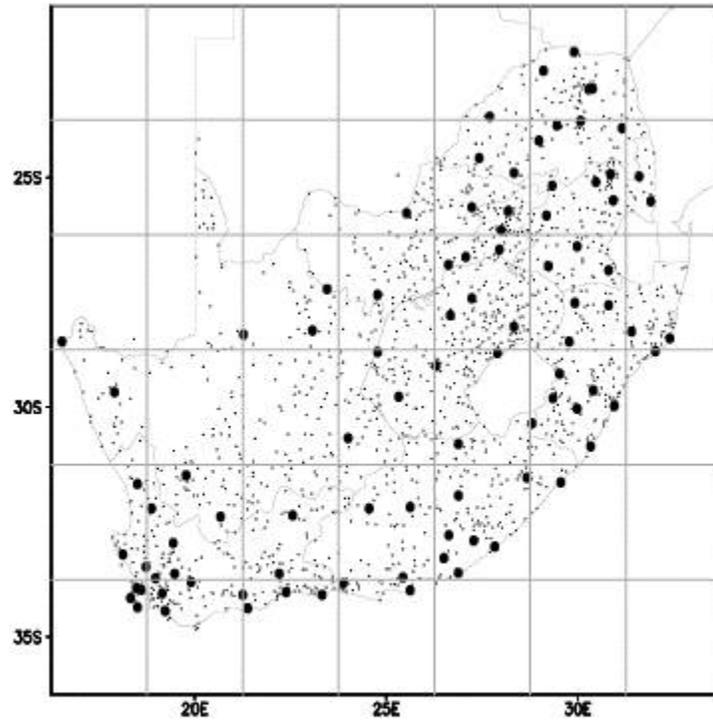
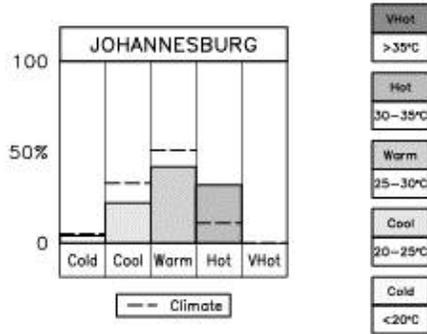
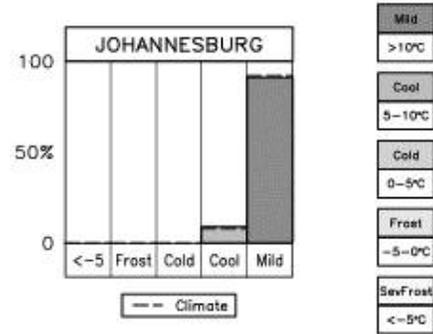


Figure 1: Location of temperature (large dot) and rainfall (small dot) stations in South Africa, in relation to the NCEP EFS model grid boxes.

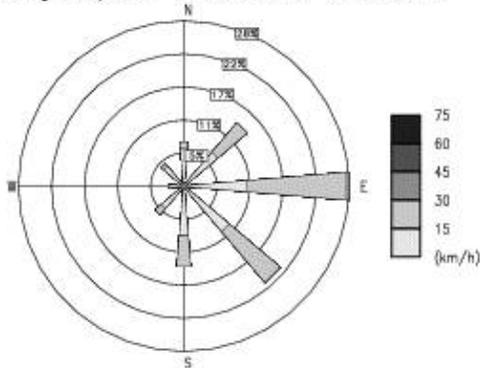
Probability distribution of Maximum Temperatures during the period 9 Nov 2005 to 15 Nov 2005



Probability distribution of Minimum Temperatures during the period 9 Nov 2005 to 15 Nov 2005



Windrose forecast for EAST_LONDON (km/h) during the period 9 Nov 2005 to 15 Nov 2005



Windrose for EAST_LONDON (km/h) November Climate

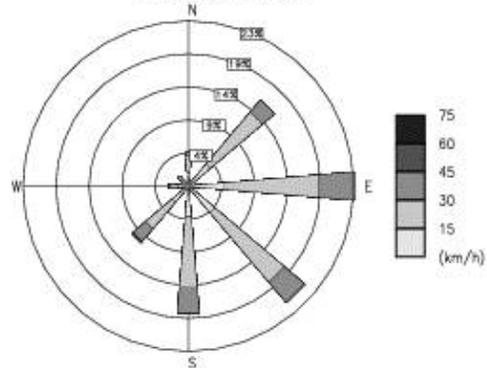


Figure 2: Examples of week-2 temperature and wind probabilistic forecasts generated directly from the NCEP EFS data. Forecasts of warmer than average maximum temperatures and normal minimum temperatures for Johannesburg, and a higher incidence of easterly winds with overall lower wind speeds relative to climate for East London, are indicated in these products.

