SAWS
Verification Bulletin

Year 1
Number 1

Preface

The main reason for doing verification of forecasts is to determine its quality and assess its value. One needs to assess errors in forecasts to develop a basis for improvement and users should be provided with this information. The aim of this bulletin is to present skill estimates of past operational forecasts issued by the SAWS, ranging from daily to seasonal. Verification is the process of comparing sets of forecasts and their corresponding sets of observations.

The first verified forecasts were the Finley forecasts of tornadoes (Murphy, A. H., 1996: The Finley Affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11**, 3-20). He created a Yes/No table for 18 regions and for the period between 10 March and 31 May 1884. A total of 2803 forecasts were used, from which 28 was Yes forecasted/Yes observed, 72 Yes forecasted/Not observed, 23 Not forecasted/Yes observed, and 2680 Not forecasted/Not observed. This gave Finley a Hit rate of 96.6%. This immediately started responses from different people. For example, Gilbert (Gilbert, G. K., 1884: Finley's tornado predictions. *Amer. Meteor. J.*, **1**, 166-172) noted that Finley's measure for accuracy took no account for the rare occurrence of tornadoes and that he would have gotten a higher Hit rate of 98.2% had he forecast no tornadoes every time. Responses like these are invited here also to help improve on the current verification system and this bulletin.

In addition to the forecasts produced by the current operational forecast systems already in place at the SAWS, developers of new forecast systems are also encouraged to contribute to this bulletin. In this manner other forecasters and researchers can respond to these new techniques, just like in Finley's time, and maybe bring up new questions, steering our minds in new directions to improve on our current forecast and verification systems.

Estelle Marx

In this issue

Seasonal Forecast Verification: 2001/02

The South African Weather Service (SAWS), the Universities of Pretoria and Cape Town, and also the International Research Institute for Climate Prediction (IRI) situated at the Lamont-Doherty Earth Observatory of Columbia University, New York, USA, produce seasonal forecast output for southern Africa every month from an array of various statistical and dynamical forecast models. The models' forecast fields are subsequently used by forecasters from the SAWS to subjectively consolidate the various forecast outputs into single forecast fields that provide guidance on the expected rainfall probabilities for upcoming 3-month seasons for southern Africa.

Seasonal forecasts should not be expressed deterministically (e.g., statements of the coming summer will be wet), since the inherent variability of the atmosphere requires that seasonal forecasts are expressed probabilistically. Two sets of forecast verifications are presented. The first set (top row) is for the first half of the summer season, October to December 2001, and the second set (bottom row) for the period January to April 2002. Four forecast maps are shown that represent statements of expected probabilities of below-normal (dry), near-normal (close to the seasonal average) and above-normal (wet) rainfall for the 2001/02 season.

Maps of the October to December (OND) forecasts show an expectation of enhanced probabilities for wet conditions. The forecast made at the longer lead-time (i.e., the OND forecast issued in July 2001) was for enhanced probabilities of wet conditions over the entire country, but the forecast issued in September 2001 suggested a reduced probability of wet conditions over the northeastern interior of the country where equal probabilities (40%) of wet and near the average conditions were forecast. The area of the largest positive rainfall anomalies of the OND 2001 season is seen to be over the central and western interior of the region. This observed rainfall spatial pattern is in closer agreement to the OND rainfall forecast produced at the shorter lead-time (issued in September 2001) than the rainfall forecast produced early July 2001 for the same season.

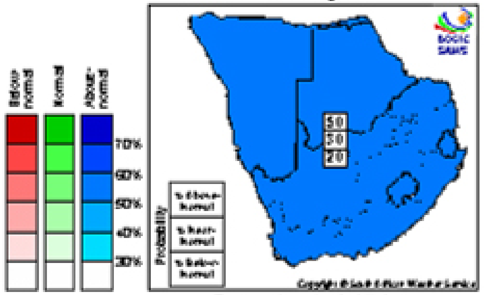
The January to March (JFM) 2002 rainfall season was associated with mostly dry conditions over the central and northeastern interior of South Africa and wet conditions over the far western areas. However, the JFM rainfall forecast produced in September 2001 was for mostly a favourable rainfall season over the entire summer rainfall region, even though the expected high probability (50%) of near the average rainfall was associated with a higher probability for dry conditions (30%) than for wet conditions (20%). Notwithstanding, the rainfall forecast produced in the beginning of the summer rainfall season (September 2001) for the second half of the season (JFM 2002) cannot be considered useful, owing to the lack of enhanced probabilities of dry conditions being forecast. Continued dry conditions were also observed during the February to April (FMA) season. The FMA rainfall forecast issued early January 2001 did show enhanced probabilities of average to dry conditions over the larger part of the region, consequently making the FMA rainfall forecast produced at a short lead-time more useful than the longer lead-time forecast of JFM rainfall.

Contributed by Willem Landman, Estelle Banitz and Anna Bartman

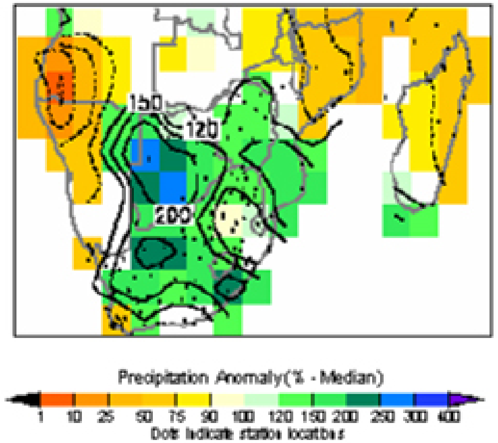
Observed Precipitation Anomaly OND 2001

Shaded ONLY for "ABOVE-Normal" & "BELOW-Normal"
[CAMS_OPI data, courtesy of NCEP/CPC]

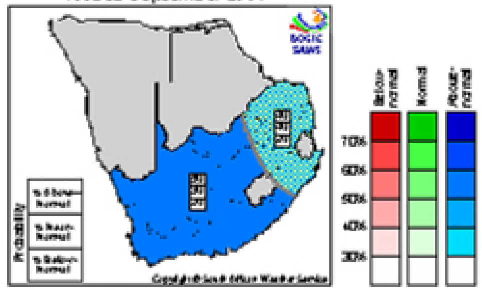
Issued July 2001



Expected total rainfall for
October + November + December 2001



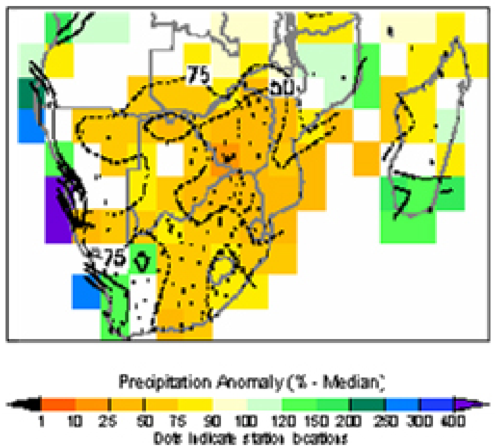
Issued September 2001



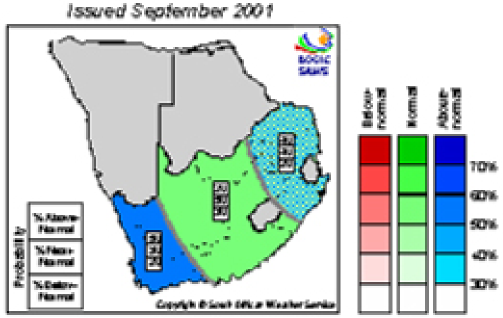
Expected total rainfall for
October + November + December 2001

Observed Precipitation Anomaly JFM 2002

Shaded ONLY for "ABOVE-Normal" & "BELOW-Normal"
[CAMS_OPI data, courtesy of NCEP/CPC]



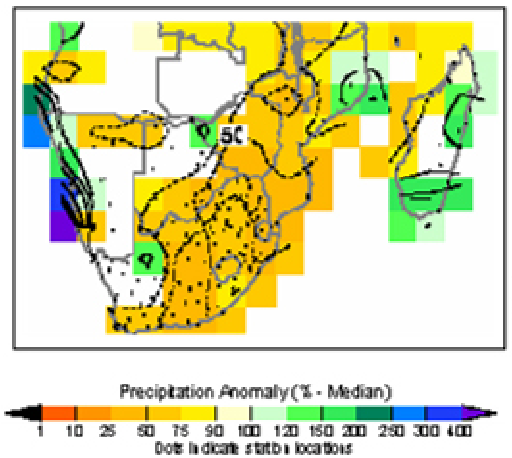
Issued September 2001



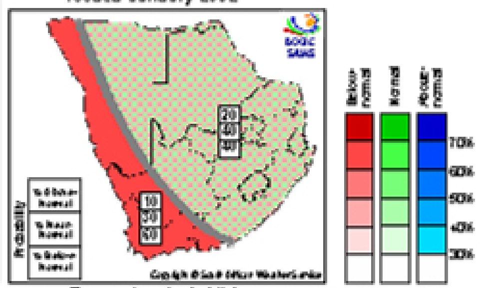
Expected total rainfall for
January + February + March 2002

Observed Precipitation Anomaly FMA 2002

Shaded ONLY for "ABOVE-Normal" & "BELOW-Normal"
[CAMS_OPI data, courtesy of NCEP/CPC]



Issued January 2002



Expected total rainfall for
February + March + April 2002

Verification of max and min temperature forecasts for the next day for "summer seasons" (6 months period 1 Oct to 31 March)

Below follows forecast verification results for 12 cities and towns in South Africa, preceded by an explanation of the graphs.

