MODEL OUTPUT STATISTICS APPLIED TO MULTI-MODEL ENSEMBLE LONG-RANGE FORECASTS OVER SOUTH AFRICA

Report to the
Water Research Commission

by

Willem A. Landman[1], Francois Engelbrecht[2], Asmerom Beraki[1],
Christien Engelbrecht[3], Maluta Mbedzi[4], Tracey Gill[1] &
Lucky Ntsangwane[1]

[1]South African Weather Service
[2]University of Pretoria
[3]Agricultural Research Council
[4]City Council of Johannesburg

WRC Report No 1492/1/08
ISBN 978-1-77005-799-9

JANUARY 2009

DISCLAIMER

EXECUTIVE SUMMARY

1. Introduction and Objectives

Improved seasonal forecasts would greatly assist managers in the fields of water resources and agriculture in adapting to climate variability and, more particularly, in preparing adequately for seasons characterized by climatic extremes. The main emphasis of this project was, therefore, to develop a multi-model forecasting system for South Africa that would be able to produce objective operational seasonal rainfall, streamflow and global sea-surface temperature (SST) forecasts skillfully.

The forecasting systems developed for rainfall and streamflow would comprise of a number of tiers including the generation of output from general circulation models (GCMs), the statistical downscaling of these output to station and catchment level and forecast combination. The statistical downscaling techniques demonstrated in WRC Report No. 1334/1/06 would be used to translate large-scale information provided by the global models to station and catchment level. Downscaling would be applied to three GCMs and the downscaled results combined using a selection of linear schemes of varying complexities. Since the occurrence of tropical cyclones over the south-western Indian Ocean is an additional factor affecting rainfall variability over southern Africa, recent efforts to predict these tropical disturbances would also have to be considered. For SST predictions three coupled model forecasts would be combined using linear methods and, thereafter, skill assessment performed.

The specific objectives of the project were:

1) To investigate the operational predictability of seasonal to interannual rainfall and streamflow over South Africa through the use of multi-model ensembles;
2) To test various multi-model linear combination schemes;
3) To investigate the operational predictability of seasonal to interannual occurrence of tropical cyclones over the south-western Indian Ocean through the use of multi-model ensembles;
4) To investigate the operational predictability of multi-model global sea-surface temperature fields; and,
5) To set up an operational multi-model prediction system at the South African Weather Service (SAWS) to the benefit of the end-users of seasonal forecast products.

2. Results and Conclusions

1. Even simple model combinations improve on the forecast skill associated with individual models. However, this is only true under certain conditions, which include the use of long training periods. Furthermore, only combining the best forecast models will improve on forecast skill. The geographical distribution of skill shows that during the season with the highest climate predictability, most of the skill is

found over the north-east and central-western regions of South Africa. Moreover, seasonal rainfall predictability seems to be largely confined to mid-summer periods which coincide with El Niño – Southern Oscillation (ENSO) seasons.

2. There is potential in predicting tracks of tropical vortices through the application of detection and tracking algorithms applied to physical model output. Operational tropical cyclone prediction on a seasonal time scale is non-existent in South Africa, but needs to be addressed. Tropical cyclone landfall significantly influences streamflow. Results obtained show that streamflows measured in rivers are potentially predictable.
3. High prediction skill for SSTs is found over large areas and the best multi-model skill is found when the SST forecasts from the different models are simply averaged.

3. Recommendations

1. Model performance evaluations presented here are based mostly on deterministic model simulations and forecasts. Multi-model ensemble systems that consider all the available members from the individual systems need to be developed.
2. The simulation and forecast data sets of the operational GCMs being run at the SAWS and University of Pretoria (UP) need to be significantly expanded to include more years, more ensemble members and more forecast lead-times.
3. Combinations schemes other than the linear methods presented here need to be tested and compared to simply averaging of the forecasts.
4. The prediction skill of SSTs over certain ocean areas such as the area south-east of Madagascar needs to be further improved in order to improve on GCM-based forecasts of rainfall over the region.
5. Modellers need to investigate whether to either use the best single-field global SST forecast (e.g., a multi-model forecast), or to use the individual SST forecasts to force atmospheric GCMs and therefore include into the forecasting system uncertainties associated with boundary conditions.

ARCHIVING OF DATA

The CCAM (5 members, 1979 to 2003) data are archived at the University of Pretoria. The SAWS also has a backup archive and has archived ECHAM4.5 data (6 members, 1979 to 2003). Both these sets were generated by forcing the GCM with observed SST anomalies. These data sets are available for research purposes. To get access to these data sets, contact Dr Willem Landman at the South African Weather Service (tel: 012-367-6003, email: willem.landman@weathersa.co.za).

KNOWLEDGE DISSEMINATION

1. The following papers are based on the results presented in the report:
    a. Landman, W.A., 2007: The influence of ENSO on operational rainfall forecast skill for South Africa. *CLIVAR Exchanges*, **12**, 26-28.
    b. Landman, W.A., 2008: ENSO and multi-model forecast skill for mid-summer rainfall over South Africa. Submitted to the *International Journal of Climatology*.
    c. Landman, W.A., F. Engelbrecht, and A. Beraki, 2008: Model output statistics applied to multi-model ensemble long-range forecasts over South Africa. In preparation for the *International Journal of Climatology*.

2. The following paper was presented at a national conference:
    a. Landman, W.A., and N. Le Roux, 2007: The influence of ENSO on operational rainfall forecast skill for South Africa. *SASAS, 13-14 September 2007*.

ACKNOWLEDGEMENTS

1. Dr. G. Green for his interest and support during the project.
2. The International Research Institute for Climate and Society (IRI) for making available their archived ECHAM4.5 data.
3. School of Bioresources Engineering and Environmental Hydrology at the University of KwaZulu-Natal for supplying the streamflow data.
4. The authors also appreciate the guidance of the Steering Committee: Dr. G. Green, Mr. E. Poolman, Prof. B. Hewitson, Prof. Johan van Heerden, and Prof. C. J. de W. Rautenbach.

TABLE OF CONTENTS

LIST OF FIGURES

Figure 3.7. Single model vs. multi-model simulation skill for the MAM season. The three single models are ECHAM4.5 (ECHAM), CCAM (CCAM) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + CCAM + SST-rainfall (ECS), ECHAM4.5 + CCAM (EC), ECHAM4.5 + SST-rainfall (ES) and CCAM + SST-rainfall (CS). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

Figure 3.8. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the MAM season. The three single models are ECHAM4.5 (ECHAM), UKMO (UKMO) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + UKMO + SST-rainfall (EUS), ECHAM4.5 + UKMO (EU), ECHAM4.5 + SST-rainfall (ES) and UKMO + SST-rainfall (US). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

Figure 3.9. Single model vs. multi-model simulation skill for the JJA season. The three single models are ECHAM4.5 (ECHAM), CCAM (CCAM) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + CCAM + SST-rainfall (ECS), ECHAM4.5 + CCAM (EC), ECHAM4.5 + SST-rainfall (ES) and CCAM + SST-rainfall (CS). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

Figure 3.10. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the JJA season. The three single models are ECHAM4.5 (ECHAM), UKMO (UKMO) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + UKMO + SST-rainfall (EUS), ECHAM4.5 + UKMO (EU), ECHAM4.5 + SST-rainfall (ES) and UKMO + SST-rainfall (US). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

Figure 3.11. Spearman rank correlations from the best multi-model for SON simulated (top panel) and forecast (bottom panel) rainfall. The multi-model simulation consists of the ECHAM4.5 and SST-rainfall models, while the multi-model forecast consists of the UKMO and SST-rainfall models. Equal weighting is used to combine the models. Negative correlations are masked out.

Figure 3.12. Spearman rank correlations (top panel) and MSESS (bottom panel) of the best multi-model for DJF simulated rainfall. The multi-model simulation consists of the ECHAM4.5 and CCAM models. Negative values are masked out.

Figure 3.13. Spearman rank correlations (top panel) and MSESS (bottom panel) of the best multi-model for DJF forecast rainfall. The multi-model simulation consists of the ECHAM4.5 and UKMO models. Negative values are masked out.

Figure 3.14. Spearman rank correlations from the best multi-model for MAM simulated (top panel) and forecast (bottom panel) rainfall. The multi-model simulation consists of the ECHAM4.5 and SST-rainfall models, while the

Niña (L) seasons as defined by the Climate Prediction Centre, USA, are also shown. Correlations between multi-model forecasts and observed values are shown on the bottom left hand side of the figure. The years on the x-axis refer to the December months of the DJF seasons.

Figure 4.2. Mean squared error skill score (MSESS) values of the best single- and best two multi-model systems, considering all of the 34 years, El Niño together with La Niña years (26), only El Niño years (14), only La Niña years (12), and only neutral years (8). Negative MSESS values imply that the performance associated with the model forecasts is worse than when using climatology as the forecasts.

Figure 5.1. Observed tracks of tropical cyclones during the La Niña seasons of 1995/96 (top panel) and 1999/2000 (bottom panel).

Figure 5.2. An example of tracks of tropical cyclone-like vortices detected when applying the objective procedure for tracking cyclones in the ECHAM4.5-RegCM3 output.

Figure 5.3. Tracks of observed (black lines) and predicted tropical systems during DJF 2007/08.

Figure 6.1. Forecast skill (Spearman rank correlations) of the ECMWF (top panel), Météo-France (middle panel) and UKMO (bottom panel) for model initialized in November to produce DJF SST forecasts. The black ellipse identifies the Indian Ocean south-east of Madagascar.

Figure 6.2. Forecast skill (Spearman rank correlations) of the multi-model system. Forecasts of the ECMWF, Météo-France and UKMO coupled models are combined with CCA. Model initialized is in November to produce DJF SST forecasts. The black ellipse identifies the Indian Ocean south-east of Madagascar.

Figure 6.3. Forecast skill (MSESS) associated with two multi-model combination schemes and three individual models for the Nino3.4 region and the equatorial domains of the Indian and Atlantic Oceans. Model initialized in November to produce DJF SST forecasts. C_MM: multi-model combination using CCA; E_MM: multi-model combination using equal weighting.

Figure 6.4. Forecast and observed index time series of the Nino3.4 and the equatorial domains of the Indian and Atlantic Oceans. The forecast time series comprise of forecasts produced by a CCA-multi-model, equal-weights-multi-model, ECMWF, Météo-France and UKMO coupled models.

Figure 6.5. Correlations (top and middle panel) between forecast and observed SST indices of the three equatorial domains. The middle panel shows correlations after detrending the data, and the bottom panel shows the correlation differences between the values of the top and middle panel. Forecasts are for DJF initialized in November. CCA_MM: multi-model

combination using CCA; EqW_MM: multi-model combination using equal weighting.

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. Background

The scientific basis for doing seasonal forecasting originates from the observation that slowly evolving sea-surface temperature (SST) anomalies influence seasonal-mean weather conditions (Palmer and Anderson, 1994). Therefore, estimation of the evolution of SST anomalies, which are often relatively predictable, and subsequently employing them in atmospheric GCMs, potentially provides means of generating forecasts of seasonal-average weather (Graham *et al*., 2000). Although GCMs, commonly configured with an effective resolution of 200-300 km, have demonstrated skill at global or even continental scale, they are unable to represent local sub-grid features, subsequently producing rainfall over southern Africa that is typically overestimated (Joubert and Hewitson, 1997; Mason and Joubert, 1997). Also, the representation of rainfall at high latitudes is complex and often not well estimated (Graham *et al*., 2000; Goddard and Mason, 2002). Such systematic biases have created the need to downscale or recalibrate GCM simulations to regional level over South Africa. Semi-empirical relationships exist between observed large-scale circulation and rainfall, and assuming that these relationships are valid under future climate conditions and also that the large-scale structure and variability is well characterized by GCMs, mathematical equations can be constructed to predict local precipitation from the simulated large-scale circulation (Wilby and Wigley, 1997). Recently, empirical remapping of GCM fields to regional rainfall has been demonstrated successfully over southern Africa (Landman and Goddard, 2002, 2003, 2005; Landman *et al*., 2001).

The inherent variability of the atmosphere requires seasonal climate simulations to be expressed probabilistically. Probabilistic forecasts are made possible through the proper use of GCM ensembles since ensemble forecasting is a feasible method to estimate the probability distribution of atmospheric states (Branković and Palmer, 2000). In addition, errors in the initial conditions as well as deficiencies in the parameterizations and systematic or regime-dependent model errors can be to a large part accounted for through ensemble forecasting (Evans *et al*., 2000). Moreover, there is inevitable growth in differences between forecasts started from very slightly different initial conditions suggesting that there is no single valid solution but rather a range of possible solutions (Tracton and Kalnay, 1993). Information contained in the distribution of the ensemble members can subsequently be used to represent forecast probabilities by calculating the percentage of ensemble members that fall within a particular category (e.g. below-normal, near-normal and above-normal). Similarly, forecast probabilities can be produced indicating the percentage of ensemble members in the upper or lower extremes, i.e. 15^{th} percentiles (Mason *et al*., 1999).

There are advantages in combining ensemble members of a number of GCMs into a multi-model ensemble since GCMs differ in their parameterizations and therefore differ in their performance under different conditions (Krishnamarti *et al*., 2000). Using a suite of several GCMs not only increases the effective ensemble size; it also leads to probabilistic simulations that are skilful over a greater portion of the region and a greater portion of the time series. Multi-model ensembles are nearly always better than any of the individual ensembles (Dirmeyer *et al*., 2003; Landman and Goddard, 2003;, Doblas-Reyes *et al*., 2000; Krishnamurti *et al*., 2000). The benefits from combining ensembles are a result of the inclusion of complementary predictive information since the forecast scheme is able to extract useful information from the results of individual models from local regions where their skill is higher (Krishnamurti *et al*., 2000). In fact, the most striking benefit obtained from multi-model ensembles is the skill-filtering property in regions or seasons when the performance of the individual models varies widely (Graham *et al*., 2000). Moreover, increased ensemble size leads to further benefits (Brown and Murphy, 1996), but the multi-model approach is only beneficial if the individual models produce independent skilful information (Graham *et al*., 2000). Multi-model ensembles have also been found to be improving forecast performance of tropical cyclones (Vijaya Kumar *et al*., 2003).

A number of ensemble combining algorithms exists. The most simple of these is the unweighted combination of ensembles from different models (Graham *et al*., 2000; Mason and Mimmack, 2002). Combining forecasts this way has been shown to improve on skill levels of individual model forecasts for southern African summer rainfall (Landman and Goddard, 2003). The improvements over the individual ensemble systems are attributed to the collective information of all the models used in the mean of probabilities algorithm.