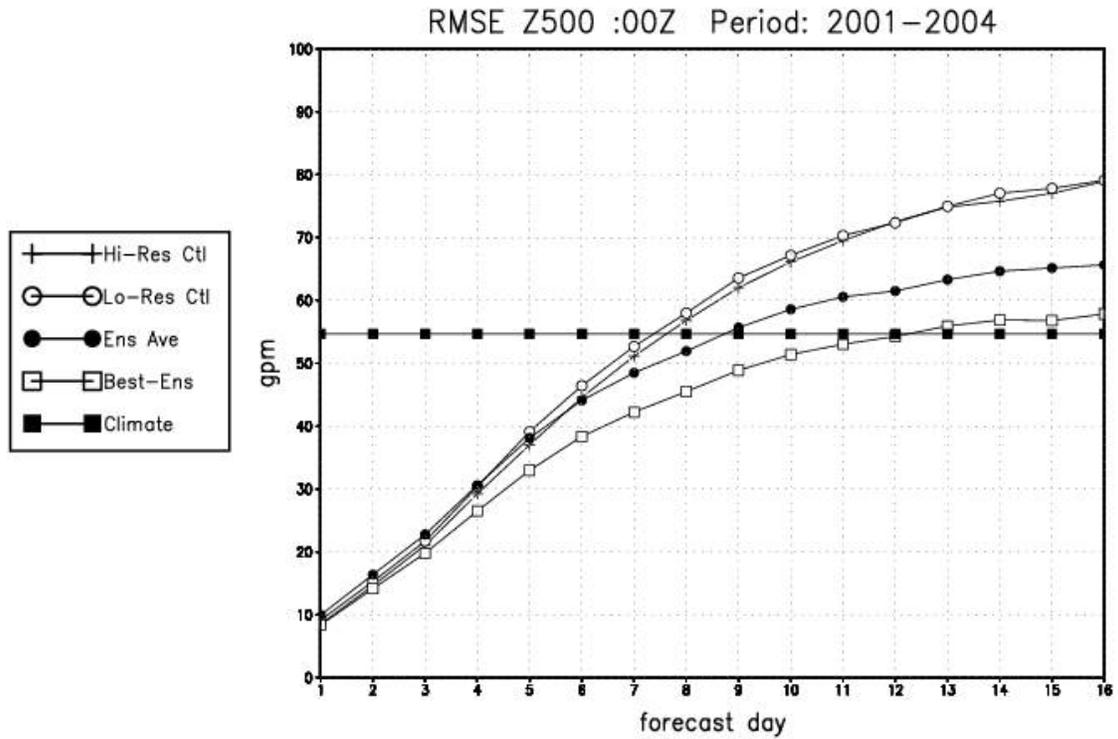


Figure 3: Root Mean Square Error (RMSE) of 500hPa geopotential height forecasts against forecast lead-time averaged over the period January 2001 to December 2004 for the domain 30°W–60°E and 0–60°S. Graphs for high-resolution control, low-resolution control, best ensemble member for each forecast case, ensemble mean, climatology and ensemble spread defined in legend.

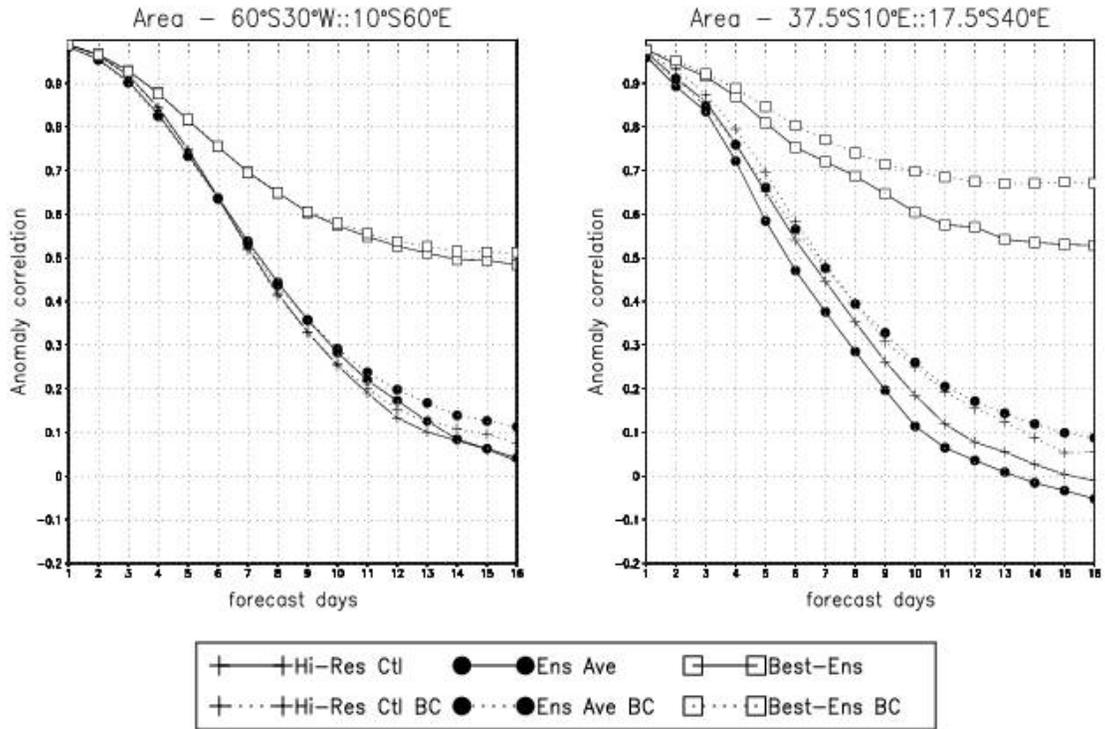


Figure 4: Anomaly Correlation Coefficients (ACC) of 500hPa geopotential height forecasts against forecast lead-time averaged over the period January 2001 to December 2004 for the domain 60°S30°W–10°S60°E (left panel) and 37.5°S10°E–17.5°S40°E (right panel). Graphs for high-resolution control, ensemble average and best ensemble member for each forecast case defined in legend. Bias-corrected scores are indicated by BC.