The top two graphs (bar graphs) for each town or city below, show the *frequency* distribution of forecast errors (forecasted minus observed temperatures), in other words the number (as a percentage) of forecasts that had errors within each error magnitude category. Note this is for the previous "summer" season only (1 October 2001 to 31 March 2002). The *observed* (measured) temperature was rounded of to the nearest whole number (e.g. 23.6 becomes 24), because the forecasted temperatures are in whole numbers. The bars show the percentage of errors (out of all forecasts) that fall in each of the 21 categories, from -10 to +10 (ten degrees Celcius too low to 10 ºC too high). For example, if the "-2" bar goes up to 23 percent, it means 23 percent of the forecasts were 2 degrees to *low*. Similarly, if bar 3 goes up to 12 percent it means that 12 percent of the forecasts were 3 degrees to *high*. Bar 0 is for the percentage of forecasts that were correct (i.e. the same as the rounded-off observed temperature).

Usually the longest bars (most frequent cases) cluster around the center category (the area of smaller errors) and the bars become shorter towards either side because larger errors occur less frequent.

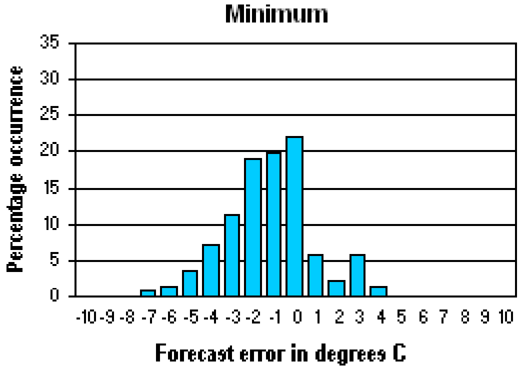
The bottom two graphs (line graphs) for each town or city show the trends in forecast errors for all "summer months" (1 October to 31 March) over the last 10 years. The top line (with diamonds) is the percentage of forecasts that were *within* 3 degrees Celcius correct (the absolute difference (higher or lower) between forecasted and rounded-off observed temperature was less or equal to 3 degrees). The second line from the top (squares) is the percentage of forecasts that were within 2 degrees correct and the bottom line (triangles) is the percentage of forecasts that were within 1 degree correct (less or equal to 1 ºC difference). For example, if the "-2 to 2" line goes through a point that corresponds to the 94/95 season on the x-axis and 80% on the y-axis, it means that for that season the percentage of forecasts that were within 2 degrees correct, was 80%. The same applies for any other point on the line – read off the percentage on the y-axis and the season on the x-axis.

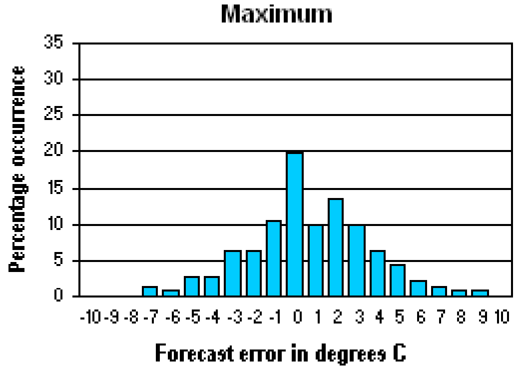
Spatial and temporal forecast accuracy changes are due to, amongst others, the topography and the varying types and frequency of weather systems from season to season.

Contributed by Louis van Hemert

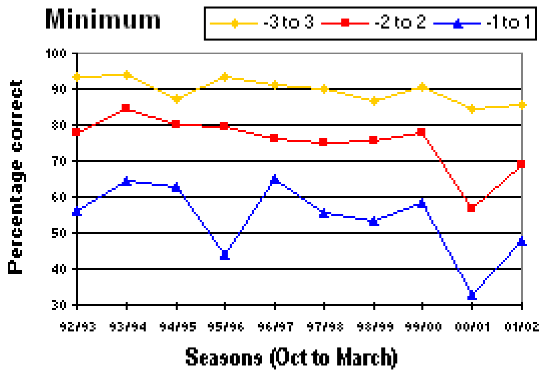
Bisho:

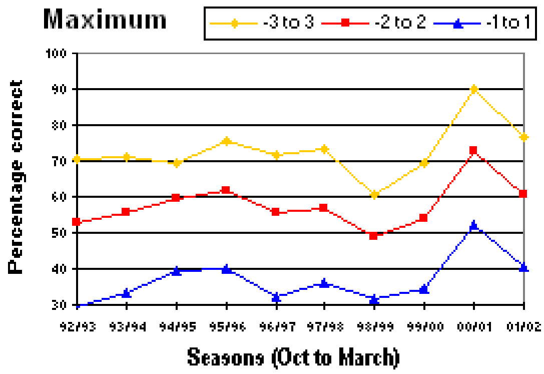
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





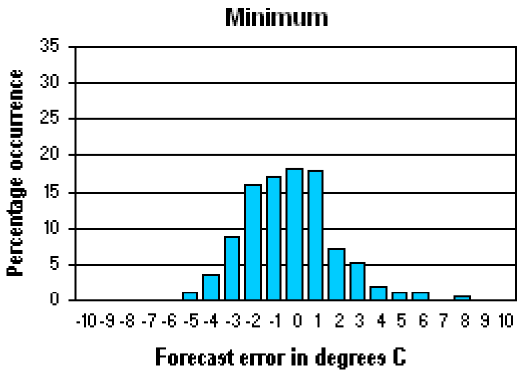
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

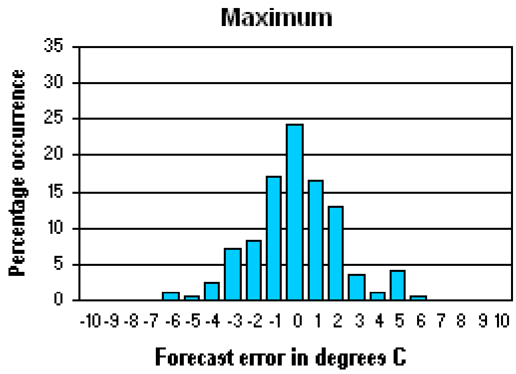




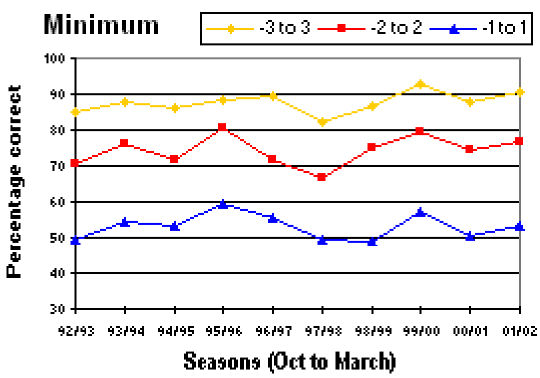
Bloemfontein:

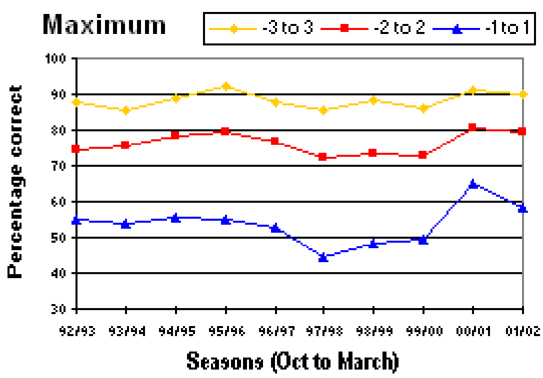
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