The resulting first study in South Africa to assess the feasibility of producing skillful multi-model forecasts for southern African seasonal rainfall was presented by Landman and Goddard (2003). Summer (December-January-February; DJF) rainfall forecast skill was assessed by combining simulations from a number of GCMs after first statistically downscaling them (using model output statistics; MOS) to homogeneous rainfall regions over a large area of southern Africa, including South Africa, Namibia and Botswana. The GCMs were forced with simultaneously observed SST anomalies, and the output was obtained from the website of the International Research Institute for Climate and Society (IRI; http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD). The model combination was done by simply averaging the downscaled results. The ranked probability skill score (RPSS; Wilks, 2006) was used as the skill measure of the probabilistic rainfall simulations. Figure 1 shows the DJF rainfall probabilistic simulation skill (RPSS) of the multi-model GCM-MOS system comprising of five atmospheric GCMs. The solid line in the figure shows the skill of the best GCM-MOS (ECHAM4.5). For the most part, combining models improved on the best model's performance.

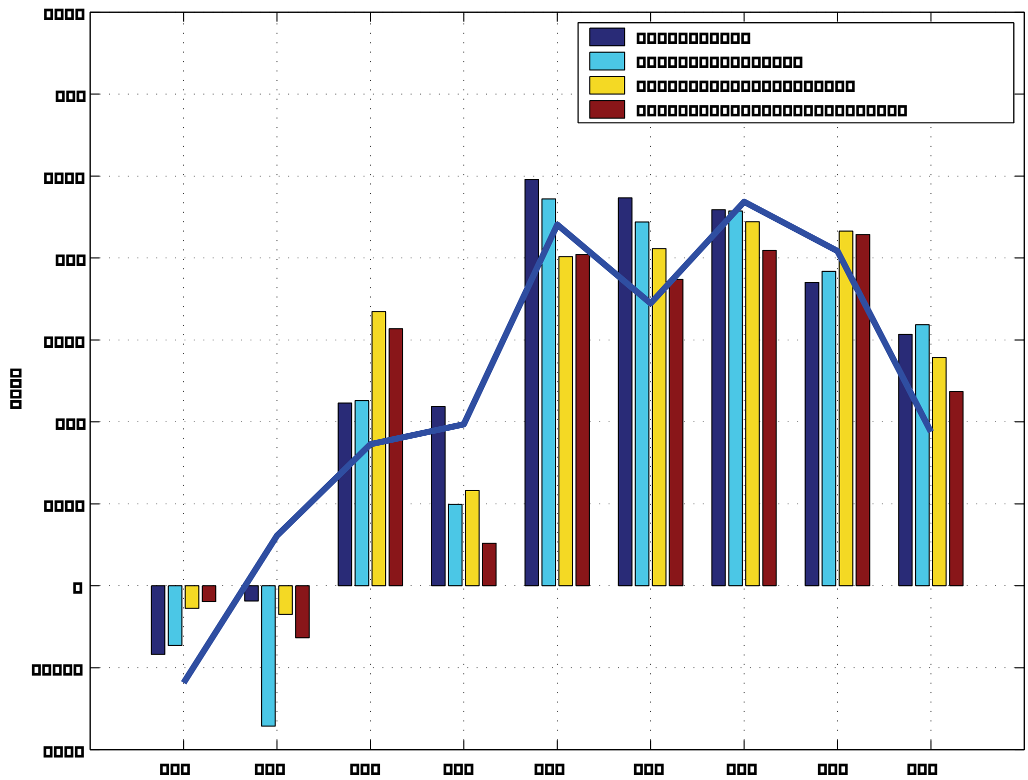

Figure 1.1. RPSS values of multi-model GCM-MOS simulations of DJF rainfall (bars) and RPSS values of the best GCM-MOS model (blue line). (SWC: south-western Cape; SCO: south coast; TRA: Transkei; KZC: KwaZulu-Natal coast; LOW: Lowveld; NEI: north-eastern interior; CIN: central interior; WIN: western interior; NWB: northern Namibia / western Botswana).

The specific objectives of the current, follow-on research project were to:

1) To investigate the operational predictability of seasonal to interannual rainfall and streamflow over South Africa through the use of multi-model ensembles;
2) To test various multi-model linear combination schemes;
3) To investigate the operational predictability of seasonal to interannual occurrence of tropical cyclones over the south-western Indian Ocean through the use of multi-model ensembles;
4) To investigate the operational predictability of multi-model global sea-surface temperature fields; and,
5) To set up an operational multi-model prediction system at the South African Weather Service (SAWS) to the benefit of the end-users of seasonal forecast products.

1.2. Organization of the report

Chapter 1 provides an introduction to the concept of multi-model forecasting and the downscaling work that has been done at the South African Weather Service (SAWS) and states the objectives of the current project. Chapter 2 describes the data used and the simulation and forecast methods, including the descriptions of the GCMs, SST-rainfall statistical model and the empirical downscaling technique. Rainfall forecast model combinations are addressed in Chapter 3, and in Chapter 4 the influence of ENSO on multi-model forecast skill is addressed. Chapter 5 shows the results of detecting and tracking tropical vortices in model data. Chapter 6 shows multi-model results from combining forecast global SST fields. Chapter 7 summarizes the research results, presents conclusions and key recommendations.

CHAPTER 2:
MODELS AND DATA

2.1. Observed rainfall data

A total of 963 rainfall stations, more or less evenly distributed over South Africa, have been used in the study (Figure 2.1). The data are available from January 1960 to December 2004. Three-month totals were calculated and the seasons considered in the analysis are September-October-November (SON), December-January-February (DJF), March-April-May (MAM) and June-July-August (JJA).

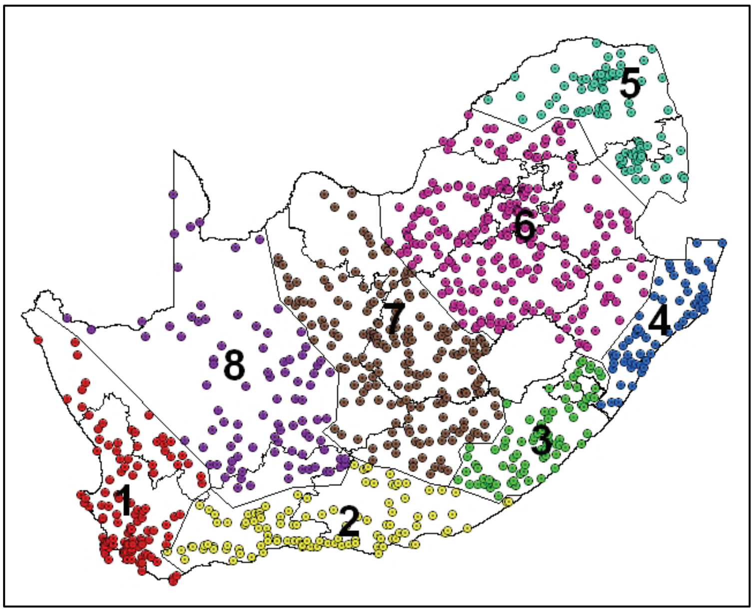

Figure 2.1. The 963 rainfall stations used in the analysis.

2.2. Streamflow data

The streamflow data were obtained from the School of Bioresources Engineering and Environmental Hydrology at the University of KwaZulu-Natal. The data are for 1946 Quaternary Catchments. The data comprise of the streamflow generated by the Agricultural Catchments Research Unit (ACRU) model in the individual catchment, the streamflow generated in the individual catchment plus the streamflow generated in all upstream catchments (this is what would be measured in the river if taking measurements was possible), and the rainfall data used in simulating flows - these are data from a local station that have been adjusted to be more representative of the catchment as a whole. The variables are all in units of mm per day. A description of the ACRU model and data can be found in Schulze *et al*. (2005a, 2005b).

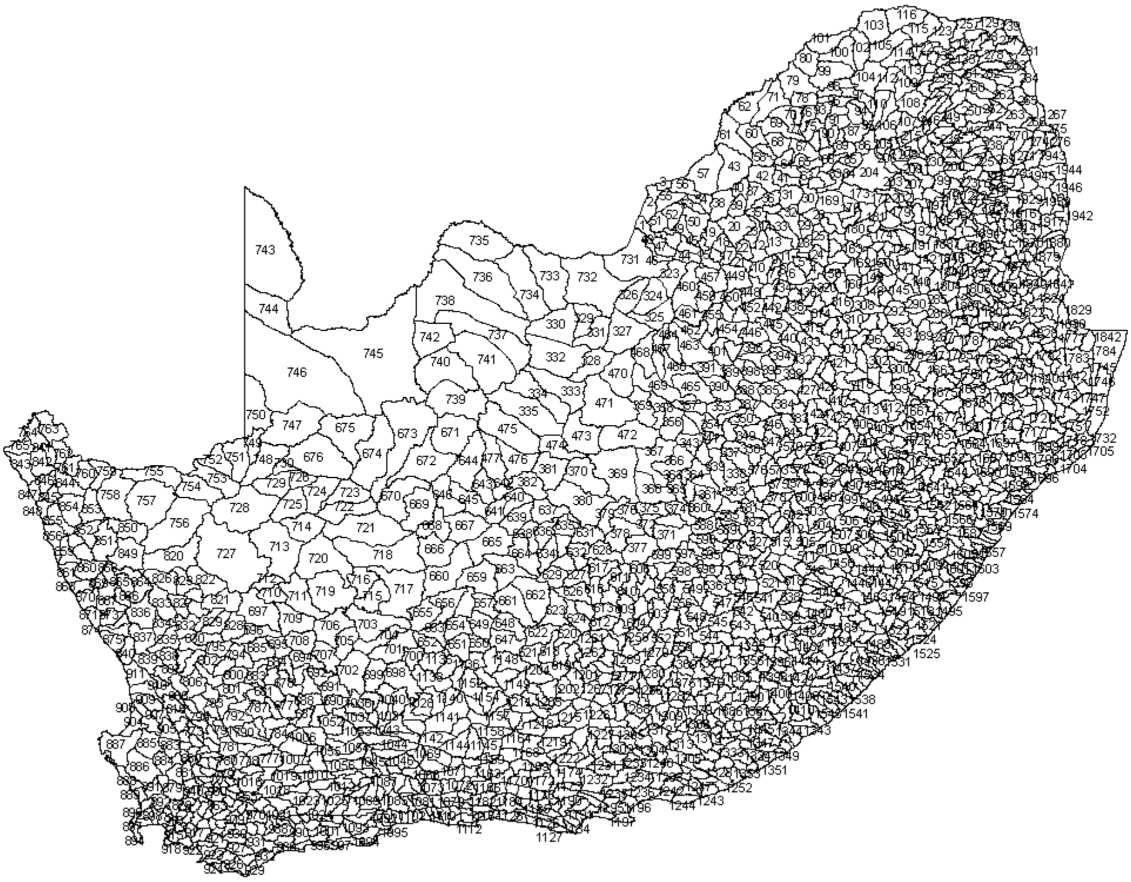

Figure 2.2. The 1946 catchments used in the analysis.

2.3. The general circulation models

The GCMs' output used here are from three coupled models from the DEMETER project (Palmer *et al.*, 2004) (UKMO, ECMWF and Météo-France), and from two atmospheric models, the ECHAM4.5 (Roeckner *et al.*, 1996) output obtained from the IRI, and the CCAM (McGregor, 1996; 2005) output obtained from the University of Pretoria (UP). The SAWS has also installed and have been running the ECHAM4.5 GCM operationally on their supercomputer as a consequence of this project. However, since the SAWS has at this point only produced multi-decadal *simulation* data from the ECHAM4.5 GCM (data obtained when forcing the GCM with simultaneously observed SST anomalies) and for only 6 ensemble members, it was decided to instead use the ECHAM4.5 data from the larger ensemble and at forecast lead-times provided by the IRI. Notwithstanding, the SAWS is in the process of significantly expanding on the ECHAM4.5 ensemble and is already in the process of providing operational ECHAM4.5 rainfall and temperature forecasts.

The ECHAM4.5 simulations performed at the SAWS have been verified, and the verification presented meets the requirements of a Standard Verification System (SVS), as stipulated by the World Meteorological Organization

(WMO), for a member centre with the capacity to run global models to be recognized as a Global Producing Centre (GPC) for Long-Range Forecasting (LRF) products. The SAWS will submit their application for GPC status later in 2008 for consideration by the Commission for Basic Systems that is scheduled to meet in 2009.

Two sets of ECHAM4.5 integrations were used here. The first ensemble of 24 runs was forced with simultaneously observed SST (Reynolds and Smith, 1994; Smith *et al*., 1996) over a period that spans several decades (1950 to present). The second ensemble contains 12 runs and was forced by persisting the SST anomalies of a month antecedent to the target period in order to facilitate a 1-month lead time for a multi-decadal data set (1968 to present). At initialization, ensemble members differ from each other by one model day at the beginning of the integration. The horizontal resolution is about 2.8°x2.8°.

Five 25-year climate simulations have been performed with the cubic-conformal atmospheric model (CCAM) by forcing the model with simultaneously observed SST anomalies. All five simulations are for the period 1979-2003, but each simulation employed different initial conditions. The model ran globally at C48 (approximately 200 km) horizontal resolution on a quasi-uniform grid. Output for a number of variables is available on a global 1° latitude-longitude grid. The simulations were performed on the Velocity-cluster at the University of Pretoria. In operational mode, CCAM is initialized using the 0Z analysis field obtained from the Global Forecasting System (GFS). A three-month seasonal forecast (having 8 ensemble members initialized on 8 consecutive days) is issued on a monthly basis. Lower boundary forcing is prescribed from persisted SST anomalies, as obtained from the GFS.

The models used from the DEMETER set (Palmer *et al*., 2004) are the UKMO, ECMWF and Météo-France fully coupled models. Each model has 9 ensemble members available, and the fields considered here are the total precipitation and 850 hPa geopotential height fields for downscaling, and the SST fields used in the section on multi-model SST predictability. All model fields are those associated with a 1-month lead-time, and the horizontal resolution is 2.5°x2.5°. The GCM data from the DEMETER set was obtained from:
http://data.ecmwf.int/data/d/demeter_mnth/1950/hindcasts/.