SD Z500 :00Z Period: 2001–2004

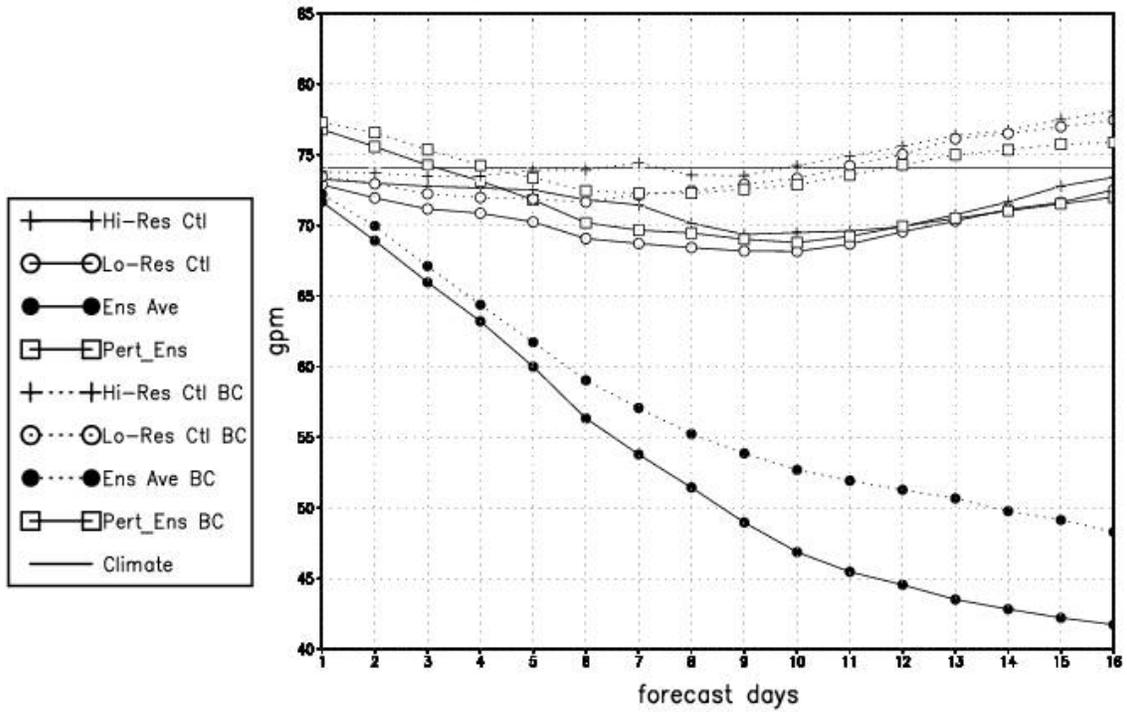


Figure 5: Standard deviation of 500hPa geopotential height forecast fields with the time-mean removed for the period January 2001 to December 2004, showing high-resolution control, low-resolution control, ensemble average and mean of perturbed ensemble members indicated in legend. Bias-corrected scores are indicated by BC.