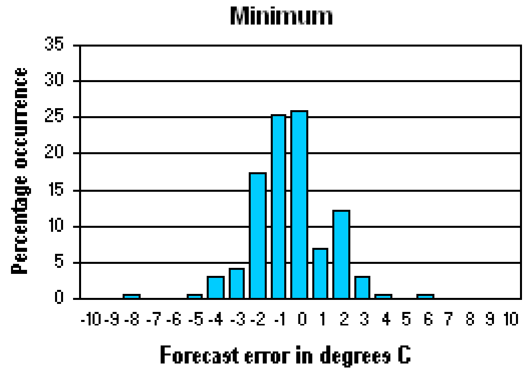
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

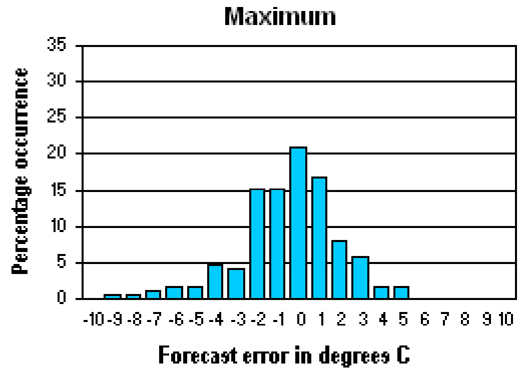




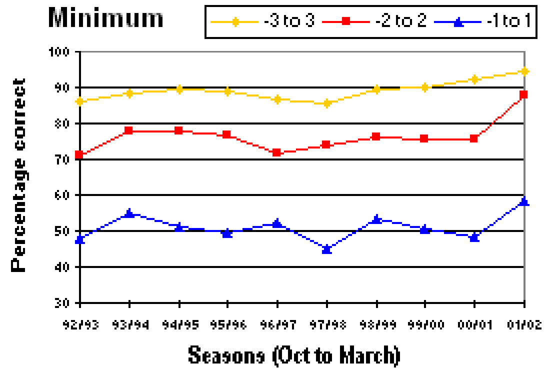
Cape Town:

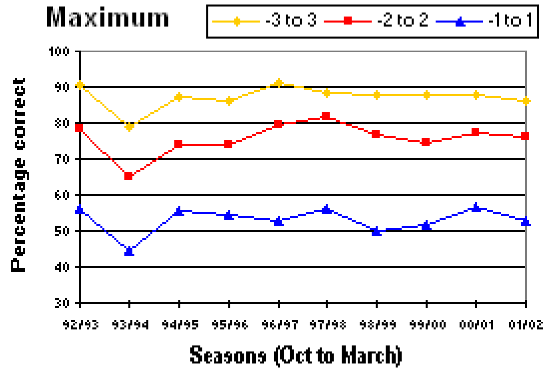
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





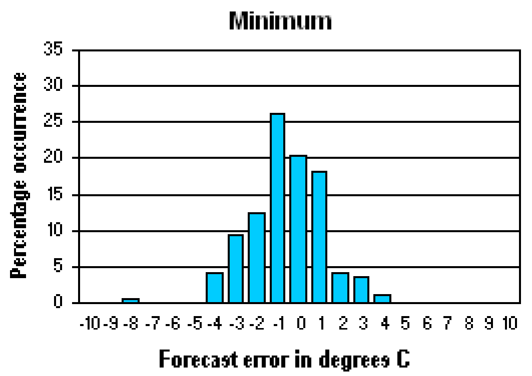
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

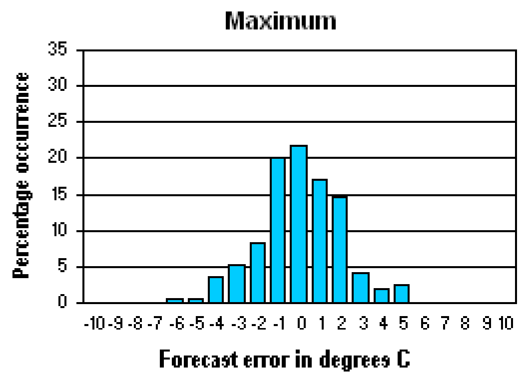




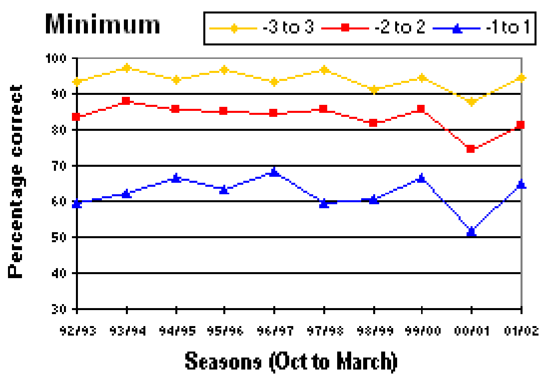
Durban:

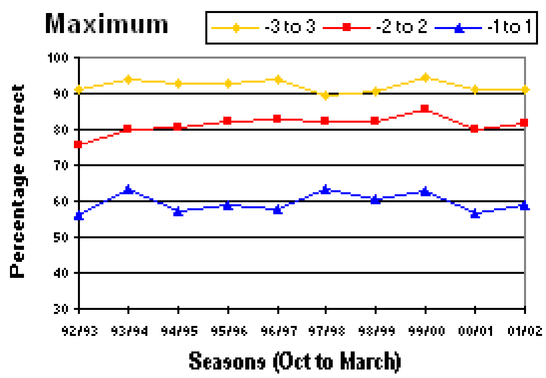
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





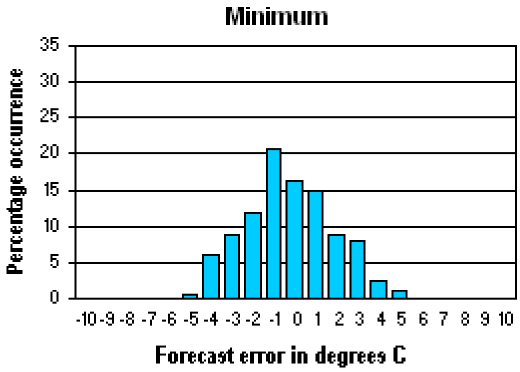
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

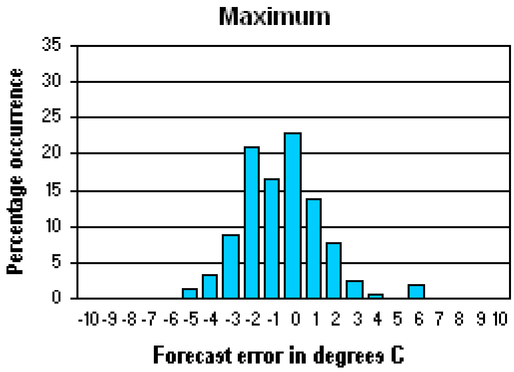




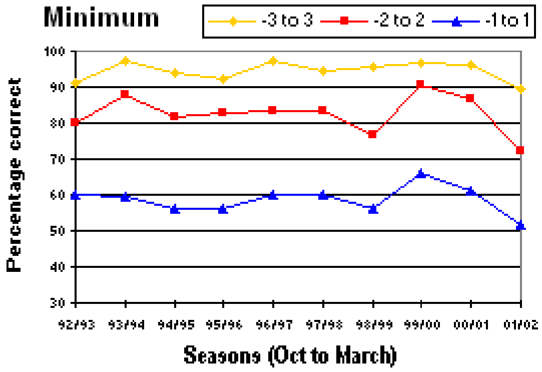
Johannesburg:

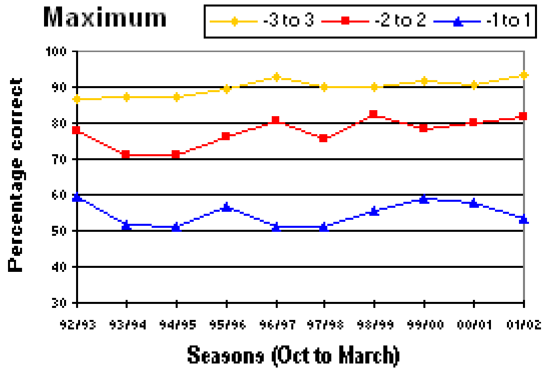
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





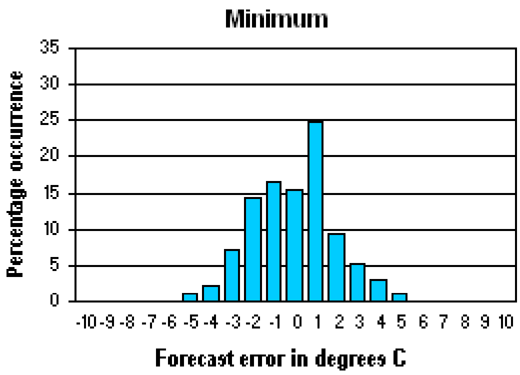
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