The ensemble means from each of the physical models is used since the ensemble mean best represents the predictable signal from the full ensemble (the noise component of each ensemble member normally is large). In addition, the ensemble mean values are considered instead of the full ensemble from each physical model (24 members from ECHAM4.5, 9 members each from the DEMETER models, and only 5 from the CCAM) in order to make sure that improved skill through combination is actually because of combination and not because of a larger ensemble size for a particular model. The assumption is also made that the number of available members from each model used to calculate the ensemble mean is adequate so that adding additional members will not show significant improvement in simulating South Africa's DJF rainfall. In fact, an ensemble mean calculated from more than 16 members produced by the ECHAM4.5 model have been found to only add marginally to the skill (Dr Lisa Goddard, IRI, personal communication).

2.4. Canonical correlation analysis

The statistically based method used in constructing the SST-rainfall model (Landman and Mason, 1999) and the MOS models (Landman and Goddard, 2002) prior to model combination is called canonical correlation analysis (CCA). Canonical correlation analysis, which is often used as a forecast technique, is a multivariate statistical methodology to determine linear combinations of two data sets (the predictor data set, e.g. sea-surface temperature, and the predictand data set, e.g. observed rainfall) that are highly correlated, and is at the top of the regression modeling hierarchy (Barnett and Preisendorfer, 1987).

Canonical correlation analysis is used to seek relationships between two sets of variables. It attempts to find the optimum linear combination between the two sets with maximum correlation being produced. Because the predictor and the predictand fields contain a large number of highly correlated variables and few observations, it is recommended that pre-orthogonalisation (Barnston, 1994) using standard empirical orthogonal function (EOF) analysis (Jackson, 1991) be performed. The predictor and predictand data sets are first standardized, resulting in correlation matrices on which the EOF analysis is performed. The standardization ensures that all the grid points and rainfall indices have equal opportunity to participate in the prediction process (Jackson, 1991; Barnston, 1994). EOF-analysis is performed separately for each of the predictor and predictand fields. The number of modes to be retained in the analysis (i.e. those that will be used in the CCA eigenanalysis problem) is determined using forecast skill sensitivity tests.

The Climate Predictability Tool (CPT) software, developed at the IRI, is used for the CCA calculations. The performance of the individual models (SST-rainfall and GCM-MOS) as well as the combined forecasts is obtained by following a 5-year-out cross-validation procedure.

2.5. Model output statistics

Model output statistics (MOS; Wilks, 2006) applied to GCM output can improve seasonal rainfall forecasts over southern Africa (Landman and Goddard, 2002; Shongwe *et al.*, 2006). MOS is applied to the ensemble mean fields of the three GCMs used in the study. For the ECHAM4.5 and CCAM a 24-year period (1979-2002), and for the UKMO a 23-year period (1979-2001) is used. Two GCM fields are considered: the total precipitation and the 850 hPa geopotential height fields. These two fields have been found to produce the best skill over the region using an older version of the ECHAM model (Landman and Goddard, 2002). The forecast fields from each GCM used in the MOS are restricted over a domain that covers an area between 10°S and 40°S, Greenwich to 60°E.

Area-averaged *Spearman ranked correlations* (Wilks, 2006) are frequently used in this report to estimate model performance. The correlation values are adjusted using the Fischer Z transformation (Wilks, 2006). The Spearman correlation is simply the Pearson (or "ordinary") correlation computed using the ranks of the data. However, this correlation is a robust (to deviations from linearity in a relationship) and resistant (to outlying data) alternative to the Pearson correlation.

Except for DJF, the best ECHAM4.5 predictor for seasonal rainfall is total precipitation. With this GCM the 850 hPa geopotential field is the best predictor for DJF rainfall. Although this skill assessment is done for ECHAM4.5 simulation data, it is assumed that the same result would be found using ECHAM4.5 data produced at a 1-month lead-time (Landman and Goddard, 2002). For CCAM, 850 hPa geopotential heights are the best predictor for DJF, MAM and JJA rainfall. For SON total precipitation produced the highest correlations. The higher DJF skill is again associated with the 850 hPa geopotential height fields when the UKMO is used. Total precipitation outscores the height fields for the remaining three seasons. The GCM fields that produced the highest correlations are subsequently used as the predictor fields in the respective GCM-MOS equations.

2.6. Combination methods

Three linear combination schemes are considered: equal weights (Mason and Mimmack, 2002), CCA (Barnett and Preisendorfer, 1987) and principal components regression (PCR; Wilks, 2006). Combination through equal weighting involves the summation of forecasts and then division by the number of models involved. For CCA and PCR the predictor fields from the GCMs and SST-rainfall models are first interpolated to a common 2.5°x2.5° grid, and then combined in order to have more than one variable (model field) at each grid-point. Some CCAM-MOS skill is lost after interpolating the data from the 1°x1° to the 2.5°x2.5° resolution.

2.7. Synopsis

Properties and data from various physical and empirical models have been described. They include the GCMs, statistical base-line model, MOS and model combinations. The observed South African station rainfall and catchment data have also been described.

CHAPTER 3
COMBINING MULTIPLE MODELS

3.1. Background

The South African Weather Service (SAWS) issues forecasts of rainfall and temperature at various time ranges, including forecasts on seasonal time scales. For this purpose, seasonal forecasts maps and output data from a number of general circulation models (GCMs) are obtained from international centres (e.g. IRI, UKMO and ECMWF) and the Universities of Cape Town and Pretoria (www.gfcsa.net). These forecasts are then subjectively combined through consensus discussion by SAWS forecasters by also considering forecasts produced with existing seasonal forecasting systems implemented at the SAWS and also those developed at the SAWS. These systems include rainfall-sea-surface temperature (SST) empirical models (Landman and Mason, 1999) and ECHAM4.5 (Roeckner *et al*., 1996) forecasts that are additionally downscaled statistically to 963 rainfall stations evenly distributed across South Africa (Landman and Goddard, 2005). In 2006 the SAWS issued an operational rainfall forecast for the 2006/07 December-January-February (DJF) season (Kgatuke *et al*., 2006) that, for the first time was produced by a regional model, RegCM3 (http://users.ictp.it/~pubregcm/RegCM3),. Six-hourly forecasts provided the initial conditions and the time dependent lateral boundary conditions were incorporated from ECHAM4.5 forecasts.

The SAWS is developing an objective forecasting system that is based on a multi-model forecasting approach. Such a system will replace the current subjective forecasting system that relies heavily on forecasters' interpretation of model output. In addition to removing the subjectivity of the current forecasting system, there are advantages in combining a number of GCMs into a multi-model ensemble since GCMs differ in their parameterizations and therefore differ in their performance under different conditions (Krishnamarti *et al*., 2000). Multi-model systems are nearly always better than any of the individual systems (Doblas-Reyes *et al*., 2000, Krishnamurti *et al*., 2000) and the multi-model approach is further enhanced if the individual systems produce independent skilful information (Graham *et al*., 2000).

3.2. Combination procedure

3.2.1. Selecting models and test periods

Four models are considered here: Two atmospheric GCMs (ECHAM4.5 and CCAM), one fully coupled GCM (UKMO)), and one empirical SST-rainfall model. Figure 3.1 shows a schematic representation of the models as well as the linear combination schemes considered. The multi-model system that is currently being developed at the SAWS is considering the inclusion of forecasts from all of these models, but it must be tested whether or not the inclusion of all of them optimises forecast skill.

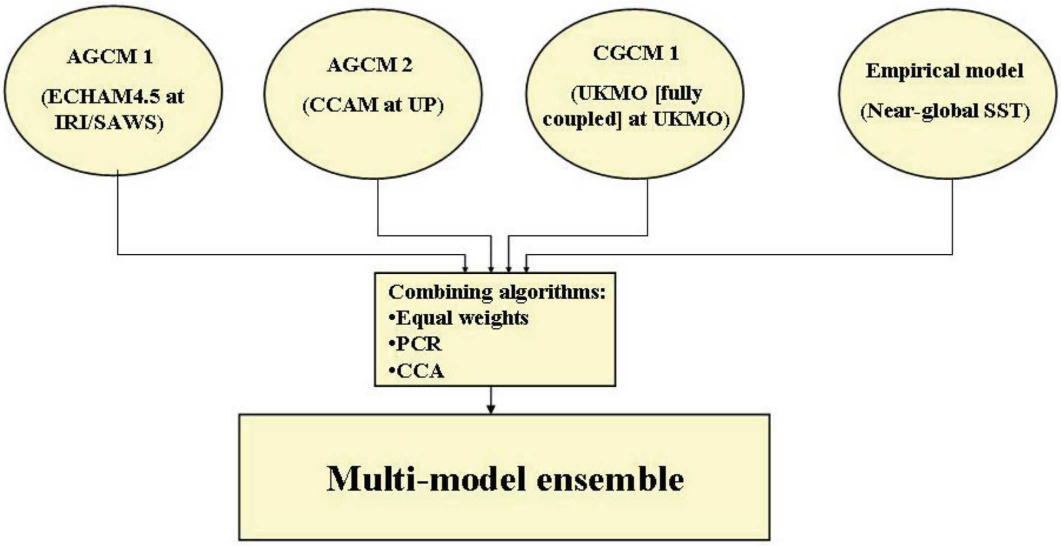

Figure 3.1. The seasonal forecast multi-model system under consideration.

Through the SAWS/UP partnership the CCAM has been configured on a cluster of computers at UP in order to generate operational seasonal rainfall forecasts every month. However, this model has up to now only been used to generate a 5-member ensemble of simulation data (simultaneously observed SST) for the period 1979 to 2003 (25 years). The ECHAM4.5 simulation and forecast data include this 25-year period. The multi-model system to be tested involving the CCAM model is therefore only considered in simulation mode (no forecast lead-time) and is tested over the said period, and consists of the ECHAM4.5, CCAM and SST-rainfall model that relates simultaneous SSTs with seasonal rainfall (e.g. DJF SSTs predicting DJF rainfall).

The UKMO model is also considered in the multi-model since the SAWS is involved in a collaborative project with the UK Met Office aimed at developing the Unified Model. One of the objectives of this project is to evaluate the HadGEM and GloSea4 seasonal forecast performance in southern Africa and to develop plans to use GloSea4 data as part of the SAWS multi-model seasonal forecasting system by early 2009. It is therefore sensible to use here the available data from the UKMO model. This data set covers the period 1959 to 2001. A multi-model system capable of making operational forecasts at a 1-month lead-time consists of the ECHAM4.5, UKMO and SST-rainfall model, all configured to produce forecasts over the 1979 to 2001 period at a 1-month lead time. This 23-year period is then comparable in length to the 24-year period used for the multi-model system based on simulation data.

3.2.2. Single vs. multi-model performance

Skill obtained from the multi-model system should outscore the skill from the individual models to justify its use. This section compares cross-validation skill (Spearman correlations) of the GCM-MOS, SST-rainfall and multi-model systems for both the simulation and forecast (at a 1-month lead-time) cases.

Figure 3.2 shows the area-averaged Spearman rank correlation values over the 963 stations for the SON season using simulation data. Skill levels are negative with no evidence that the multi-model system significantly improves on the simulations of the single models. The same result is found combining forecast data (Figure 3.3). This result indicates the challenges in predicting the spring season that is associated with the onset of the summer rainfall period over South Africa.

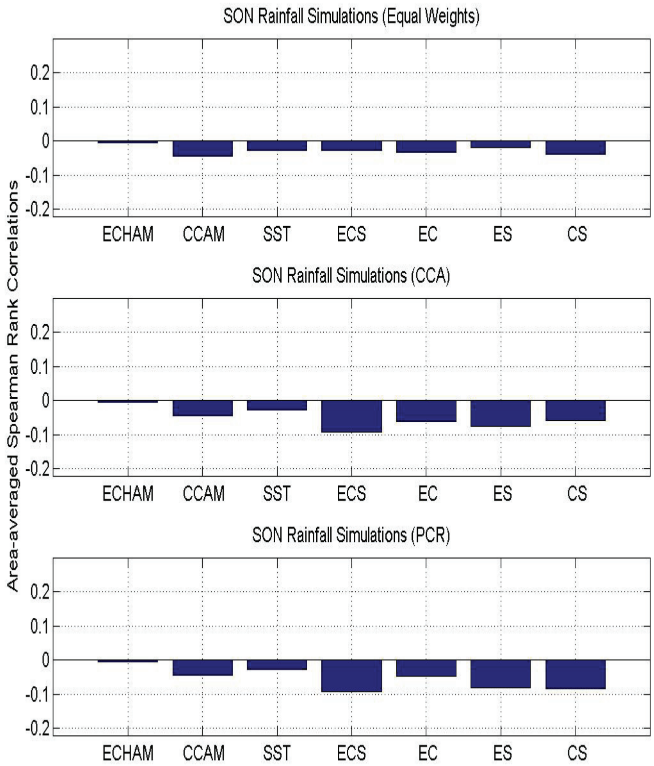

Figure 3.2. Single model vs. multi-model simulation skill for the SON season. The three single models are ECHAM4.5 (ECHAM), CCAM (CCAM) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + CCAM + SST-rainfall (ECS), ECHAM4.5 + CCAM (EC), ECHAM4.5 + SST-rainfall (ES) and CCAM + SST-rainfall (CS). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

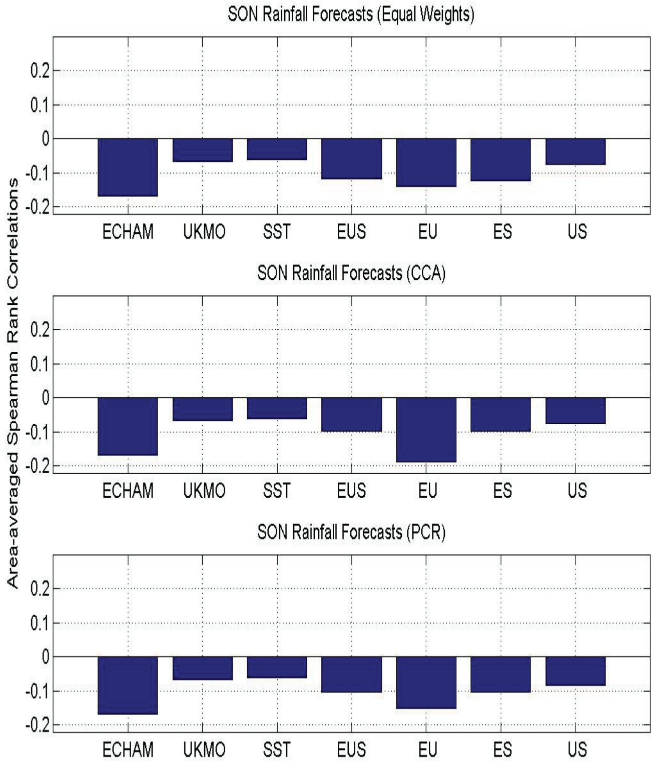

Figure 3.3. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the SON season. The three single models are ECHAM4.5 (ECHAM), UKMO (UKMO) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + UKMO + SST-rainfall (EUS), ECHAM4.5 + UKMO (EU), ECHAM4.5 + SST-rainfall (ES) and UKMO + SST-rainfall (US). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

Positive and much higher area-averaged Spearman correlations are found for DJF (Figures 3.4 and 3.5) as opposed to SON. For the DJF simulation data (Figure 3.4) the ECHAM4.5-MOS model outscores the other two single models and is comparable in skill to the multi-model simulations. In addition, using equal weights to combine the models produces higher skill levels than using either CCA or PCR for model combination. Moreover, for both the simulation and forecast data the SST-rainfall model's inclusion in any of the multi-models made the simulations worse, which is in contrast to what was found for South America (Coelho *et al.*, 2006).