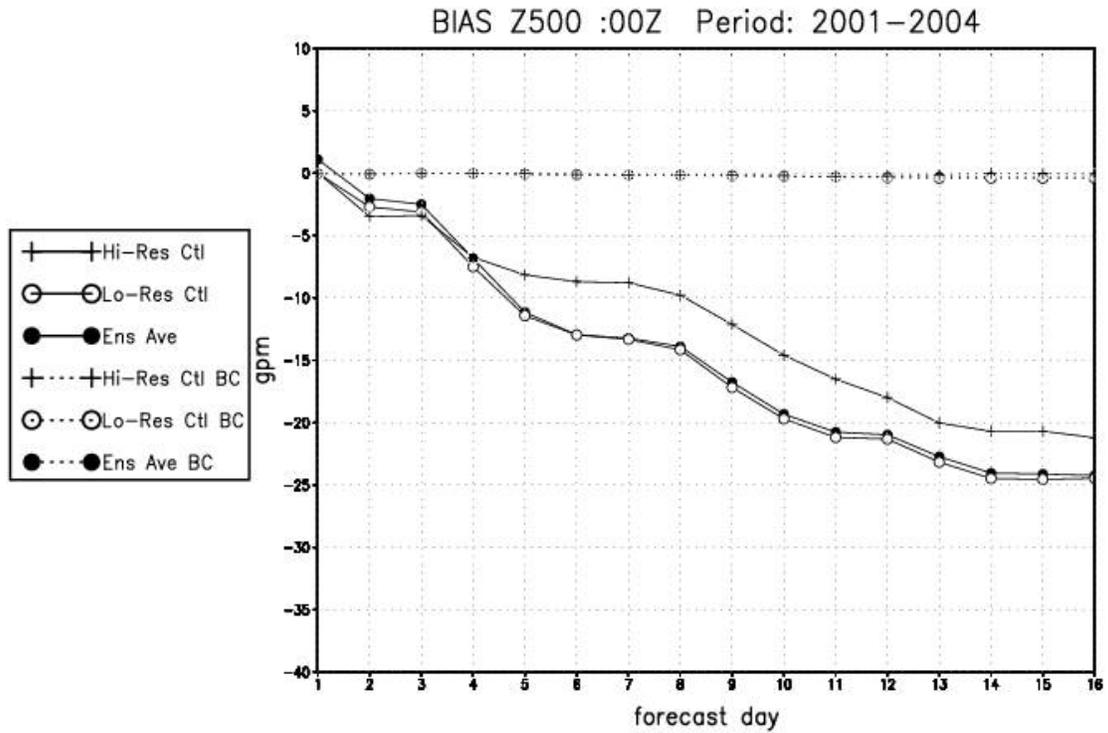


Figure 6: Average bias of 500hPa geopotential height forecast fields for the period January 2001 to December 2004, showing high-resolution control, low-resolution control and ensemble average before and after bias correction (indicated by BC in the legend).

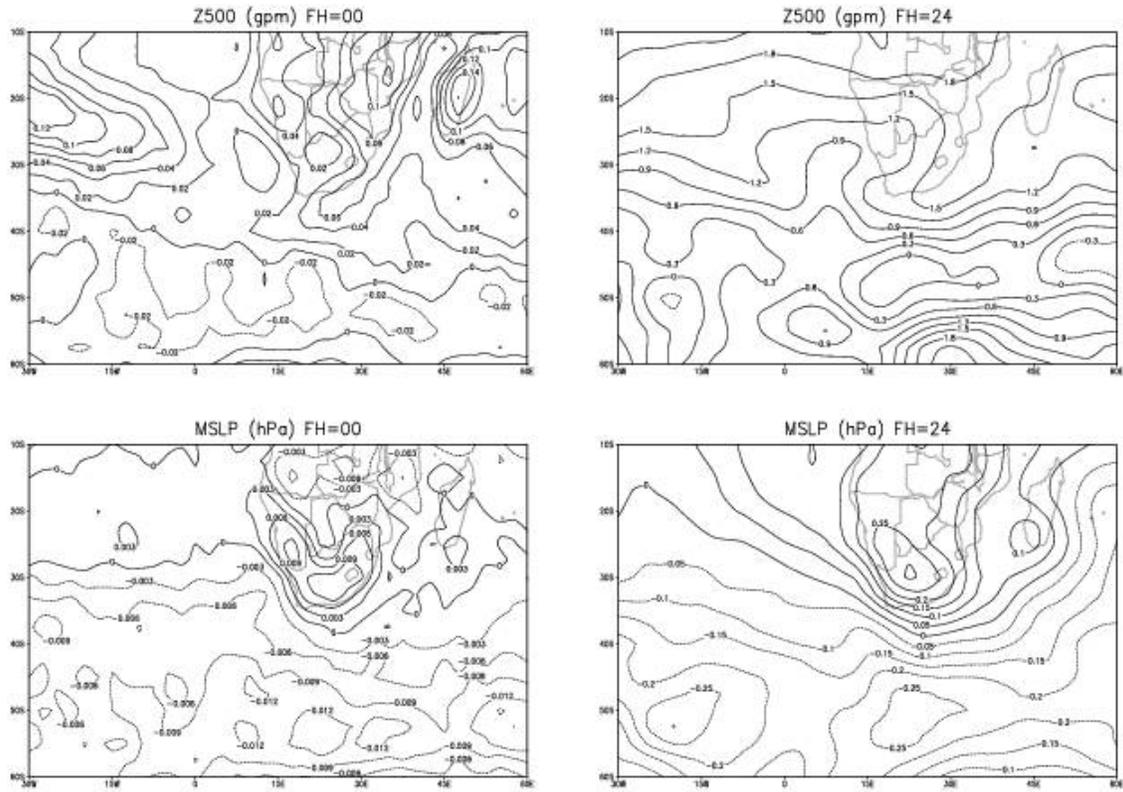


Figure 7: Spatial difference of the average of the perturbed ensemble members minus the low-resolution control analysis (left) and 24-hour forecast (right) for 500hPa geopotential height (top) and sea level pressure (bottom). Negative contours are stippled.