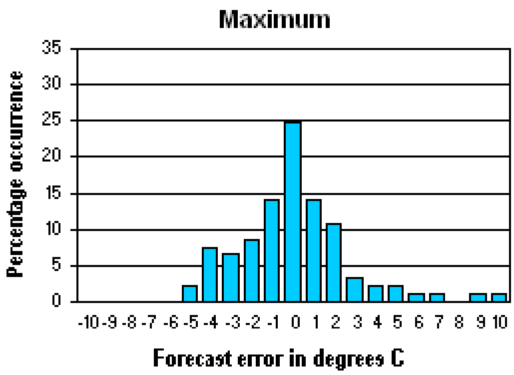




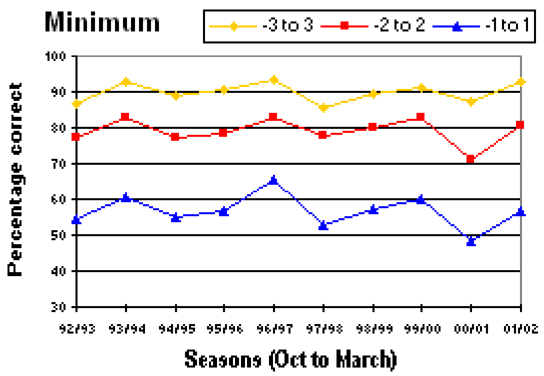
Klerksdorp:

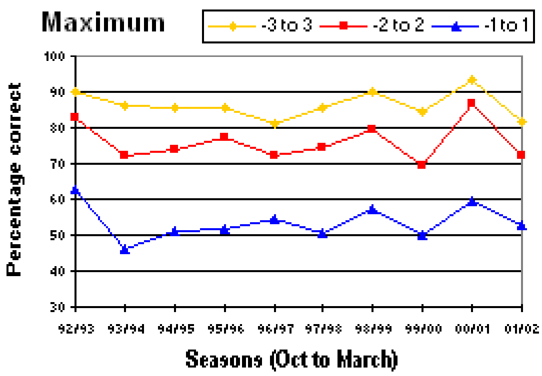
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





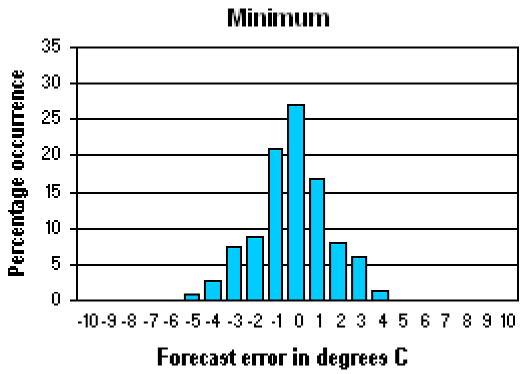
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

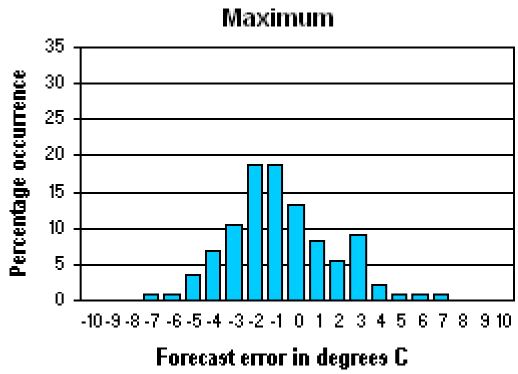




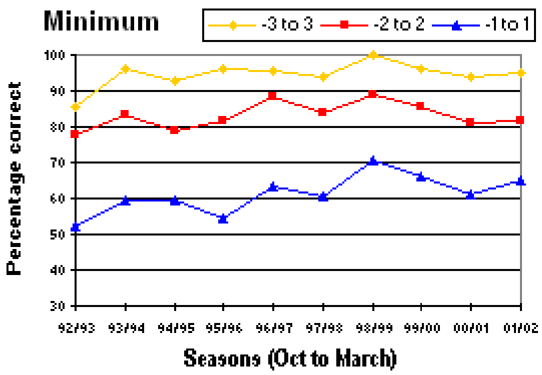
Nelspruit:

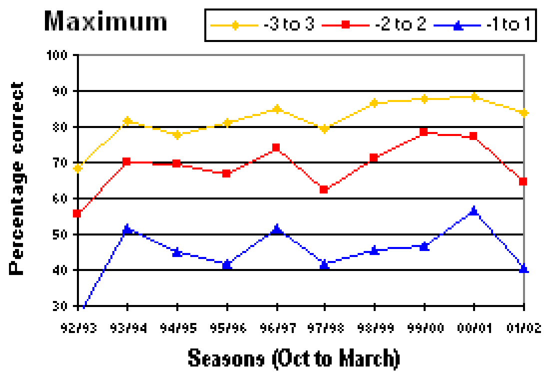
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





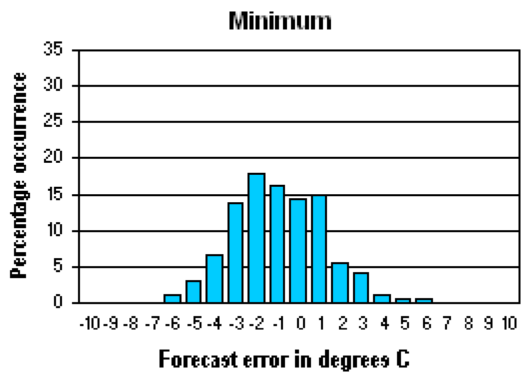
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

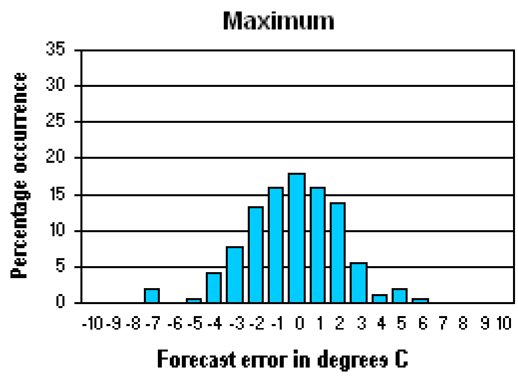




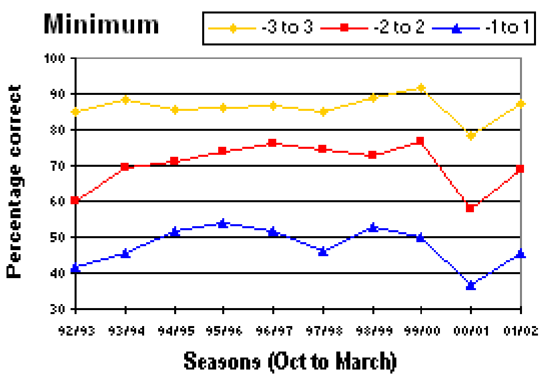
Port Elizabeth:

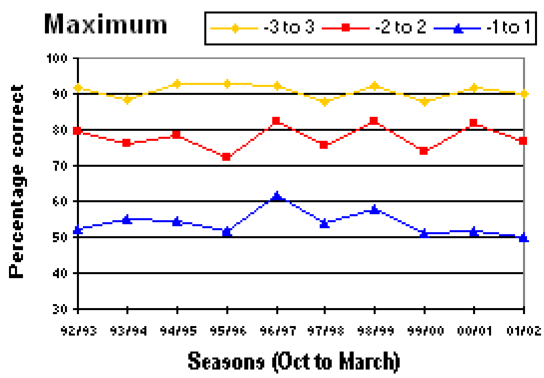
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





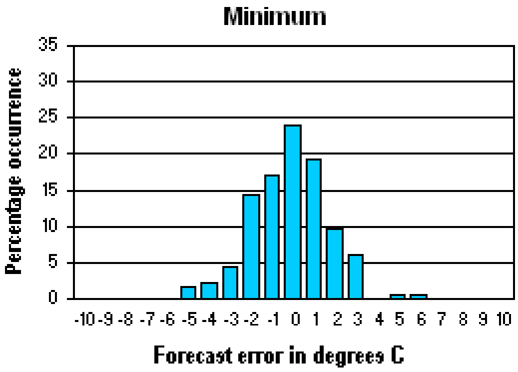
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