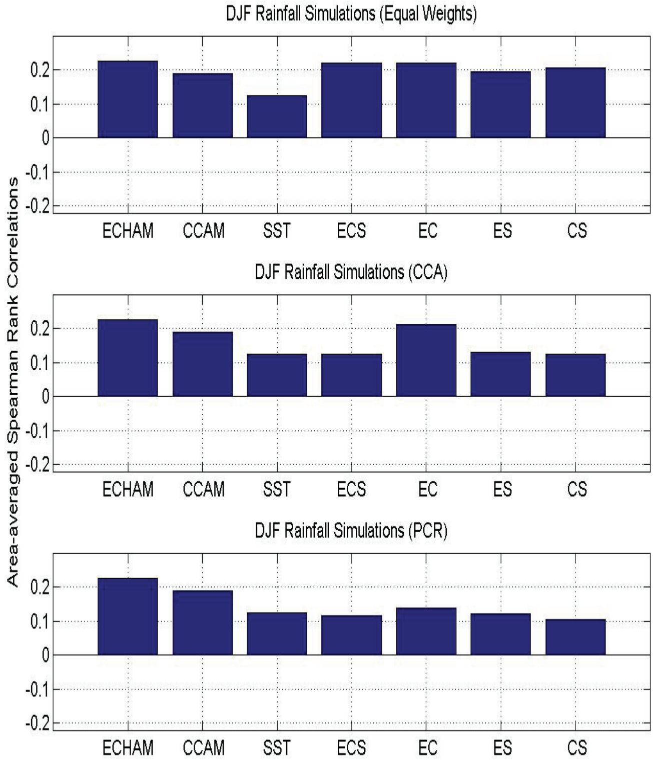

Figure 3.4. Single model vs. multi-model simulation skill for the DJF season. The three single models are ECHAM4.5 (ECHAM), CCAM (CCAM) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + CCAM + SST-rainfall (ECS), ECHAM4.5 + CCAM (EC), ECHAM4.5 + SST-rainfall (ES) and CCAM + SST-rainfall (CS). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

The single ECHAM4.5-MOS model has somewhat higher correlation scores than the best multi-model (0.2259 for ECHAM4.5 vs. 0.2210 for the ECHAM4.5 + CCAM multi-model). Similar results are seen for the DJF forecast data (Figure 3.5) where the best single model (UKMO-MOS) also outscores the best multi-model (ECHAM4.5 + UKMO). This result is in contrast to what is normally expected when models are combined. However, Figure 3.6 shows the verification results for DJF rainfall when a period of 34 years (1968/69-2001/02) as opposed to the 23-years of Figure 3.5 is considered. Here we see that for the longer period model combination indeed produces higher correlation scores than the models on their own.

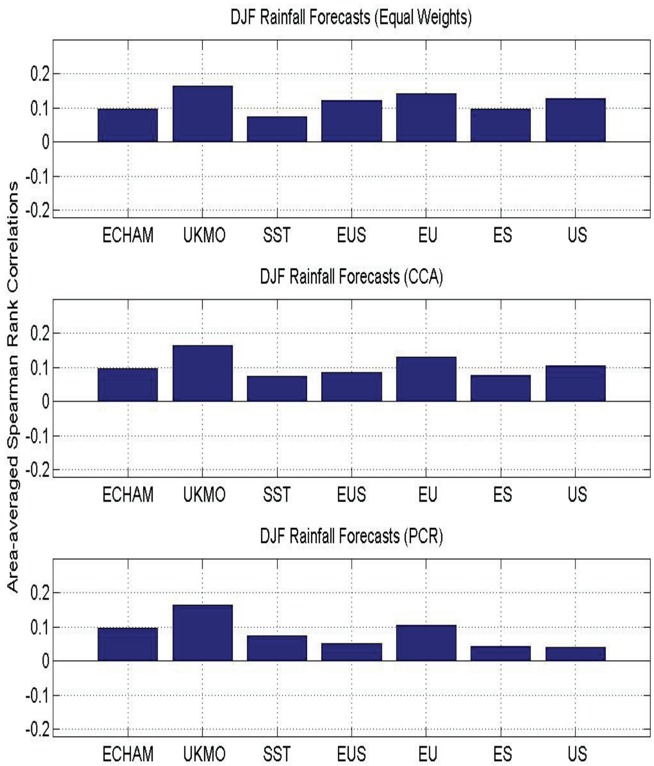

Figure 3.5. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the DJF season. The three single models are ECHAM4.5 (ECHAM), UKMO (UKMO) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + UKMO + SST-rainfall (EUS), ECHAM4.5 + UKMO (EU), ECHAM4.5 + SST-rainfall (ES) and UKMO + SST-rainfall (US). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

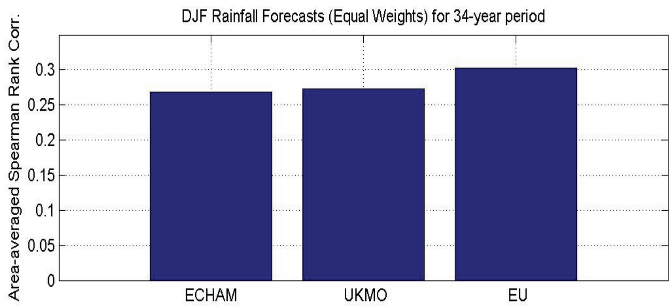

Figure 3.6. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the DJF season for the 34-year period of 1968/69-2001/02. The two single models are ECHAM4.5 (ECHAM) and UKMO (UKMO). The multi-model is ECHAM4.5 + UKMO (EU) and is obtained by averaging the forecasts.

Skill levels found for the MAM and JJA season are just as disappointing as is seen for the SON season (Figures 3.7 to 3.10). However, the scores associated with the CCAM model are higher, and in most cases positive, when using the original 1° x 1° horizontal resolution. In other words, some of the predictive signal from the CCAM is lost through the interpolation to a common 2.5° x 2.5° grid.

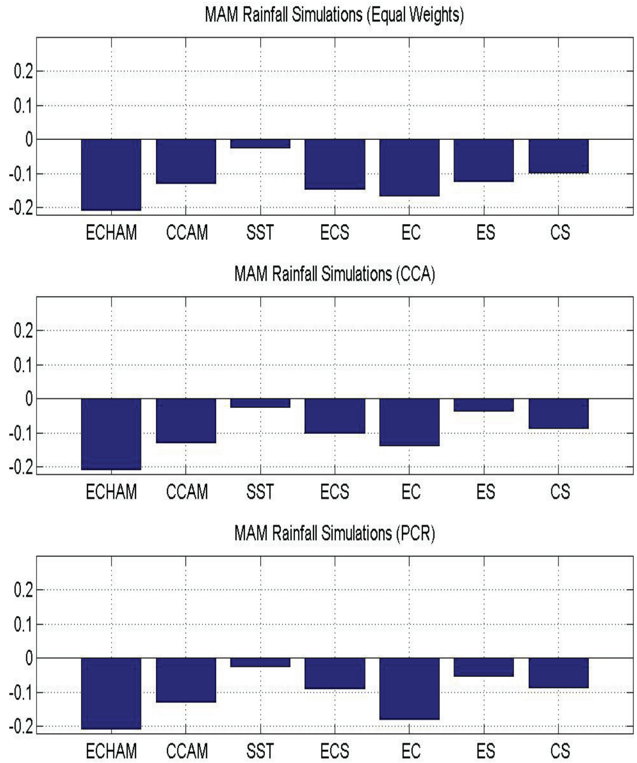

Figure 3.7. Single model vs. multi-model simulation skill for the MAM season. The three single models are ECHAM4.5 (ECHAM), CCAM (CCAM) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + CCAM + SST-rainfall (ECS), ECHAM4.5 + CCAM (EC), ECHAM4.5 + SST-rainfall (ES) and CCAM + SST-rainfall (CS). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

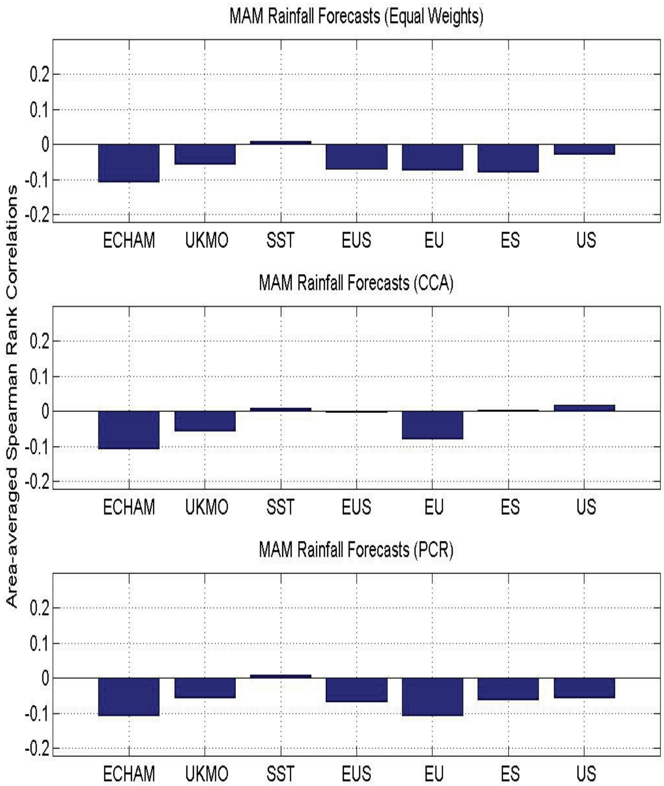

Figure 3.8. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the MAM season. The three single models are ECHAM4.5 (ECHAM), UKMO (UKMO) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + UKMO + SST-rainfall (EUS), ECHAM4.5 + UKMO (EU), ECHAM4.5 + SST-rainfall (ES) and UKMO + SST-rainfall (US). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

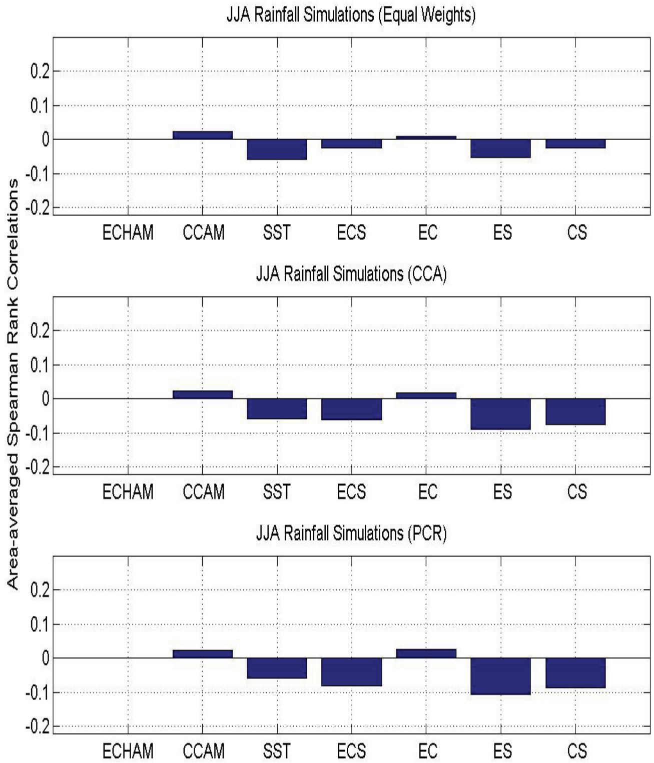

Figure 3.9. Single model vs. multi-model simulation skill for the JJA season. The three single models are ECHAM4.5 (ECHAM), CCAM (CCAM) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + CCAM + SST-rainfall (ECS), ECHAM4.5 + CCAM (EC), ECHAM4.5 + SST-rainfall (ES) and CCAM + SST-rainfall (CS). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

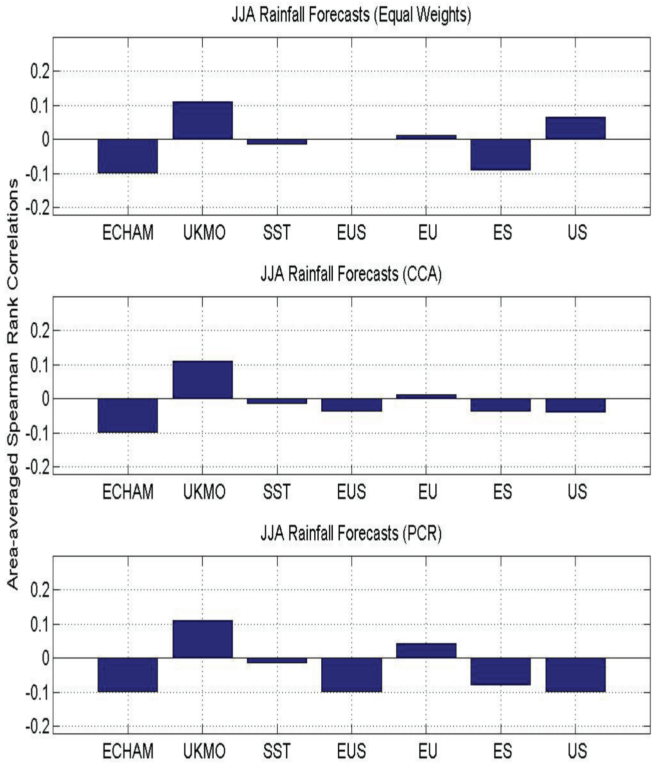

Figure 3.10. Single model vs. multi-model forecast skill (at a 1-month lead-time) for the JJA season. The three single models are ECHAM4.5 (ECHAM), UKMO (UKMO) and SST-rainfall (SST). The various multi-models are ECHAM4.5 + UKMO + SST-rainfall (EUS), ECHAM4.5 + UKMO (EU), ECHAM4.5 + SST-rainfall (ES) and UKMO + SST-rainfall (US). Three combination schemes are presented: Equal weights (top panel), CCA (middle panel) and PCR (bottom panel).