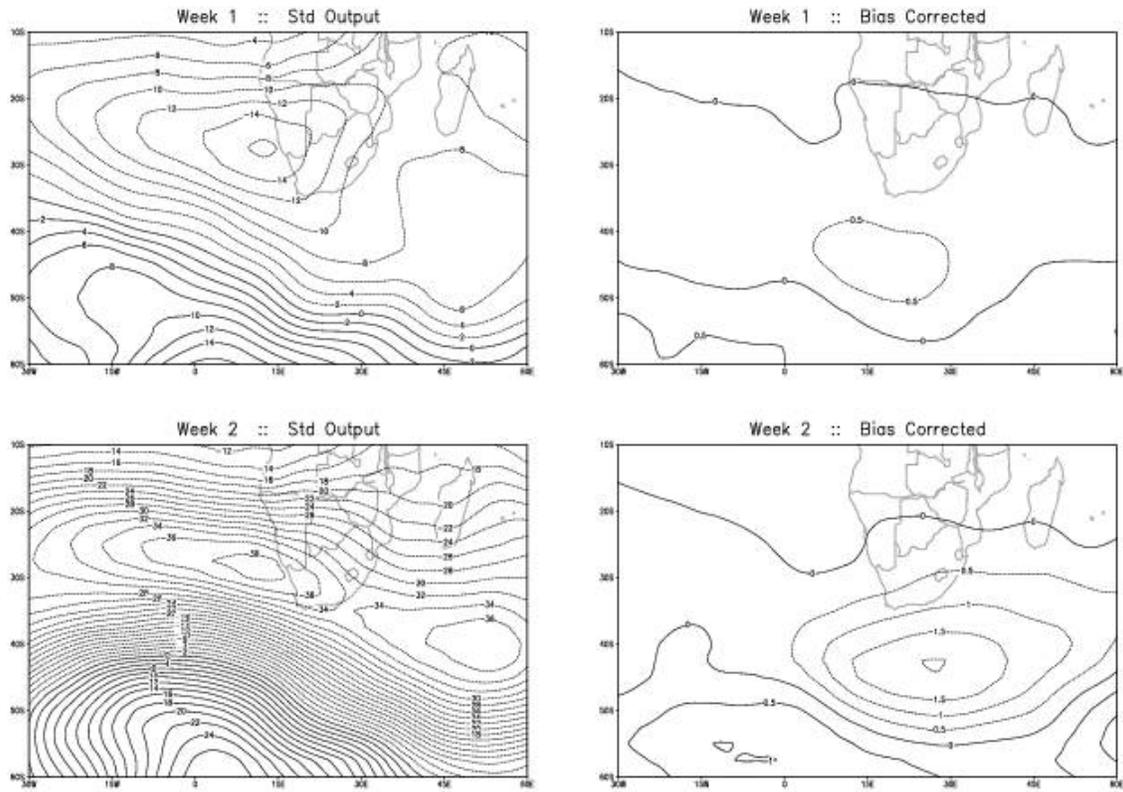


Figure 8: Spatial bias of 500hPa height forecasts (gpm) before (left) and after (right) bias correction for week 1 (forecast days 1 to 7) (top) and week 2 (forecast days 8–14) (bottom) lead-time. Negative contours are stippled. Bias (un)corrected maps have a contour interval of (2) 0.5 gpm

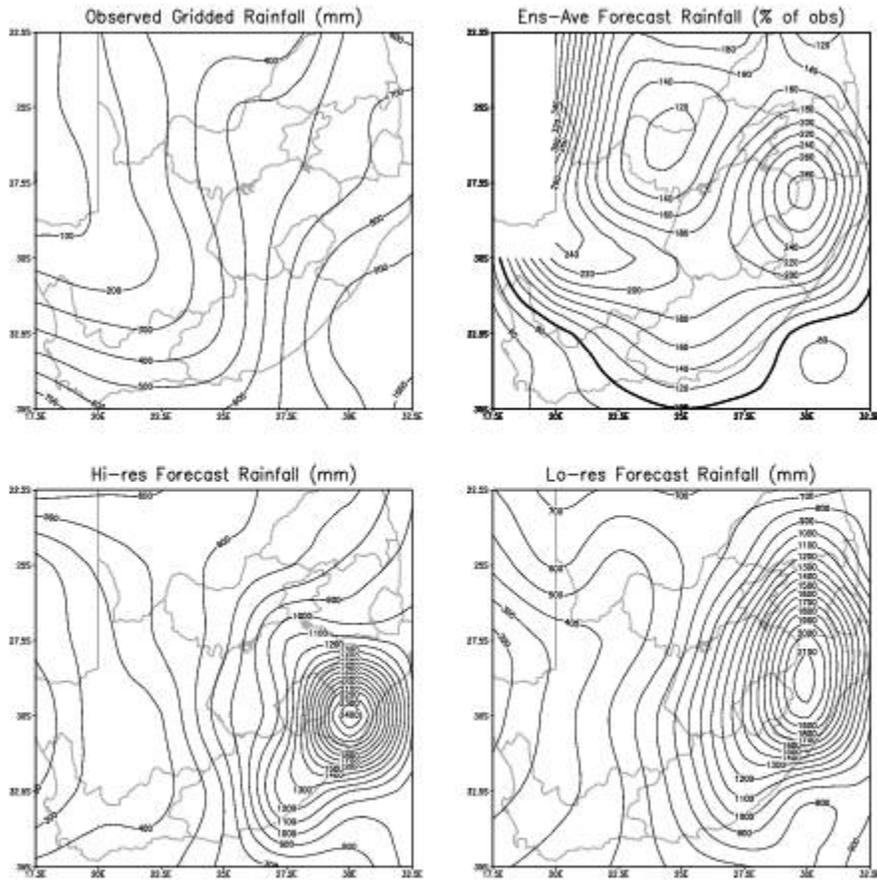


Figure 9: Annual average rainfall for the period July 2000 to June 2005 for observed gridded rainfall (top left), ensemble average 5-day forecast rainfall as a percentage of observed (top right), high-resolution 5-day control forecast (bottom left) and low resolution 5-day control forecast (bottom right).