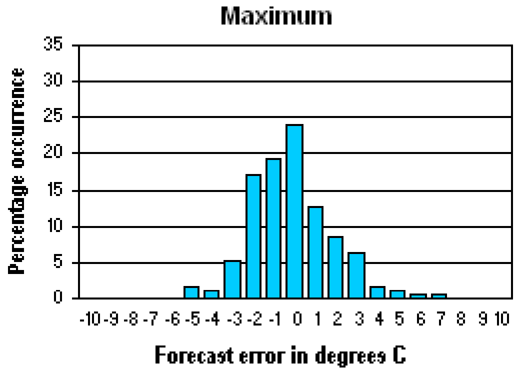




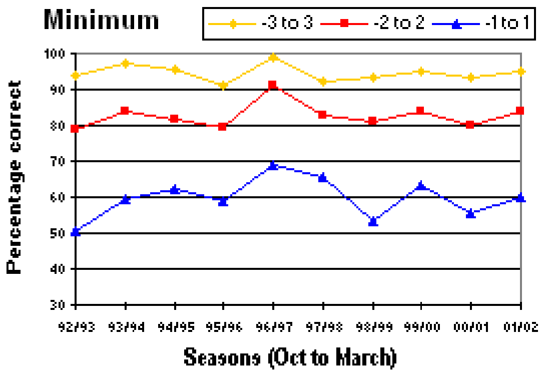
Polokwane:

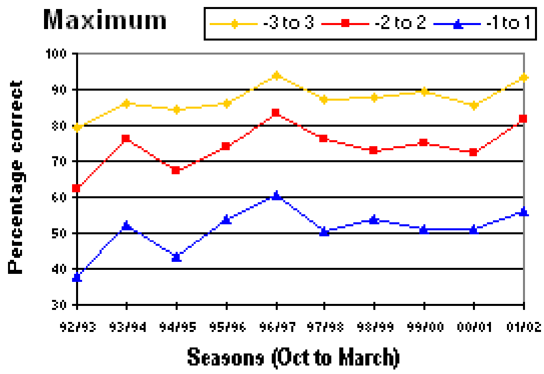
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





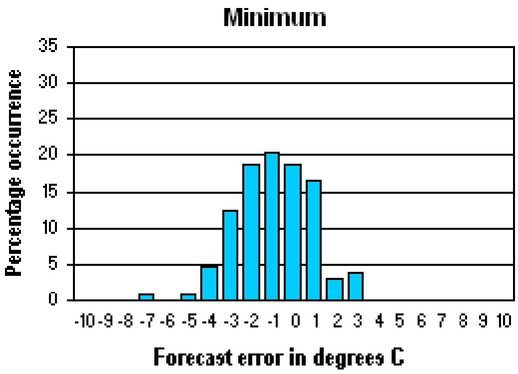
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

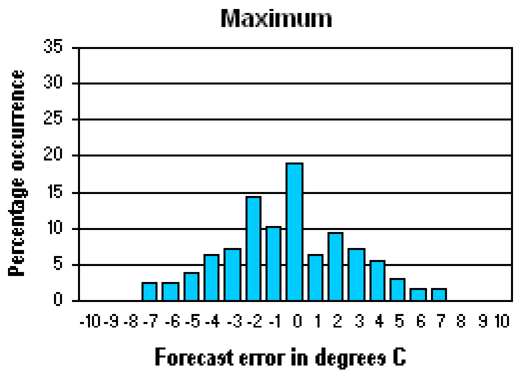




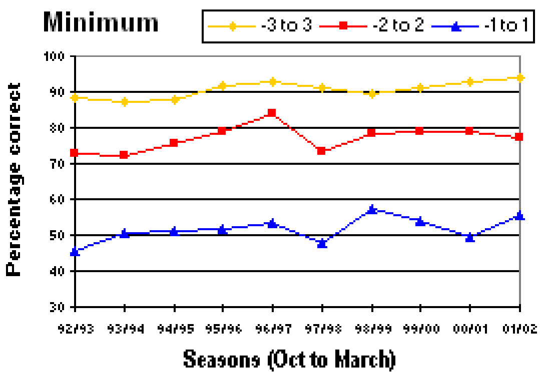
Pietermaritzburg:

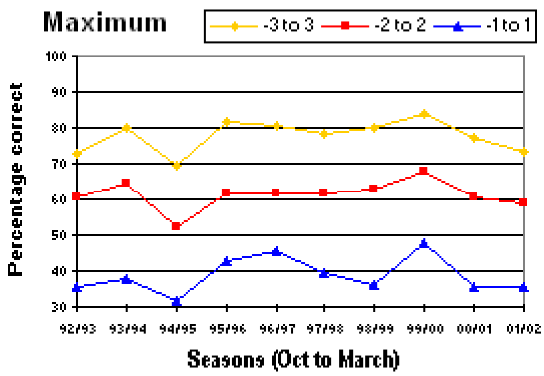
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





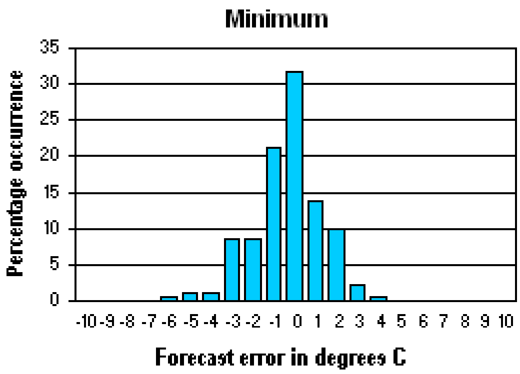
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

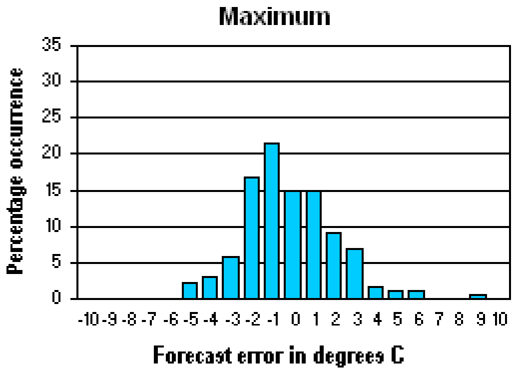




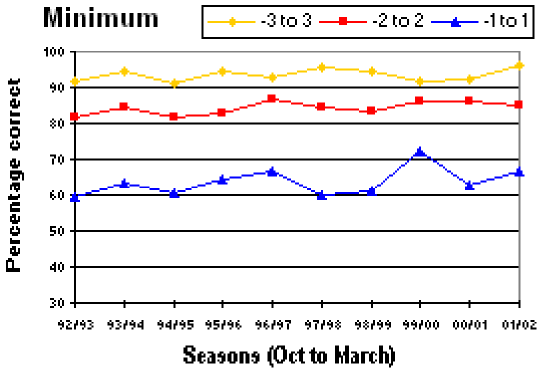
Pretoria:

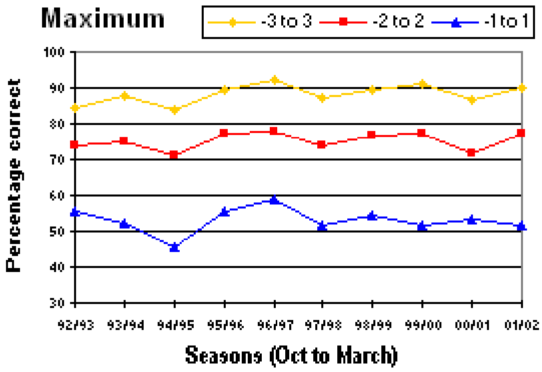
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





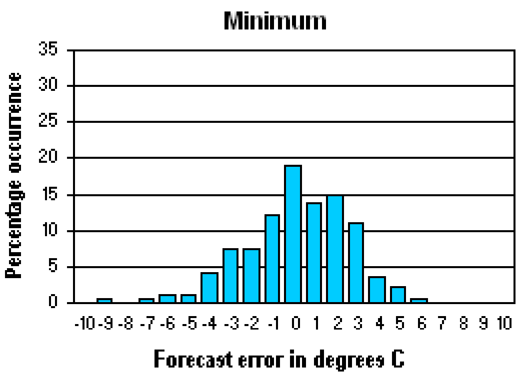
Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:

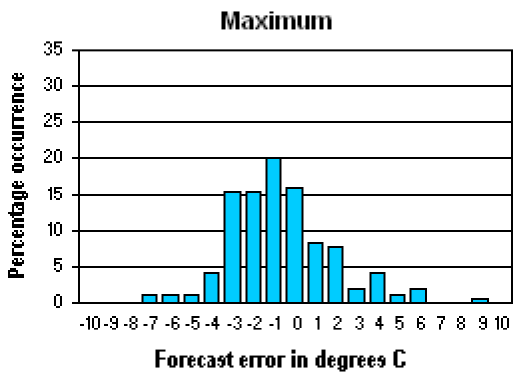




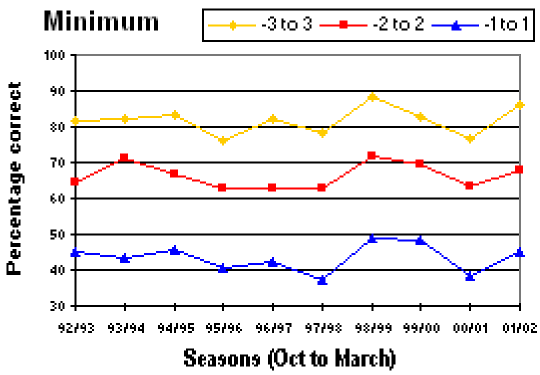
Upington:

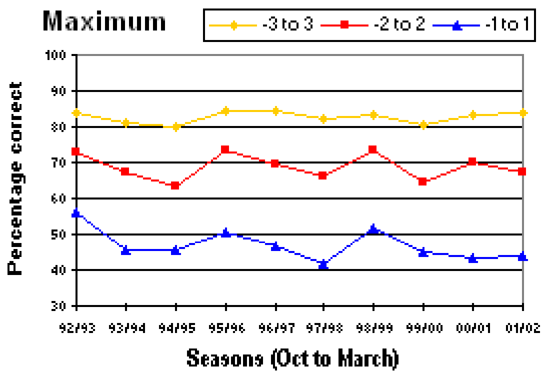
Frequency distribution of forecast errors for the season 1 Oct 2001 to 31 Mar 2002:





Percentage of forecasts within 3, 2 and 1 ºC correct for the past 10 seasons (1 Oct to 31 Mar) 1992 to 2002:





Verification of Eta Precipitation forecasts for the summer rainfall season October 2001 to March 2002

The precipitation forecast generated by a numerical weather prediction model is considered to be one of the potentially most useful products the model can provide. The SAWS currently runs a Limited Area Model (LAM) referred as the Eta model. Case studies on heavy rainfall events over South Africa involving the Eta model found that the model is generally well equipped to identify the geographical distribution of heavy rainfall but can underestimate the amount of rainfall by 50%. (De Coning et al., 1998, Dyson and Van Heerden, 2001).

Eta model precipitation forecasts are verified against rainfall data obtained from the South African Weather Service rain gauges. At present only those stations of which the rainfall totals are available in real time are used. The SAWS Eta model is initialized twice daily at 00 UTC and 12 UTC. It covers Southern Africa and surrounding oceans from approximately 9°S to 48°S and 13°W to 53°E. The model has a horizontal resolution of half a degree (approximately 48 km) with 38 vertical layers. The 24-hour Eta rainfall forecast, which corresponds with the rain gauge observation period, is evaluated for both the 00 UTC and 12 UTC model runs. Eta provides two precipitation forecasts, namely total precipitation and convective precipitation. The rain gauge data used in this verification originated from SAWS synoptic and daily reporting rainfall stations. There are approximately 140 synoptic stations in South Africa. The synoptic stations are distributed over the entire country but with a higher concentration in Gauteng, Kwazulu Natal and the South Western Cape. The daily reporting rainfall stations are not available for verification at present but will be incorporated into the results.

2x2 contingency tables (Yes/No) were calculated for 7 different thresholds. The thresholds are 0.1mm, 1mm, 2mm, 4mm, 8mm, 16 mm, and 25 mm per 24 hours. Six scores are calculated from these tables, namely, the Frequency Bias Index (FBI), the Equitable Threat Score (ETS), the True Skill Score (TSS), the False Alarm Ratio (FAR), the Probability of Detection (POD), and the Hit Rate (H). A definition of each score can be found in the Appendix. The yes/no nature for each category are then tested at each point or station, which will explain the model's behavior at different thresholds. For example, at low thresholds, it explains the model's behavior at light rain events or the area coverage of rain (Ebert and McBride, 2000). At higher thresholds it explains the model's behavior at peak values.

In Figure 2 to Figure 7, 5 synoptic stations were chosen in different climate regions and show the scores calculated for the 00 UTC model run. The Cape Town station is located in a winter-rain region, which is characterized by temperate wet winters, and warm dry summers. The Durban station is located in a moist subtropical region, which are characterized by rainfall throughout the year, with high annual mean values. The Nelspruit station again is located in a tropical region with one cool wet season. The Upington station is located in a desert region characterized by a dry climate with occasional showers. The Kimberley station is located in a sahel region, which is characterized by a short wet season of 3 months.

At Cape Town the model seems to perform well at low thresholds, looking at all six scores, but at higher thresholds the skill does deteriorate. From the TSS and POD score for Cape Town it is clear that the model does relatively well in forecasting the occurrence of rain at high thresholds, but the model also tends to overforecast rainfall (FBI) and makes too much false alarms (FAR), which causes a draw back in the overall skill of the model at high thresholds.

At Durban the model performs relatively well for low thresholds. For higher thresholds there is a moderate decrease in the overall skill (ETS) and the ability of the model to forecast the occurrence of rainfall (TSS). There is also an increase in false alarms, but less obvious than in the case of Cape Town.

At Nelspruit, the accuracy of the forecasts (H) is quite low, relative to the other stations. The ability of the model to forecast the occurrence of rainfall (TSS) also seems to be lower than for the other stations. The answer for this is not clear, because the FAR and FBI for Nelspruit doesn't show a tendency of the model to overforecast more or make more false alarms relative to the other stations. High POD values contradict with the low TSS values which doesn't give a clear conclusion on the skill of the forecast.

At Upington, it is clear that there is more model skill in forecasting higher thresholds, looking at the TSS and POD. The reason for this might be that Upington experience occasional showers and when such occasional shower does occur the model does fairly well in detecting it.

Scores for Kimberley seems to be relatively constant over the range of thresholds, except at the lowest thresholds, below 4 mm, where an increase in skill and accuracy are present.

Overall, wetter regions (blue lines) seem to have lower FAR than the drier regions (red lines, green line for summer only). Looking at the Hit Rate, the wetter regions show lower values, decreasing faster with higher thresholds than for the drier regions.

Contributed by Estelle Marx

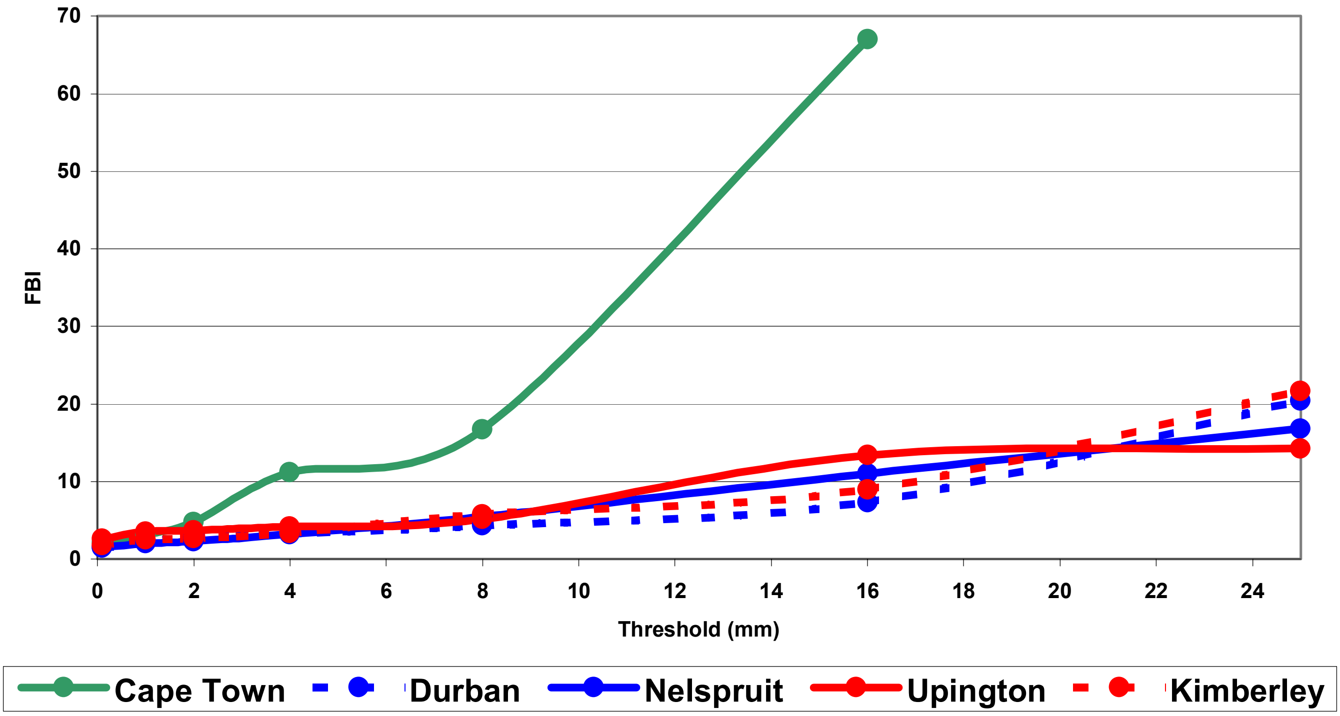


Figure 1 The Frequency Bias Index for 5 synoptic stations from Oct 2001 to March 2002.