3.2.3. Geographical distribution of skill

The geographical distribution of skill of the multi-models with the best (highest area-averaged value) Spearman rank correlations are considered in this section. Figure 3.11 show the Spearman rank correlations at the 963 stations for SON rainfall simulations and forecasts by combining the model output through equal weighting.

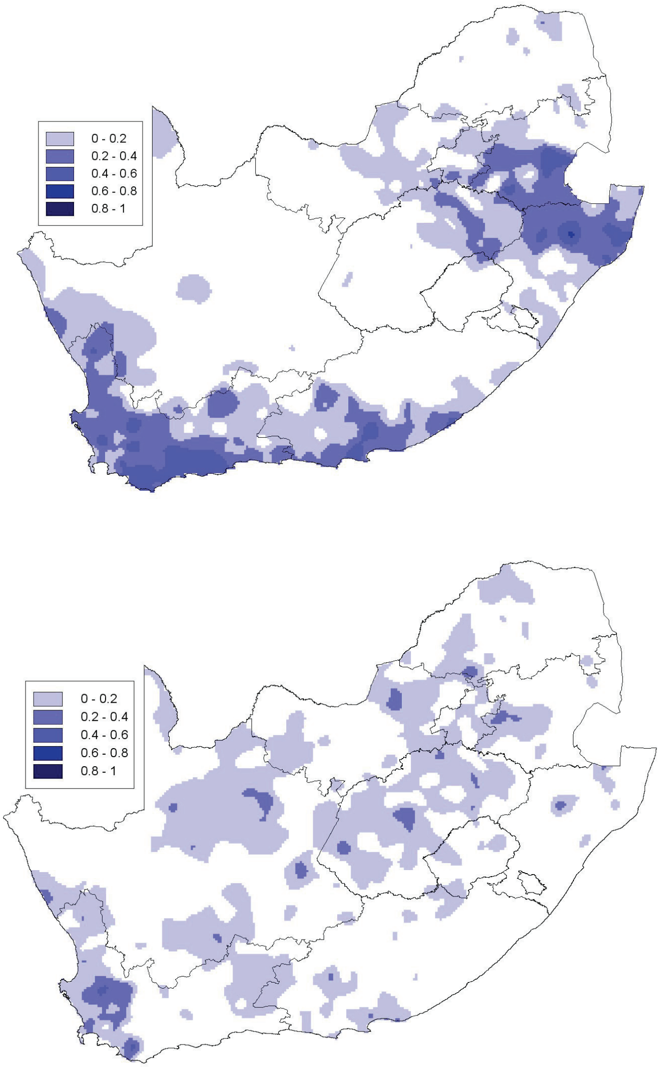

Figure 3.11. Spearman rank correlations from the best multi-model for SON simulated (top panel) and forecast (bottom panel) rainfall. The multi-model simulation consists of the ECHAM4.5 and SST-rainfall models, while the multi-model forecast consists of the UKMO and SST-rainfall models. Equal weighting is used to combine the models. Negative correlations are masked out.

Equal weighting also produced the best multi-model simulations and forecasts for DJF rainfall. Figures 3.12 and 3.13 show the Spearman rank correlation values and mean-squared-error skill scores (MSESS; Wilks, 2006) respectively for DJF rainfall. MSESS are calculated here using climatology as the reference forecast.

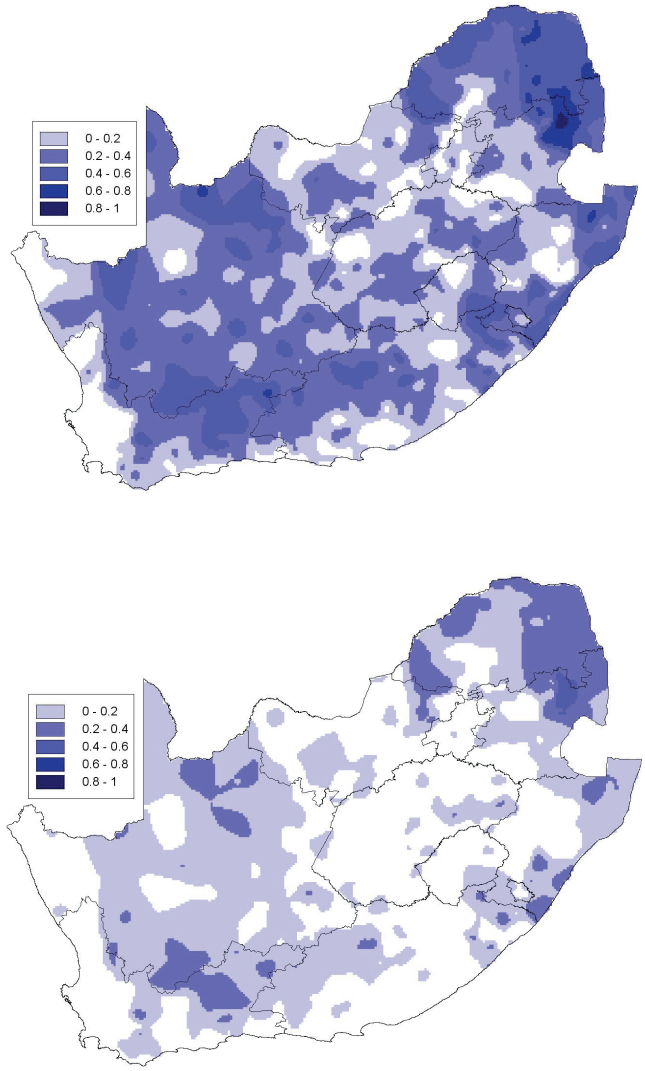

Figure 3.12. Spearman rank correlations (top panel) and MSESS (bottom panel) of the best multi-model for DJF simulated rainfall. The multi-model simulation consists of the ECHAM4.5 and CCAM models. Negative values are masked out.

The geographical distribution of skill shown on the maps of Figure 3.12 is similar for both skill measures: The best skill values are found over the north-eastern and western parts of the country, with the lowest skill values found over the central parts. This pattern is repeated when verifying the DJF forecasts made at a 1-month lead-time (Figure 3.13).

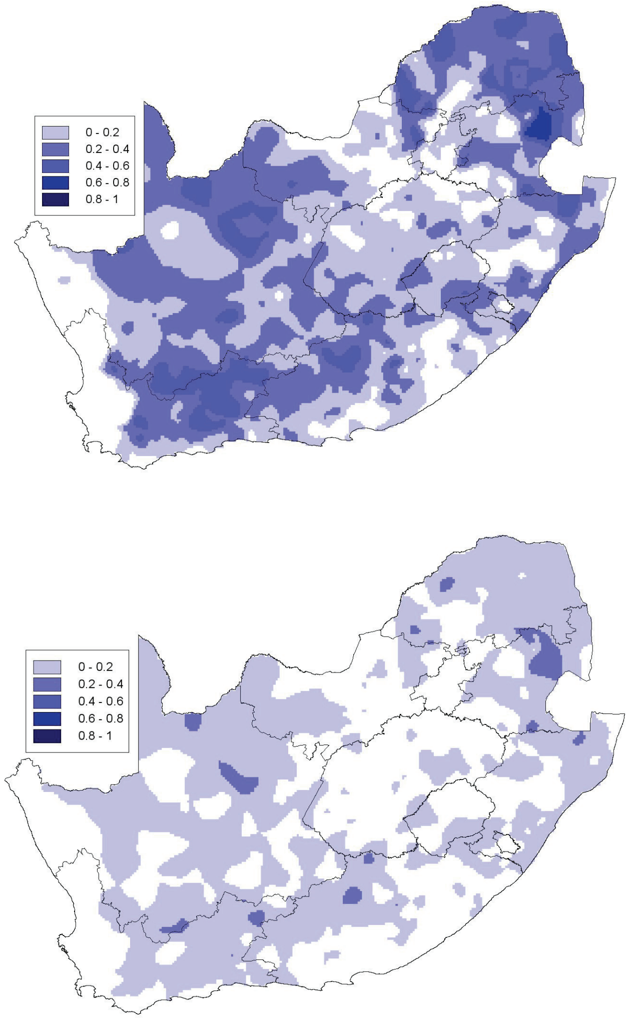

Figure 3.13. Spearman rank correlations (top panel) and MSESS (bottom panel) of the best multi-model for DJF forecast rainfall. The multi-model simulation consists of the ECHAM4.5 and UKMO models. Negative values are masked out.

For the MAM season, using CCA to combine the models produced the best results. Figure 3.14 shows the Spearman rank correlation values for the 963 stations. The highest skill values are found over the central and western interior regions. These areas receive most of their rainfall during the second half of the austral summer, so the skill pattern presented in Figure 3.14 demonstrates the usefulness of the multi-model system over these areas.

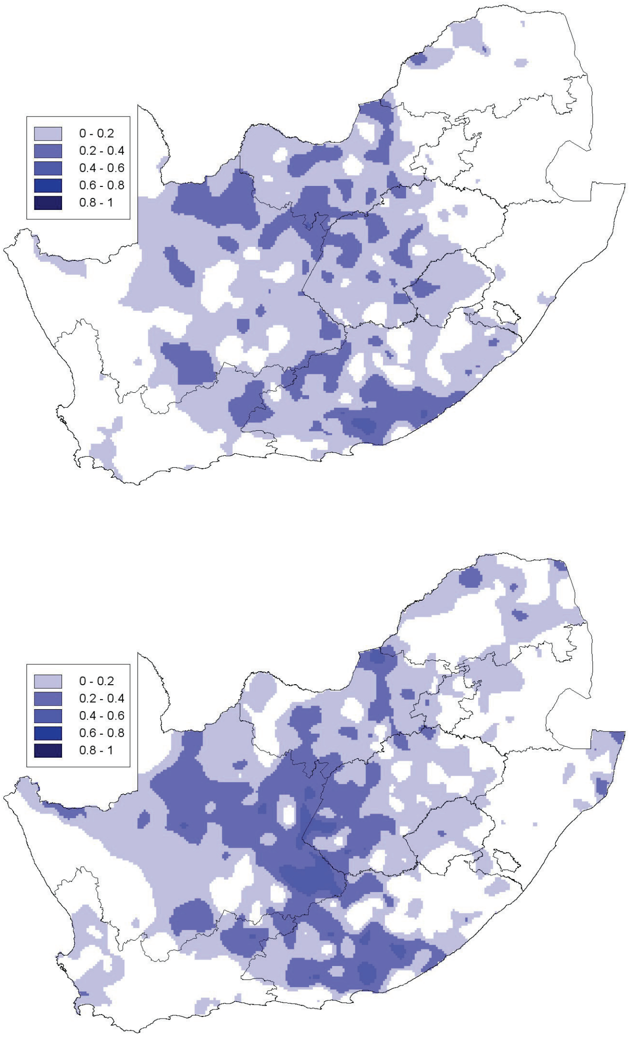

Figure 3.14. Spearman rank correlations from the best multi-model for MAM simulated (top panel) and forecast (bottom panel) rainfall. The multi-model simulation consists of the ECHAM4.5 and SST-rainfall models, while the multi-model forecast consists of the UKMO and SST-rainfall models. CCA is used to combine the models. Negative correlations are masked out.

The JJA rainfall skill maps of Figure 3.15 shows areas of positive skill values over large areas of the summer rainfall region. Low forecast skill is found over the winter rainfall regions of the south-western Cape. However, it was recently demonstrated that a relationship exists between the Antarctic Oscillation and winter rainfall over western South Africa (Reason and Rouault, 2005), but the domain selected for the GCM downscaling does not extend far enough to the south in order to include this oscillation.

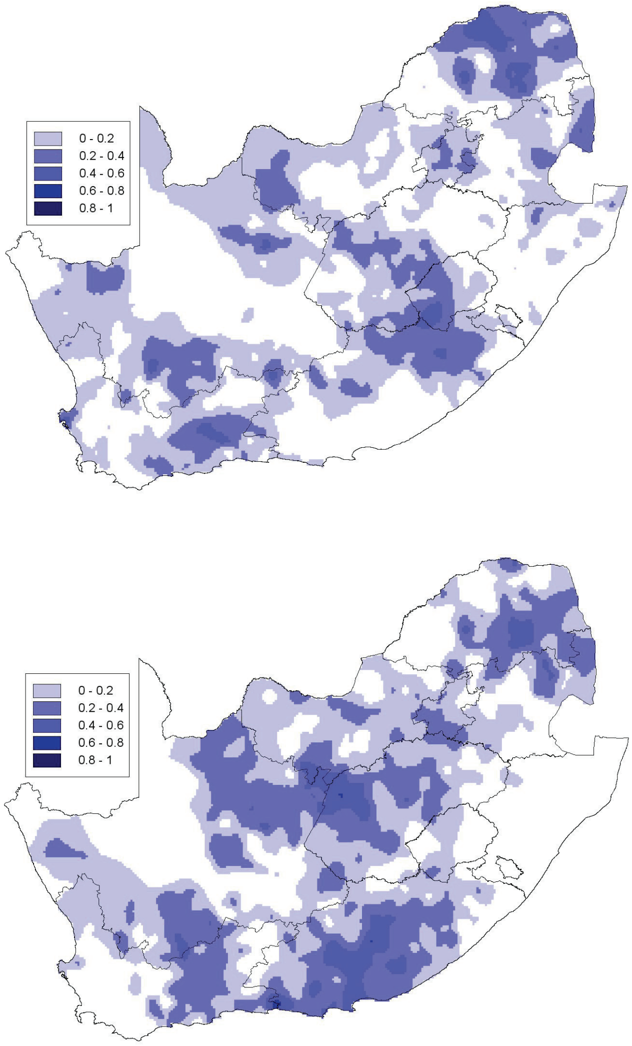

Figure 3.15. Spearman rank correlations from the best multi-model for JJA simulated (top panel) and forecast (bottom panel) rainfall. The multi-model simulation consists of the ECHAM4.5 and CCAM models, while the multi-model forecast consists of the UKMO and SST-rainfall models. PCR is to combine the simulation models, while equal weighting is used to combine the forecast models. Negative correlations are masked out.