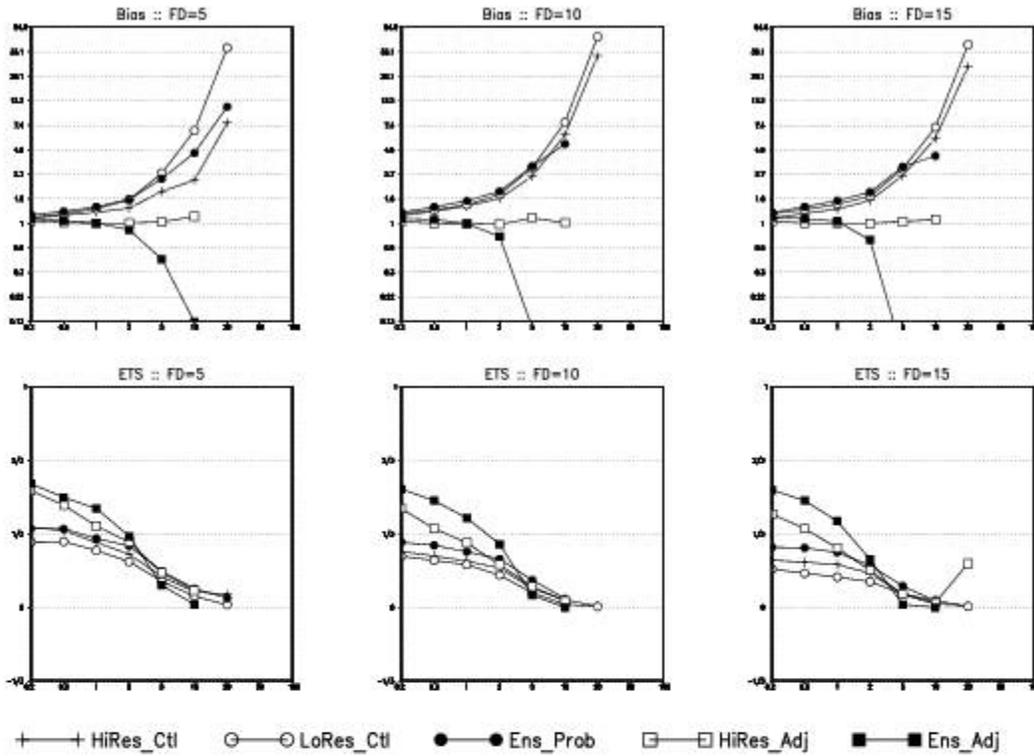


Figure 10: Summer rainfall area (27.5°S 30°E) bias score – cases forecast divided by cases observed (top row) and equitable threat score (bottom row) as a function of rain threshold for forecast lead-times of 5 days (left column), 10 days (middle column) and 15 days (right column) for high and low-resolution control, ensemble probability (>50% taken as a categorical yes forecast) and frequency adjusted high-resolution control and ensemble probability.

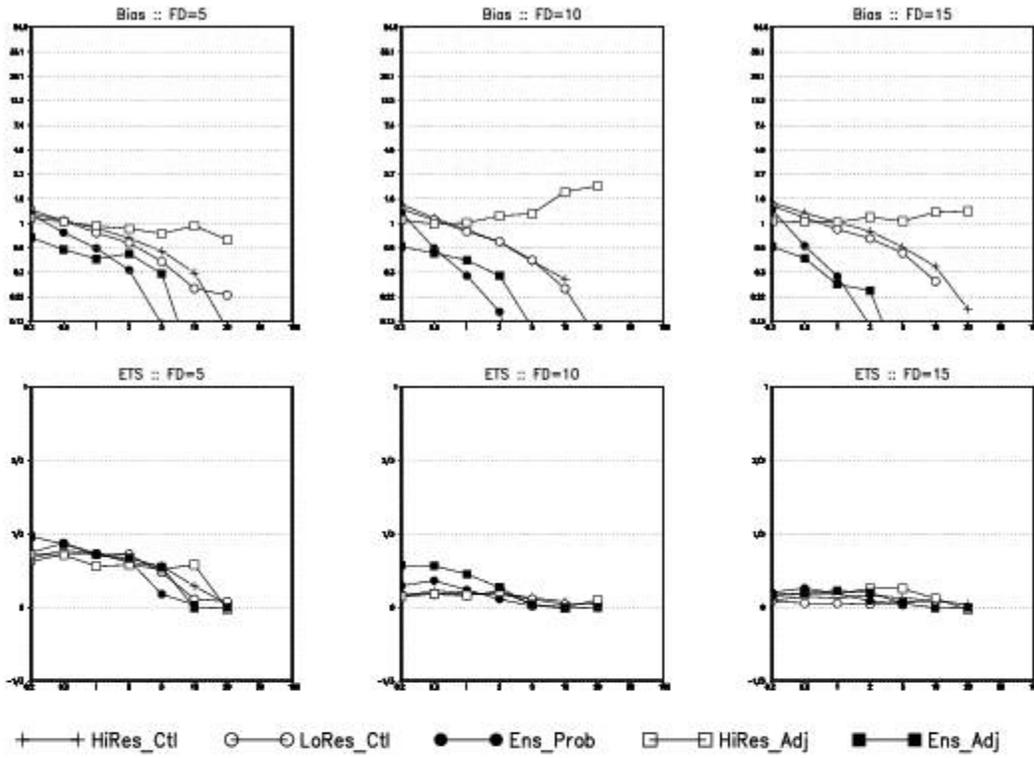


Figure 11: As for figure 10 but for the winter rainfall area (35°S 17.5°E)