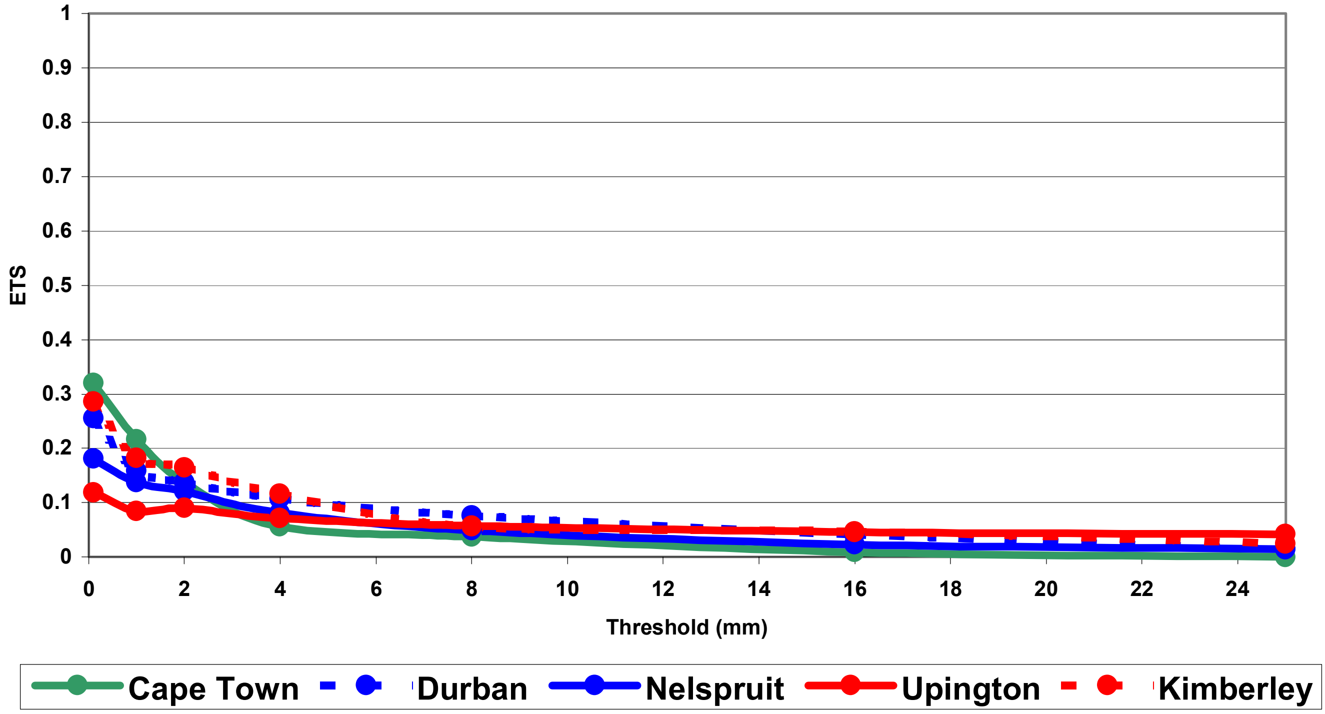


Figure 2 The Equitable Threat Score for 5 synoptic stations from Oct 2001 to March 2002.

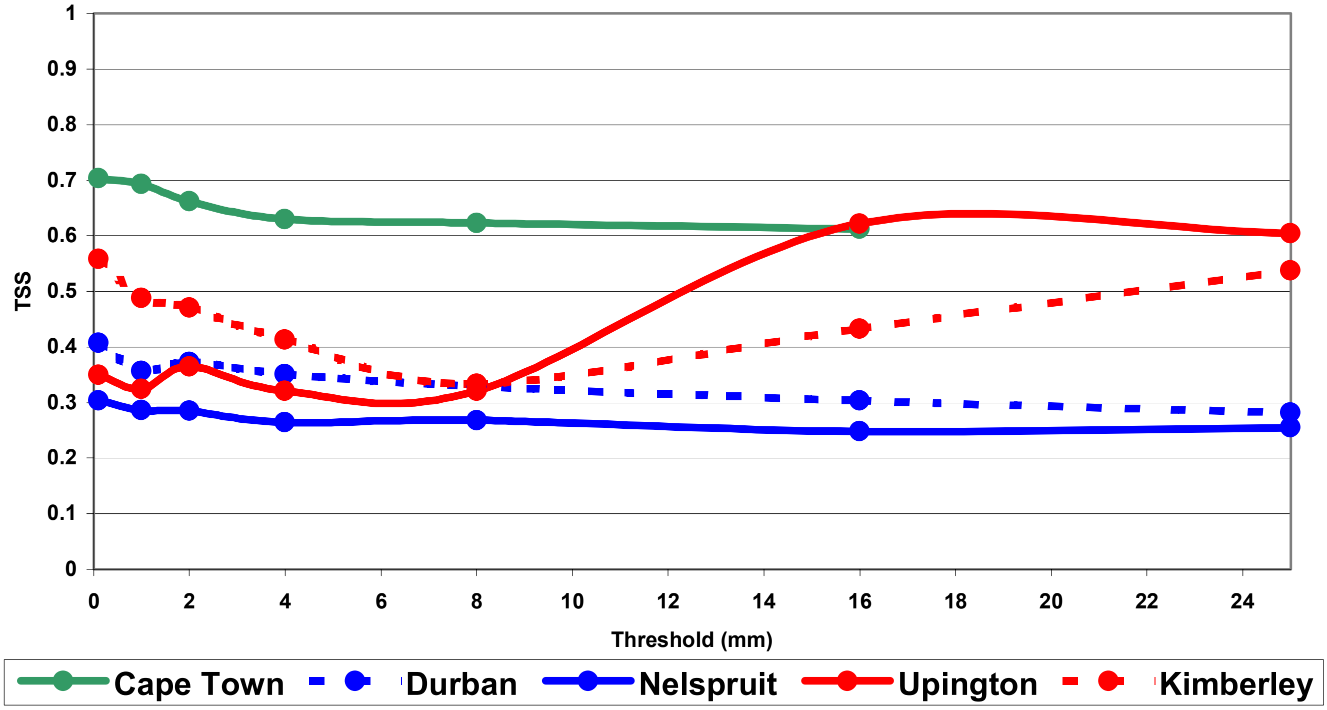


Figure 3 The True Skill Score for 5 synoptic stations from Oct 2001 to March 2002.

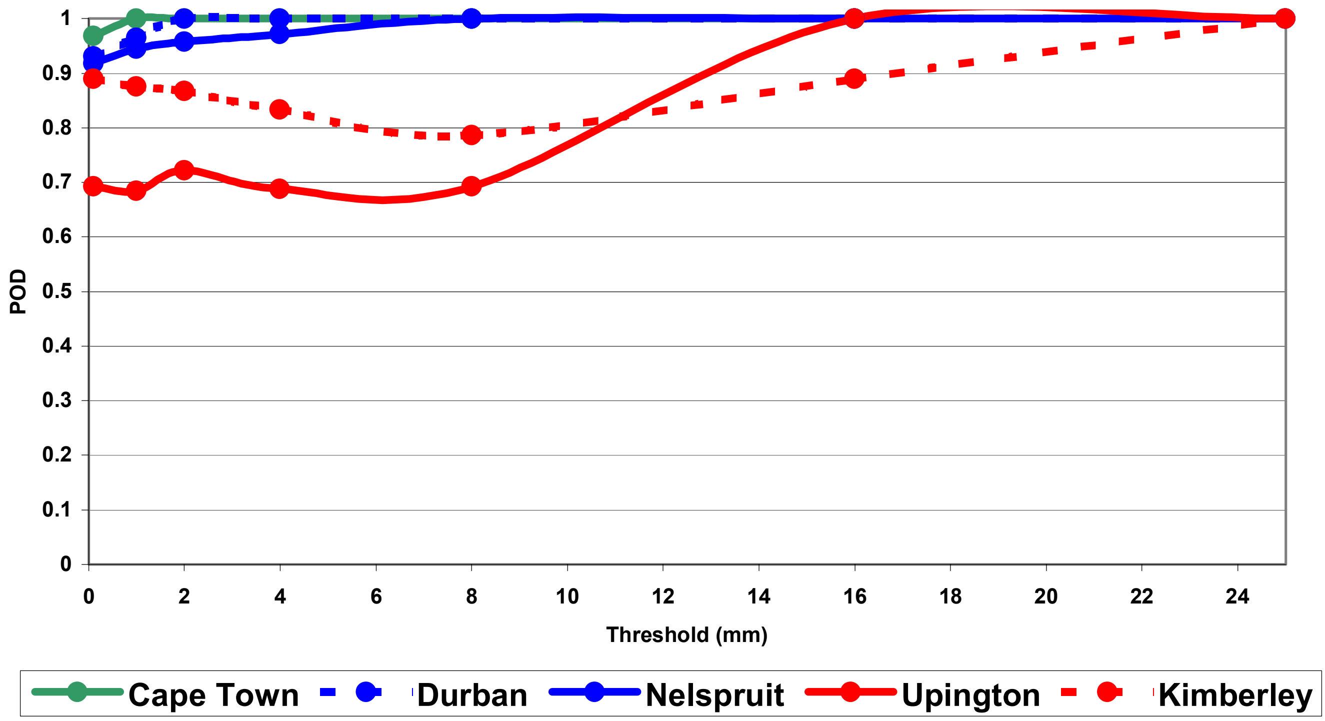


Figure 4 The Probability of Detection for 5 synoptic stations from Oct 2001 to March 2002.