3.2.4. Retro-active forecasting of DJF rainfall

This section addresses the predictability of DJF rainfall using a verification approach that is more relevant to a true operational forecasting environment. Cross-validation (even with a large cross-validation window) may indicate biased skill levels. To make sure skill levels are not biased, model validation should be conducted over a verification period that is independent of the model training or climate period. This approach is called retro-active forecasting (e.g. Landman *et al*., 2001) and involves model data excluding any information following the year to be simulated or forecast. Retro-active forecasting will be applied to both DJF simulation and forecast data. Unfortunately, with the relatively small data sets, the MOS equations are developed on even shorter training periods which may have a detrimental effect on the stability of the predictor-predictand relationship and consequently on model performance.

The retro-active procedure is applied to the GCM-MOS models and initially employs the first 14 (13) DJF seasons for the simulation (forecast) data training period. The training period is progressively lengthened by one year after estimation of the next year's DJF seasonal rainfall. For example, for estimating the DJF 1993/94 (1992/93) season's rainfall with the simulation (forecast) data, the GCM-MOS model training period is 1979/80 to 1992/93 (1979/80 to 1991/92). The next training period for simulation (forecast) data is 1979/80 to 1993/94 (1979/80 to 1992/93), in order to estimate DJF rainfall for 1994/95 (1993/94) rainfall. This procedure is repeated until the full 10-year period from 1993/94 to 2002/03 (1992/93 to 2001/02) of simulation (forecast) MOS data have been obtained. Figure 3.16 shows the area-averaged DJF rainfall totals for the simulation (left panel) and forecast data (right panel). During this period the extreme rainfall during the 1995/96 and 1999/2000 seasons is vastly underestimated. These two seasons have been associated with tropical cyclone landfall that caused widespread rainfall over a large part of the summer rainfall region (e.g. Bonita in 1996 and Eline in 2000).

The associations between the forecast and observed time series are poor for this retro-active test period. One of the reasons may be the short training period involved. Figure 3.17 shows the retro-active forecast results using a longer training period: The shortest training period is 24 years (as opposed to 13 years before) and the longest is 33 years (as opposed to 22 years before). There is an improvement in skill when longer training periods are involved. This result is in agreement with what has been found for the skill levels associated with the 34-year training period presented in Figure 3.6.

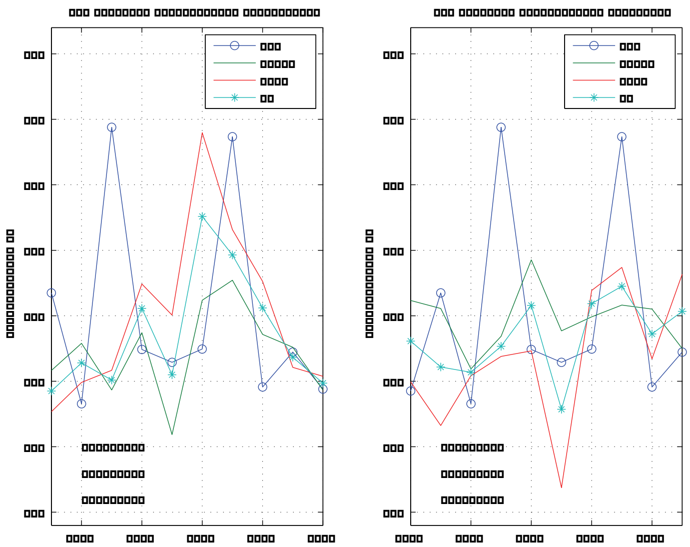

Figure 3.16. DJF rainfall GCM-MOS and multi-model simulations (left panel) and forecasts (right panel) using the retro-active procedure explained in the text. The multi-model (MM) values are obtained by simply averaging the time series of the GCM-MOS models. Correlations between the DJF estimated values and the observed are shown in the bottom left hand corners.

The retro-active multi-model (ECHAM4.5 + UKMO) forecasts over the 10-year period are evaluated next. Probabilistic forecasts are generated by the Climate Predictability Tool (CPT) software using ensemble mean information.

The probability forecasts for three categories of above-normal, near-normal and below-normal are calculated for the 10-year retro-active period of 1992/93 to 2001/02. Probabilistic forecast performance is evaluated using relative operating characteristic (ROC; Wilks, 2006) scores. Figure 3.18 shows the ROC scores over the 963 stations for above-normal DJF rainfall (top panel) and below-normal (bottom panel) rainfall. The respective ROC patterns show sporadic values higher than 0.5 and the geographical distribution of skill presented in Figures 3.12 and 3.13 is not evident here. This poor retro-active forecasting performance is attributed to the short training periods involved.

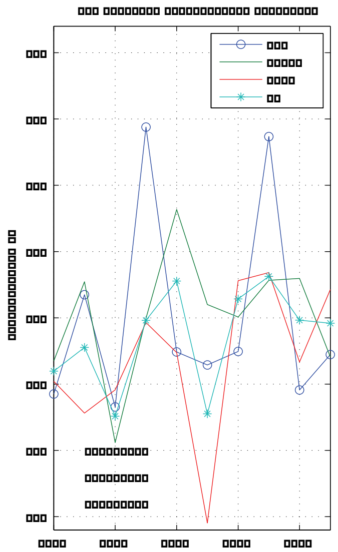

Figure 3.17. DJF rainfall GCM-MOS and multi-model forecasts using the retro-active procedure explained in the text, but for a minimum training period that is 11 years longer than the one used to make the forecasts presented in Figure 3.16. The multi-model (MM) values are obtained by simply averaging the time series of the GCM-MOS models. Correlations between the DJF estimated values and the observed are shown in the bottom left hand corner.

3.2.5. An operational forecast system

One of the objectives of the project was to set up an operational multi-model prediction system at the SAWS. This section explains how the very first operational multi-model forecast the SAWS issued was compiled and then presents the forecasts. This forecast was issued on the 31st of March 2008 and made use of an 8-member ensemble CCAM forecast produced at UP, and a 24-member ECHAM4.5 forecast produced at the IRI.
MOS equations relating 850 hPa geopotential height fields to April-May-June station rainfall totals were compiled based on the ensemble mean values of the ECHAM4.5 and CCAM GCMs, respectively. The two sets of MOS equations subsequently used the latest forecast 850 hPa geopotential height data from the two GCMs to respectively produce 24-member and 8-member ensemble sets of downscaled rainfall forecasts at 963 rainfall stations.

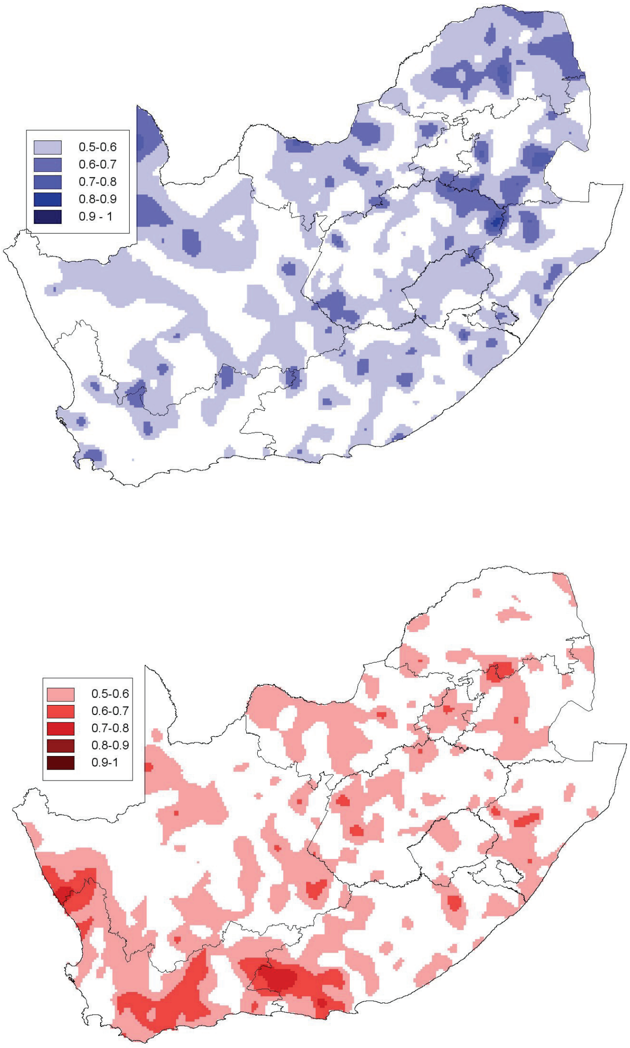

Figure 3.18. Relative operating characteristic (ROC) scores for the multi-model (ECHAM4.5 + UKMO) DJF rainfall 10-year (1992/93 to 2001/02) retro-active forecasts. The top panel shows ROC scores for above-normal rainfall and the bottom panel for below-normal rainfall. ROC scores below 0.5 are shaded out.

Multi-model results presented here show that the use of simple averages is a sound way to combine forecasts. Each operational forecast member is first evaluated according to its own MOS model's climatology in order to produce a forecast of above-normal, near-normal or below-normal at each station. The downscaled forecasts from the two models are then combined by simply giving equal weight to each of the 32 members in the grand ensemble. The number of hits per category is then calculated to determine the probability of a particular category occurring at each station.

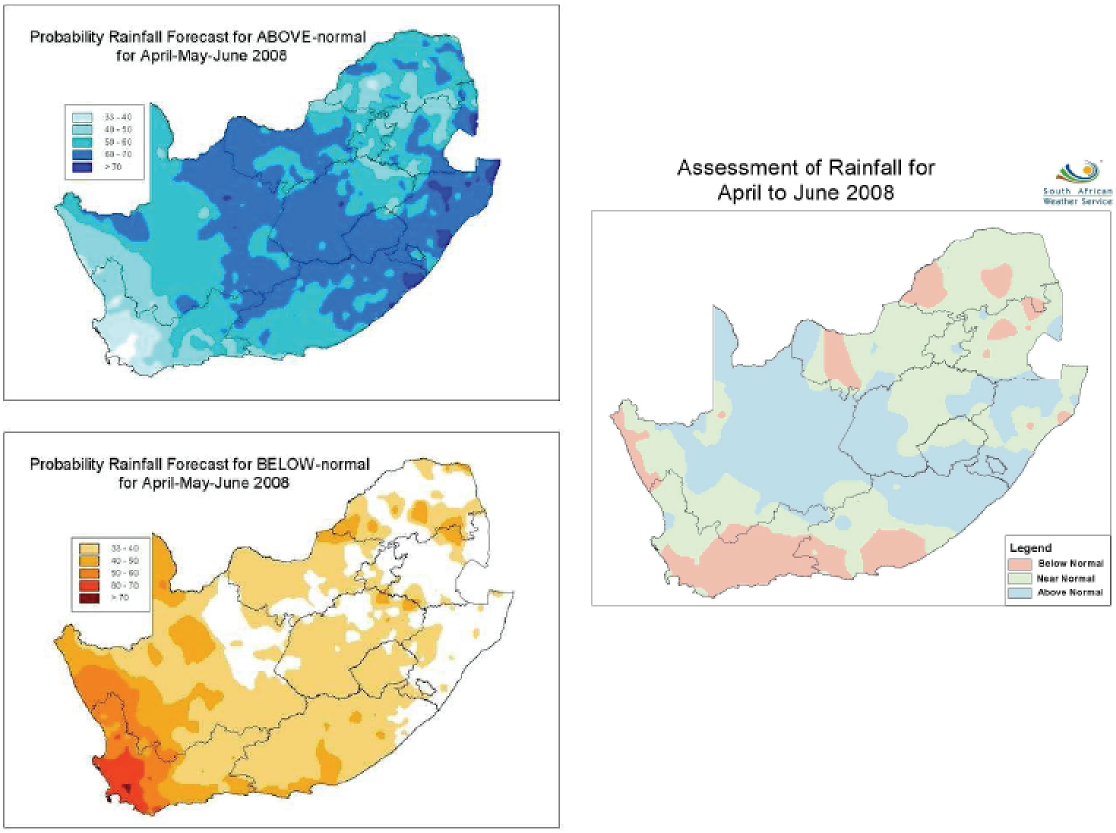

Figure 3.19. The first multi-model probability forecast issued by the SAWS on 31 March 2008. The top left map shows probabilities for the above-normal rainfall to occur, and the bottom left map for below-normal rainfall to occur during the April-May-June (AMJ) 2008 season. The map on the right shows the observed AMJ 2008 rainfall categories.

Figure 3.19 shows the first ever operational objective multi-model seasonal rainfall forecast issued by the SAWS. This forecast was issued on the 31st of March 2008 for the April-May-June 2008 season. The figure also shows the observed rainfall for the three categories of above-normal, near-normal and below-normal. Take note that the areas observed to have received above-normal rainfall totals coincide to a large extent with the areas of forecast high probabilities of occurrence of above-normal rainfall.

3.2.6. A multi-model forecast application

The predictability of streamflow using the multi-model system introduced above is investigated next. The streamflow data are only available until December 1999, so the maximum available test period covers the 31 years from 1968/69 to 1998/99. As with the station rainfall data, the 850 hPa geopotential height fields of the ECHAM4.5 and UKMO GCMs are statistically downscaled to the catchments shown in Figure 2.2. CCA is used to set up the MOS equations that reflect the association between the models' forecast height fields and the accumulated streamflow of each catchment. A 5-year-out cross-validation window is employed. Figure 3.20 shows the area-averaged Spearman rank correlation scores for the two individual models of the multi-model system as well as that of the multi-model system. The latter system is obtained by simply averaging the forecasts of the two GCM-MOS models. As was found with predicting rainfall over sufficiently long testing periods, the multi-model forecasts outscore the performance of the individual models.