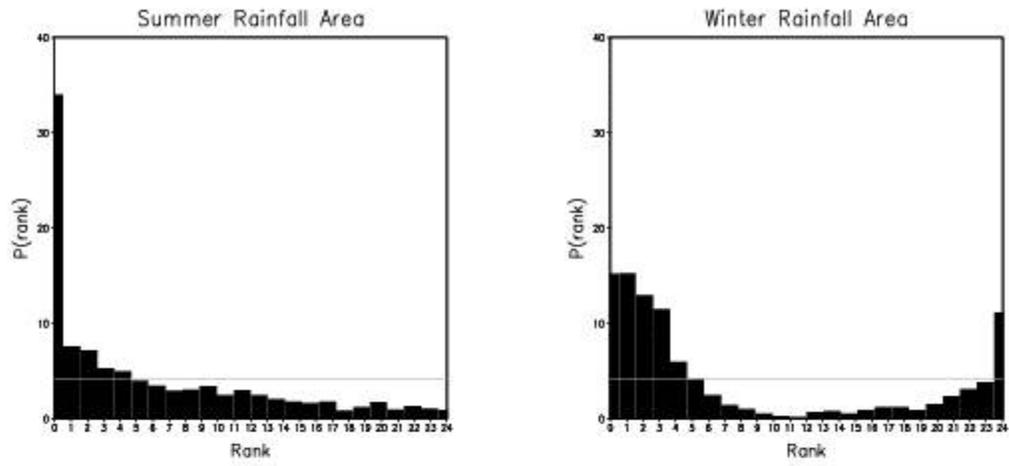


Figure 12: Talagrand diagram for NCEP 5-day forecasts (all ensemble members) for the summer rainfall area (left) and winter rainfall (right) area of South Africa for the period July 2000 to June 2005. The gray line shows a perfect distribution.

Period: MJJAS 2004 + 2005	Events=138 Total=296
---------------------------	----------------------

FxProb<4200gpm Grid Box :: 35S 17.5E

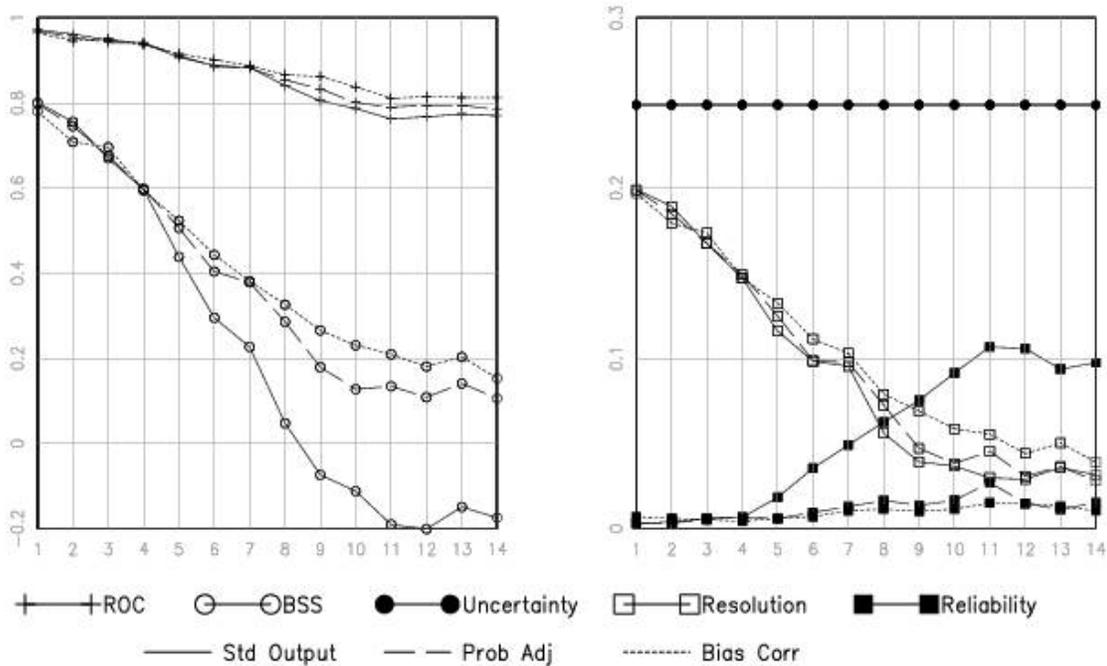


Figure 13: Brier Skill Score decomposition and Relative Operating Characteristic (ROC) of forecasts of the probability of the 850–500hPa thickness field dropping below 4200gpm at the grid point 35°S17.5°E for the May–September months of 2004/5. Solid lines represent standard output, dashed line calibrated 850 and 500hPa height fields and dotted line frequency adjusted probabilities.