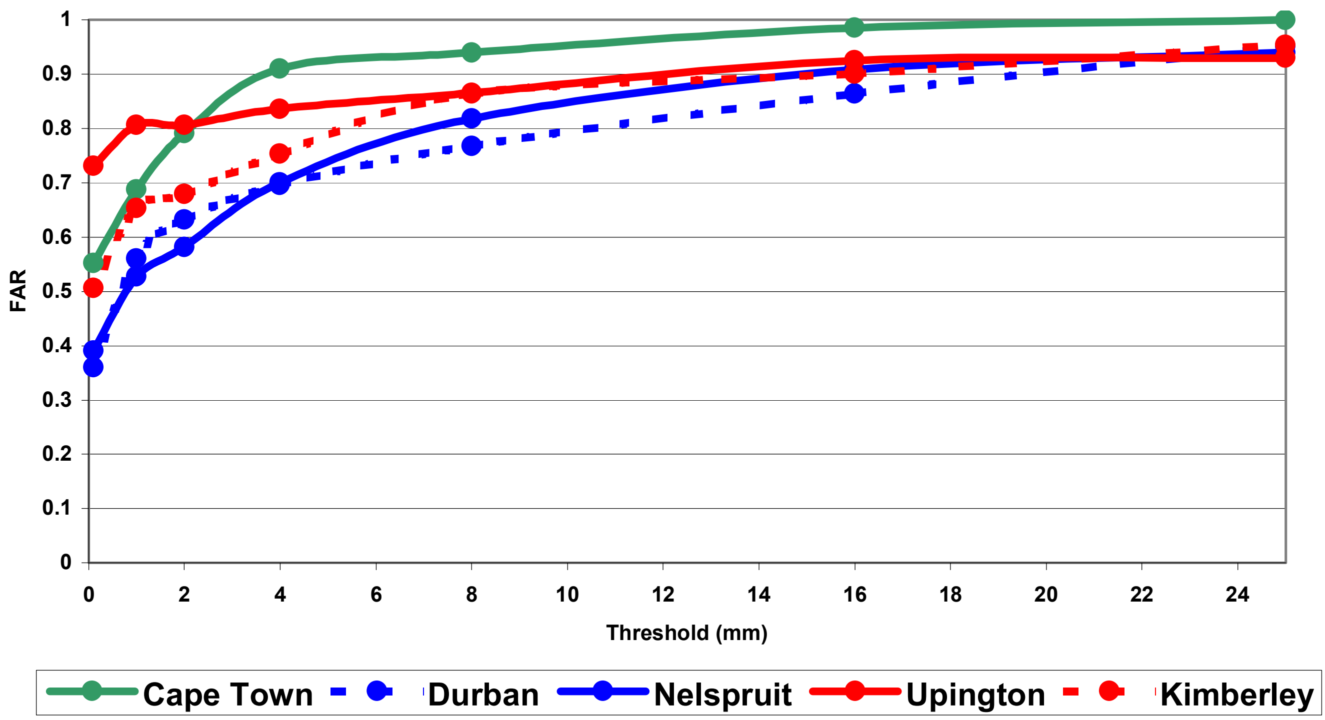


Figure 5 The False Alarm Ratio for 5 synoptic stations from Oct 2001 to March 2002.

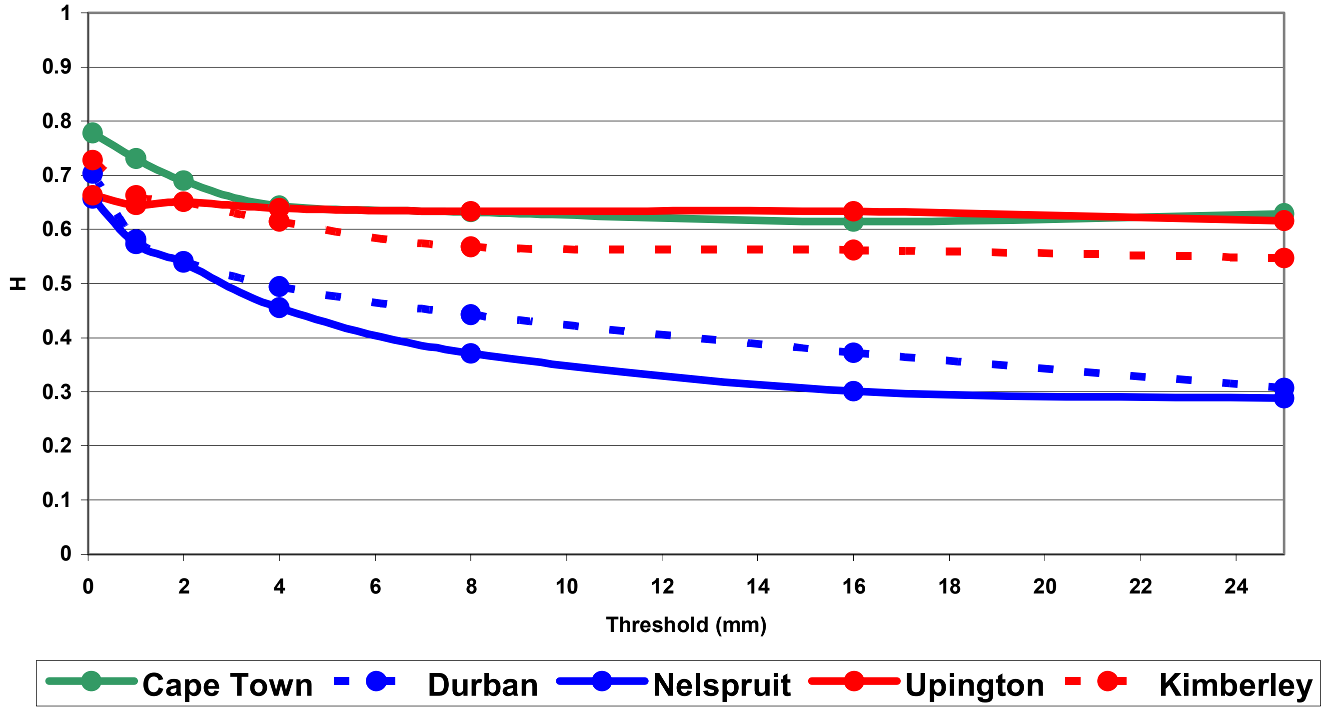


Figure 6 The Hit score for 5 synoptic stations from Oct 2001 to March 2002.

Appendices

A: Categorical Verification Scores

Frequency Bias Index (FBI)

- is a measure of the relative frequency of occurrence in the forecast with respect to observations. The closer to unity the better the forecast is. This means for FBI larger than 1, the occurrence of rainfall at a certain threshold is overforecasted.

Equitable Threat Score (ETS)

- is a good estimate for overall forecast skill. The equitable threat score can vary from a small negative number to 1.0, where 1.0 represents a perfect forecast. This is basically the ratio of the correct forecast area to the total area of the forecast and observed precipitation.

True Skill Score (TSS)

- is a measure to determine the ability of the model to correctly predict the occurrence of rainfall, regardless of the particular frequency or distribution of observed rain. A perfect TSS score will be equal to 1.0.

Probability of Detection (POD)

- measures the success of the forecast in correctly predicting the occurrence of rain. POD = 1 is a perfect score.

False Alarm Ration (FAR)

- measures the fraction of rainfall forecasts, which were actually not observed. FAR = 0 is a perfect score and can range to 1.

Hit Score (H)

- is a measure of forecast accuracy, which is the ratio of correct forecast to the total number of forecasts. Again a perfect score is reach at unity.