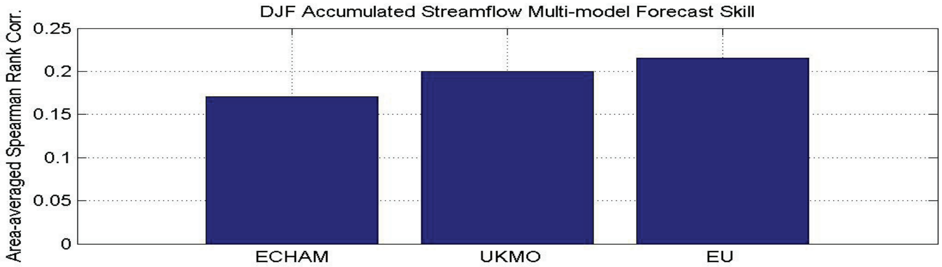

Figure 3.20. Single model vs. multi-model forecast skill (at a 1-month lead-time) for accumulated DJF streamflow for the 31-year period of 1968/69-1998/99. The two single models are ECHAM4.5 (ECHAM) and UKMO (UKMO). The multi-model is ECHAM4.5 + UKMO (EU) and is obtained by averaging the forecasts.

Multi-model forecast skill is also represented by mean-squared-error skill scores (MSESS) for each catchment (Figure 3.21). Take note of the positive MSESS values found along the major river systems such as the Orange and Vaal rivers.

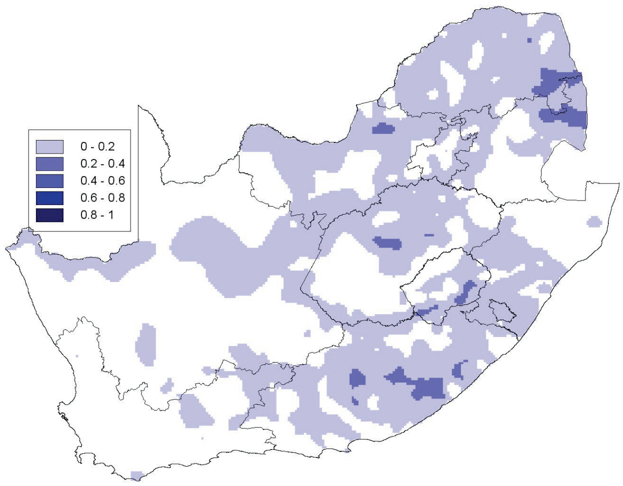

Figure 3.21. MSESS predicting accumulated DJF streamflow within the catchments shown in Figure 2.2. The multi-model simulation consists of the ECHAM4.5 and UKMO models. Negative values are masked out.

3.3. Synopsis

The SAWS has been using output from a large number of locally and internationally run models to compile the seasonal forecast subjectively. Now they are developing an objective multi-model forecasting system and the work presented in this chapter led to the first issuance of such a forecast in March 2008.

Four models and their combination have been tested here in both forecast (at a 1-month lead-time) and simulation mode. Low predictability has been found for the SON, MAM and JJA seasons, but useful forecast skill is associated with the DJF season. Model combinations did not seem to outscore the individual models conclusively, but skill assessments of combined forecasts made with longer training periods indicate improvement in skill over the single models. More importantly, multi-model systems that did not include SST-rainfall model forecasts for DJF outscored those that did.

The best linear combination approach seems to be equal weighting. There are certain advantages in such a simple approach. If the pool of candidate models is large, the problem of which subset of these models should be included in the combination process arises. One approach could be to only use those models with the best skill. However, the probability of identifying highly, but spuriously, skilful models increases as the pool of candidate models is expanded, i.e. if numerous models are used in the combination process, the probability of finding at least one model that gives spuriously "skilful" predictions increases. This problem is known as multiplicity (Wilks, 2006), and can lead to large errors in the combined forecasts since a large weight could be assigned to a model that is unsound. With equal weighting such models' forecasts will not be overemphasised. If the models' forecasts are themselves correlated with each other, errors in estimating the model weights can become considerable. The errors in these estimated weights can give poor predictions when new model forecasts are applied to the model combination scheme. Whereas multiplicity results in bad forecasts because of the inclusion of inappropriate models in the combination scheme, multicolinearity can cause bad forecasts even when the correct models are included simply because the weights are poorly estimated. Assigning an equal weight to all the models partly eliminates this problem.

The geographical distribution of skill shows restricted areas where the multi-model system can be of use. For SON very little use can be derived from the multi-models which places some doubt over the models' ability to predict the onset of the summer rainfall season. During DJF, however, large areas are identified where the multi-model can potentially provide useful forecasts. These areas of positive skill are restricted to the central-western interior and north-eastern regions, and are similar for two verification parameters (Spearman rank correlation and MSESS). MAM rainfall forecast skill is on average lower than that seen from DJF skill assessments, but the areas where skill is found during MAM are associated with summer rainfall seasonal maxima during the second half of the rainfall season. JJA skill assessments show large areas of positive skill over the predominantly summer rainfall regions, with little skill found over the peninsula and adjacent region where mainly winter rainfall occurs.

Retro-actively producing forecasts for a 10-year period shows that the multi-model failed to predict the excessively wet seasons of 1995/96 and 1999/2000, and although improvement in skill is found when longer training periods are involved, these wet extremes are not captured by the models. One of the reasons why these peaks have been missed is because these wet seasons were accompanied by tropical cyclone landfall that inundated the area with large volumes of rain. Operational forecasting of seasonal rainfall totals should therefore take into consideration the likelihoods associated with tropical cyclone development over the south-western Indian Ocean and the possibility of their landfall.

The chapter concludes by showing an application of multi-model forecasts. Streamflow measurements, as simulated by a hydrological model, have been predicted skillfully, especially along major river systems, using the 850 hPa heights as predictors in two GCM-MOS models and then combined using averages. Again, the multi-model outscores the individual models.

CHAPTER 4
ENSO AND MULTI-MODEL FORECAST SKILL FOR MID-SUMMER RAINFALL OVER SOUTH AFRICA

4.1. Background

An association exists between South Africa's summer seasonal rainfall and the equatorial Pacific Ocean. However, this link is not always strong since the association in the middle to late austral summer season is higher than earlier in the summer rainy season (e.g. Tyson and Preston-Whyte, 2000). Notwithstanding, in the mid-summer months South Africa tends to be anomalously dry during El Niño years and anomalously wet during La Niña years. Indian and Atlantic Ocean SST also have a statistically detectable influence on South African rainfall variability (e.g. Mason, 1995; Reason *et al*., 2006). Moreover, while the El Niño-Southern Oscillation (ENSO) has a control on rainfall variability over the southern African region, Indian Ocean SST anomalies, sometimes varying independently of ENSO, are important for atmospheric GCMs to skilfully simulate southern African seasonal rainfall (e.g. Washington and Preston, 2006). This section aims to show how South African summer rainfall forecast skill, produced by a state-of-the-art multi-model forecasting system, may be primarily determined by the state of the equatorial Pacific Ocean.

4.2. Data and method

The DJF seasonal total rainfall for a large number of SAWS stations evenly distributed across South Africa is the set of predictands in a model output statistics (MOS; Wilks, 2006) approach that uses a combination of large-scale total rainfall fields produced by a range of physical models as the set of predictors. The physical forecast models' output used here are from three coupled models from the DEMETER project (Palmer *et al*., 2004) (UKMO, ECMWF and Météo-France) and from one atmospheric model, the ECHAM4.5, obtained from the IRI. The MOS equations are developed using canonical correlation analysis (CCA; Landman and Goddard, 2002). The forecast fields from each GCM used in the MOS are restricted over a domain that covers an area between 10°S and 40°S, Greenwich to 60°E. A CCA model that uses near-global SSTs as predictor is considered as a fifth model in designing a multi-model seasonal rainfall forecasting system for South Africa. A one-month lead-time is imposed for each of the five models, i.e. forecasts made in November for a DJF rainfall forecast.

MOS is subsequently applied to the ensemble mean total rainfall forecast fields from each of the four physical models. Using MOS to recalibrate GCM output produces improved forecast skill for DJF rainfall over southern Africa (Landman and Goddard, 2002). The individual performances of the MOS models and the rainfall-SST statistical model are tested using a 5-year-out cross-validation design over the 34-year period from 1968/69 to 2001/02. The best of the five models is considered to be the one with the highest area-averaged cross-validation correlation value and the worst model the one with the lowest value. The correlation values are adjusted using the Fischer Z transformation (Wilks, 2006). Station data of only those stations that get most of their rainfall during the austral summer are used (758 stations), therefore excluding stations over the southern and south-western parts of South Africa.

A number of forecast combining algorithms exists. In this chapter CCA is used for model combination. The CPT software is again used for this purpose. First, all the models' output is combined in one predictor field for a five-model multi-model system. A multi-model system of four models is considered next by discarding the model with the lowest area-averaged correlation. This backward elimination of the models is performed by each time discarding the model with the lowest correlation value until only two models are considered in a multi-model system. In addition to these multi-model systems, each physical model's forecasts are also individually combined with the forecasts of the rainfall-SST statistical model. Finally, the best two physical models are also combined with the rainfall-SST statistical model. Nine multi-model systems are subsequently considered. The performance of the multi-model systems is tested using a 5-year-our cross-validation design over the 34-year test period. Model performance is estimated by the correlations between observed and predicted values and through the calculation of the mean squared error skill score (MSESS; Wilks, 2006, eq. 7.54). The MSESS is computed over the 758 stations to obtain an accuracy measure for field forecasts. The reference field forecast is the climatological average field.

4.3. Results

The ECHAM4.5-MOS model is found to perform the best over the 34-year test period. In descending order of performance, the next four models are the Météo-France-MOS model, the UKMO-MOS model, the rainfall-SST model and the ECMWF-MOS model. Figure 4.1 shows the various multi-model cross-validated forecasts and the observed values in mm, both averaged over the 758 summer rainfall stations. The correlation values between the area-averaged multi-model forecasts over the 34-year period and the observed area-averages are shown on the bottom left-hand side of the figure. The lowest correlation is found when combining the Météo-France forecasts with the rainfall-SST model forecasts (MM6=MFST). In fact, the considered multi-model systems did not benefit from the inclusion of the rainfall-SST model forecasts, which is in contrast to what has been found in a study for South America (Coelho *et al*., 2006). MM3 (E5MFUO=ECHAM4.5+Météo-France+UKMO) is associated with the highest correlation value (0.5902), followed by MM4 (E5MF=ECHAM4.5+Météo-France). The same result is obtained using either the MSESS as a skill measure or calculating the area-averaged adjusted correlation values: The MM3 is the most skilful multi-model system. This system is also skilful in predicting the upward or downward trend from one season to the next – the ability to predict if the next season will be wetter or drier than the previous season (e.g. will 1983/84 be wetter or drier than 1982/83?). For the 33 cases, the season-to-season trend is predicted accurately 27 times, which is equivalent to a hit rate of 82%. Of the 6 misses, 5 of them occurred when the following season is associated with neutral conditions in the equatorial Pacific Ocean. Also seen on Figure 4.1 is that the intensity of the El Niño related observed drought years of 1968/69, 1972/73, 1982/83 and 1994/95 are skilfully captured by the multi-model forecasts, while the intensity of the extremely wet conditions observed during the La Niña seasons of 1973/74, 1975/76, 1995/96 and 1999/2000 are not well captured by the forecasts (El Niño and La Niña years as defined by the Climate Prediction Centre, and shown as “E” and “L” on Figure 4.1).

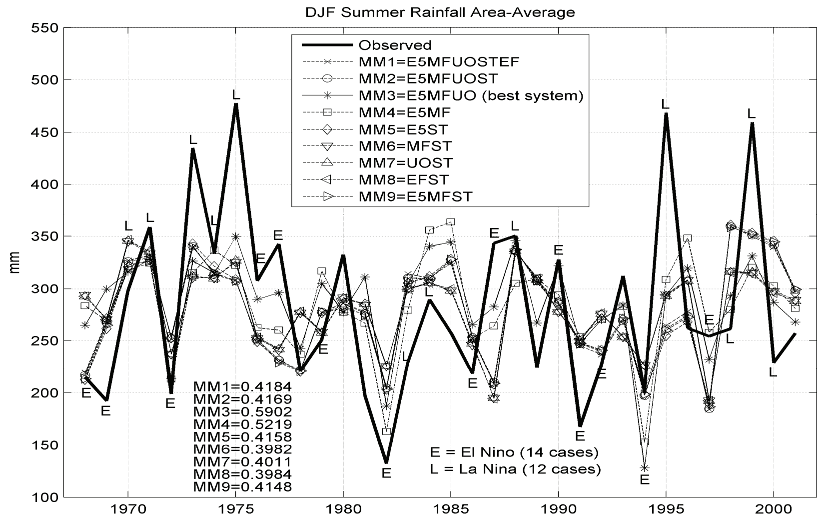

Figure 4.1. Observed versus multi-model forecast area-averaged values for DJF total rainfall in mm over the 34-year test period. Forecasts from nine multi-model systems (MM1 to MM9) are presented. The individual models considered in the various multi-models are ECHAM4.5 (E5), Météo-France (MF), UKMO (UO), SST-rainfall (ST) and ECMWF (EF). El Niño (E) and La Niña (L) seasons as defined by the Climate Prediction Centre, USA, are also shown. Correlations between multi-model forecasts and observed values are shown on the bottom left hand side of the figure. The years on the x-axis refer to the December months of the DJF seasons.

Figure 4.2 shows the MSESS values of the best single model (ECHAM4.5-MOS) and the best two multi-model systems. The multi-model system outperforms the ECHAM4.5 forecasts, which justifies the use of the multi-model system over the single-model system. For each multi-model system the MSESS values are separately recalculated for El Niño, La Niña, El Niño together with La Niña, and neutral years. The skill levels associated with El Niño seasons (14 in total) are higher than that of La Niña seasons (12 in total). Negative MSESS values are found (performance is worse than using climatology as a forecast) for both the multi-model systems shown in Figure 4.2 when only neutral years (8 in total) are considered, indicating that the multi-model is better able to predict DJF rainfall over South African during El Niño and La Niña seasons. The bias or mean error (defined here in a spatial context similar to the calculation of the MSESS) shows that the forecast errors associated with La Niña seasons (-360mm, i.e. the forecasts are too dry) are larger than the forecast errors associated with El Niño seasons (+224mm, i.e. the forecasts are too wet). An underestimation of the observed anomalies is to be expected since the variance of MOS-corrected forecasts is less than that of the observations. The reason why the bias during La Niña seasons is the larger is attributed to the low forecast amplitudes during the very wet 1973/74, 1975/76, 1995/96 and 1999/2000 seasons.