Period: MJJAS 2004 + 2005	Events=20 Total=296
---------------------------	---------------------

FxProb<4100gpm Grid Box :: 35S 17.5E

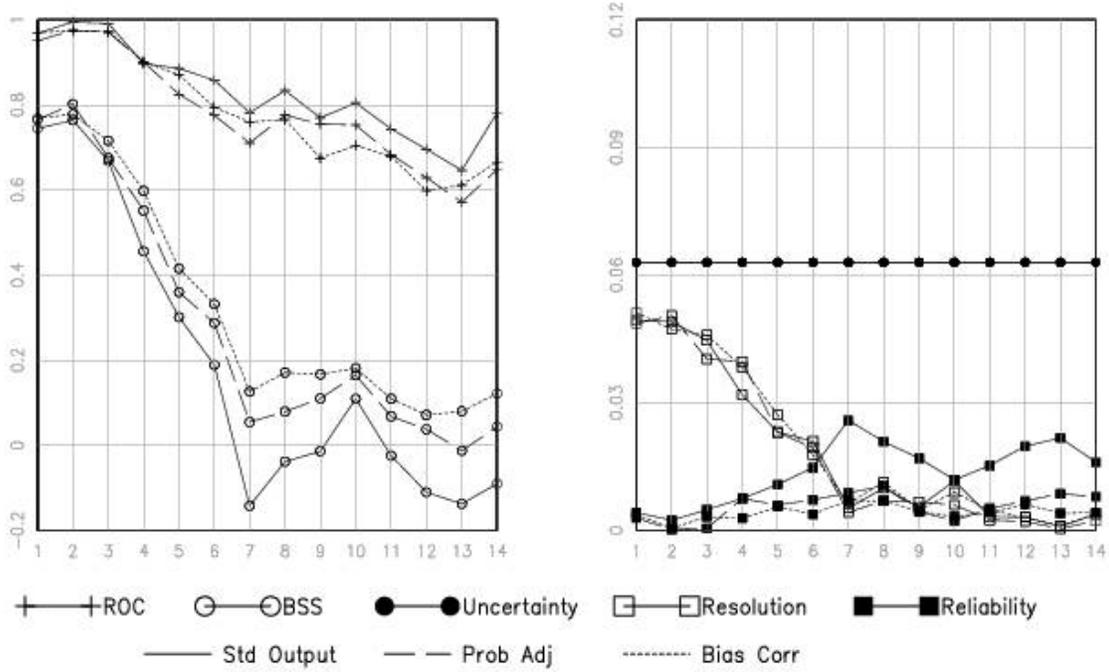


Figure 14: As for figure 13 but for thickness dropping below 4100gpm.

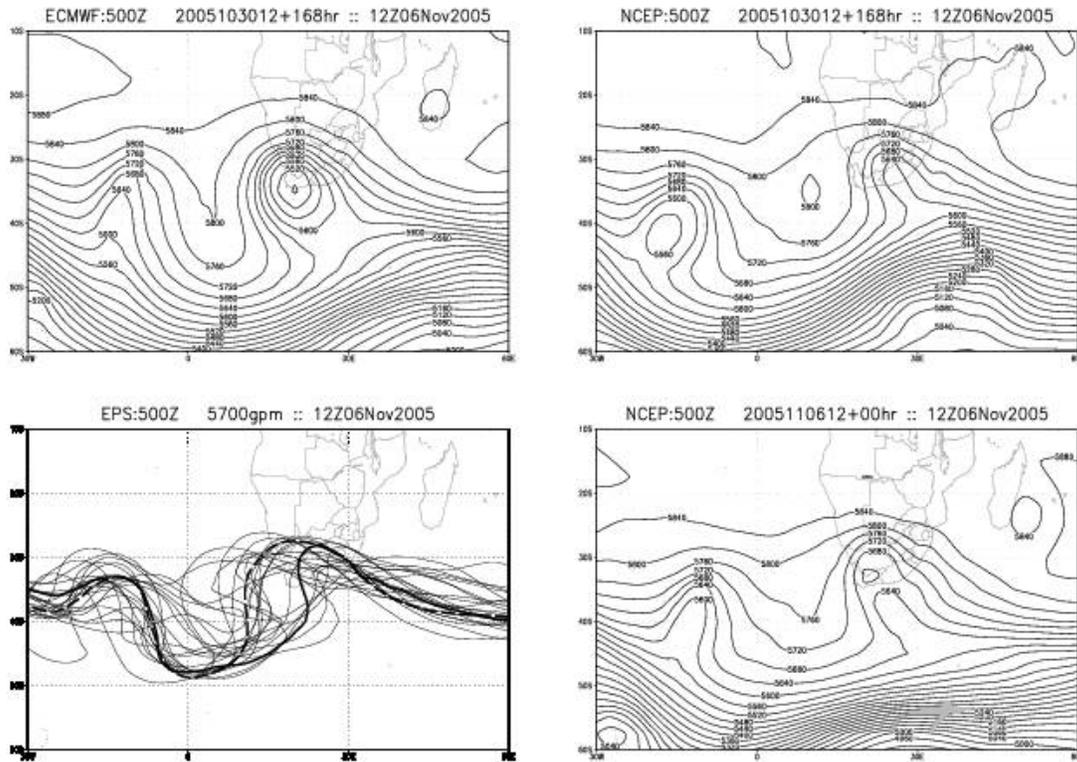


Figure 15: 7-Day 500hPa geopotential height forecast for 12Z06November 2005 issued by ECMWF high-resolution control (top left), NCEP GFS high-resolution control (top right), spaghetti diagram of 5700gpm contour of NCEP 23-member ensemble suite initialized on 2005103100 and 2005103112 (bottom left) and NCEP GFS analysis for 2005110612 (bottom right). Solid (dashed) bold line denotes the 00Z (12Z) high-resolution control forecast on 31Oct2005.