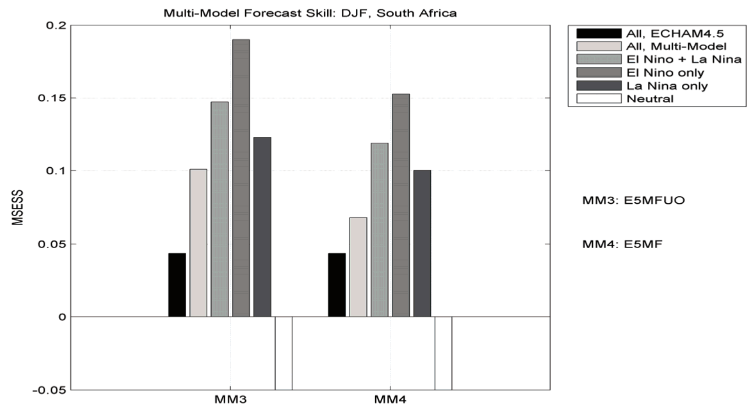

Figure 4.2. Mean squared error skill score (MSESS) values of the best single- and best two multi-model systems, considering all of the 34 years, El Niño together with La Niña years (26), only El Niño years (14), only La Niña years (12), and only neutral years (8). Negative MSESS values imply that the performance associated with the model forecasts is worse than when using climatology as the forecasts.

CCA pattern and time series analysis (Barnett and Preisendorfer, 1987) of MM3 (E5MFUO) and DJF rainfall of the 758 stations suggests that the dominant mode of variability (three canonical modes produce best forecast results) is related to ENSO since the correlations between the Oceanic Niño Index (ONI; www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml) and the leading canonical vectors of the predictor and predictand are respectively 0.66 and 0.59, both values being statistically significant at the 99% level. This high association explains why the multi-model performs the best during ENSO seasons.

Results presented here are for the DJF seasons only, since the tropical atmosphere dominates during this season creating the prospects of useful forecast skill then. Moreover, the DJF season is important for crop farmers. Notwithstanding the meteorological and societal importance of the DJF season, the multi-model system was also tested during the austral spring season of September to November and during the autumn season of March to May (not shown), but skill levels lower than those presented here for DJF were obtained, with particularly low skill when predicting spring rainfall. In addition, no clear improvement of the multi-model system over the single-model forecast was found during the spring season.

4.4. Synopsis

Forecast rainfall fields retro-actively produced over a 34-year period from the ECHAM4.5 atmospheric GCM and three DEMETER fully coupled models have been statistically combined and downscaled to make 1-month lead-time DJF rainfall forecasts for 758 South African austral summer rainfall stations. The best skill is obtained during ENSO years, while negative skill is found during neutral years (Figure 4.2). Even though DJF rainfall is found to be predictable (positive MSESS values) using the multi-model system described here, the extremely wet seasons of the 1970s and 1990s are not properly captured by the multi-model system, and in general the multi-model system has a dry bias when predicting rainfall for La Niña seasons. Moreover, forecasts are too wet during El Niño seasons. There is however a smaller bias associated with El Niño seasons which explains why DJF rainfall forecasts for these seasons outperform rainfall forecasts during La Niña seasons.

Since the results presented here suggest that the dominant mode of variability responsible for model skill is associated with the equatorial Pacific Ocean, the question may arise what added benefit there may be in running multi-model systems, especially if these systems include elaborate and expensive fully-coupled models, over simple statistical models that use, for example, the Niño3.4 SST index as predictor. The individual GCM-MOS systems, except for the ECMWF system, outscore the rainfall-SST model (dominant CCA mode is also related to ENSO), and the inclusion of the latter model in the multi-models does not improve on multi-model forecast skill. This result suggests that the GCM downscaled forecasts include additional forecast information that cannot be derived from SSTs alone, which justifies the use of multi-models, especially during El Niño seasons. Moreover, the best multi-model MOS equations consist of three modes of variability of which only the most dominant one is ENSO related.

The evidence presented here that the multi-model system performed the best during ENSO years can be useful in an operational forecast setting. Mid-summer rainfall forecasts for South Africa during ENSO seasons can be made with greater confidence as opposed to predicting rainfall anomalies during neutral seasons, and may be considered a first approach in providing guidance on the forecast skill associated with a coming rainfall season.

The section has demonstrated that multi-model systems are able to provide useful operational mid-summer rainfall forecasts since they outperform forecasts of climatology for the most part, especially during ENSO seasons and in particular during El Niño seasons. However, operational centres such as the SAWS are required to produce seasonal forecasts ever year, regardless of the state of the equatorial Pacific Ocean. This chapter has thus pointed towards still outstanding challenges for seasonal climate modelling in the region. These include the production of skilful rainfall forecasts during neutral seasons and the ability to predict the possibility of extremely wet mid-summer seasons during La Niña years.

CHAPTER 5
TROPICAL VORTEX PREDICTION WITH GLOBAL MODELS

5.1. Background

The retro-active DJF rainfall simulations and forecasts presented in Figures 3.16 and 3.17 show that the GCM-MOS multi-models presented do not seem to have the ability to capture extreme seasons when they are associated with tropical cyclone landfall. Figure 5.1 shows the observed tracks over the south-western Indian Ocean during the 1995/96 and 1999/2000 seasons. Although most of the tracks are located to the east of Madagascar, they can migrate across the island and sometimes even develop within the channel between the island landmass and the subcontinent.

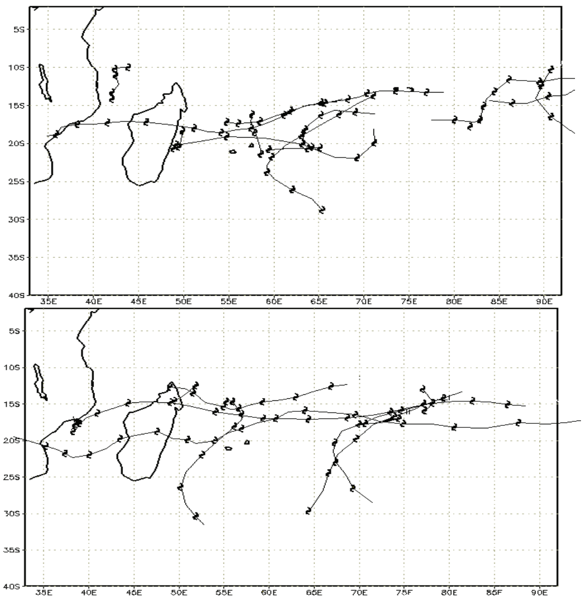

Figure 5.1. Observed tracks of tropical cyclones during the La Niña seasons of 1995/96 (top panel) and 1999/2000 (bottom panel).

Although GCMs have the ability to simulate tropical vortices that have good similarities with observed tropical cyclones, the coarse resolution of GCMs result in, for example, a lack of the presence of an eye and associated eyewall (Bengtsson *et al*., 1995). Moreover, the vortex tracks generated by GCMs are usually located too far to the east over the southern Indian Ocean (Vitart *et al*., 1997).

In this chapter tracking algorithms applied to atmospheric fields generated by global models are introduced and simulated tracks presented. Two approaches to generate the atmospheric fields to which tracking algorithms are applied are presented: nesting a regional climate model into simulated coarse resolution GCM output, and using a fine resolution GCM.

5.2. Nesting approach

5.2.1. The ECHAM4.5-RegCM3 system

One of the most currently economically viable methods for developing countries to do research and predict meteorological systems is through the use of regional climate models (RCMs). The implementation of this method has enabled the simulation of tropical cyclones that appear to be more realistic than those generated by general circulation models (Walsh, 1997). RCMs are good for studying sub-GCM grid processes. In this section, the RegCM3 RCM is nested within the large-scale fields of the ECHAM4.5 GCM. An objective algorithm for detecting and tracking tropical cyclones is subsequently applied to the RegCM3 output, similar to what has been done before for the south-western Indian Ocean (Landman *et al*., 2005).

The positions of the eastern and northern boundaries are chosen in such a way that they do not affect the simulated storms. The western and the southern boundaries do not have any influence on the storms propagating from east-to-west. Therefore, the selected model domain extends from about the equator to about 40°S, and from Greenwich to 70°E. The model is run at a horizontal resolution of 60km and a time step of 150s. Four ensemble members generated by ECHAM4.5 are used as the forcing fields for the RegCM3. The 4-month simulations are made over a 21-year period for the months of November, December, January and February. Initial and boundary conditions are derived by standard interpolation procedures from the ECHAM4.5 data grid to the RegCM3 grid.

5.2.2. Detection and tracking algorithm

The objective procedure for tracking model tropical cyclones used in this study has been adopted from Vitart *et al*. (1997). The tropical cyclone detection algorithm first locates the position of the intense vortices with a warm core for each model day as follows:

1. A local maximum of vorticity larger than 2.0 $X10^{-5}$ s^{-1} at 850 hPa is located.
2. The closest local minimum sea-level pressure is defined as the center of the storm.
3. The closest local maximum of average temperature between 500 and 200 hPa is located and is defined as the center of the warm core. The distance between the center of the warm core and the center of the cyclone must not exceed 2° latitude. From the center of the warm core the temperature must decrease by at least 0.5°C in all directions within a distance of 8° latitude.
4. The closest local maximum thickness between 1000 and 200 hPa is located. The distance between this local maximum and the center of the cyclone must not exceed 2° latitude. From this local maximum the thickness must decrease by at least 50 m in all directions within a distance of 8° latitude.

After the cyclones are located for each day, an objective procedure is applied to find cyclone trajectories as follows:

1. For a given storm, it is examined whether there are cyclones that appear on the following day at a distance of less than 800 km.
2. If there is no such cyclone, then the trajectory is considered to have stopped. Otherwise, in 95% of the cases there is only one appropriate cyclone in the following day, which is then considered to belong to the same trajectory as the initial storm, and the first step is repeated. For the remaining 5% of cases, there is more than one cyclone within 800 km, and a first preference is given to the cyclones located in the southwestern quadrant relative to the initial cyclone. Next, the closest cyclone is chosen as belonging to the same trajectory as the initial cyclone. The first step is then repeated.

To be considered as a model tropical cyclone trajectory, a trajectory must last at least 2 days and have a maximum wind velocity within an 8° circle centered on the middle of the cyclone, which must be larger than 17 m/s during at least 2 days.

5.2.3. Example of tracked vortices

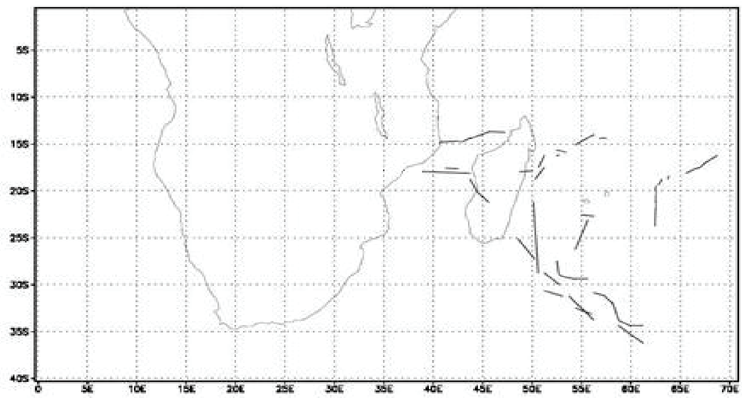

Figure 5.2. An example of tracks of tropical cyclone-like vortices detected when applying the objective procedure for tracking cyclones in the ECHAM4.5-RegCM3 output.

Figure 5.2 shows the tracks of simulated tropical cyclone-like vortices of one of the ensemble members obtained from the ECHAM4.5-RegCM3 system. The period of simulation is from 1 December 1993 until February 1994 (November 1993 was used for model spin-up).

5.3. Fine resolution GCM

5.3.1. The CCAM

The CCAM forecasting system described earlier is used to predict tropical system tracks for the summer season of 2007/2008. The ensemble forecast has 4 members, which are initialized on 4 consecutive days (14, 15, 16 and 17 November 2007). Initial conditions are obtained from the 0Z analysis fields provided by the Global Forecasting System (GFS). The model runs at C48 (approximately 200 km) horizontal resolution in a quasi-uniform grid. Persisted SSTs, obtained from the GFS, are used as lower boundary forcing. Forecast output is available on a 1°x1° latitude-longitude grid.

5.3.2. Identification and tracking of tropical systems

The oceanic area of interest in the track analysis is the Indian Ocean west of 70°E, south of 5°S and north of 35°S. All tropical circulation systems (tropical storms or tropical cyclones) as forecast by CCAM over this region during December 2007 to February 2008 are identified and their tracks compared to the corresponding observed tropical system tracks. NCAR-NCEP reanalysis data (Kalnay *et al*., 1996) is used to identify the observed tracks. Note that in the tracking methodology employed here, a tropical system is defined as a closed mid-level low that exists for at least 5 consecutive days. In the case of CCAM data the track analysis is carried out at the 750 hPa level, with the 600 hPa level used in the case of NCEP data.

The CCAM data are regridded to the 2.5° resolution of the NCEP reanalysis data before the analysis of forecast tracks is performed. The reason for this is twofold:

1. Tracking lows objectively is cumbersome at high spatial resolution (Blender and Schubert 2000) – primarily because of the presence of short-lived, mesoscale lows at these resolutions.
2. The nature of cyclone tracks determined from objective procedures is dependent on the spatial resolution at which the analysis is performed (Blender and Schubert 2000). It is therefore essential for the forecast and observed tracks to be identified from data sets of similar resolution.

The tracking method is based on identifying local geopotential minima at each time interval of a given data set. The movement of each of these minima is then tracked by making use of a nearest-neighbour approach (e.g. König *et al*., 1993; Sinclair, 1994; Blender *et al*., 1997).

5.3.3. Discussion

During DJF of the 2007/8 season, four tropical systems were observed to have reached tropical cyclone status over the area of interest, with another two systems reaching tropical storm status (according to the National Climatic Diagnostic Center (NCDC) classification – see Table 5.1). When applying the tracking methodology to the NCEP reanalysis data, four of these systems are identified and tracked. Note that the tracking methodology specifies that only systems having a lifetime duration of at least 5 days are identified as tropical systems (this condition is necessary to eliminate the numerous short-lived local geopotential minima that constantly develop and decay in the subtropics). This criterion disqualified the tropical cyclone Fame as well as tropical cyclone Hondo from being tracked. In the case of Fame, the system weakened whilst over Madagascar, and temporarily lost the property of being a local geopotential minimum (at least in the NCEP reanalysis data). Hondo developed to the east of the forecast area, existed within the area of interest for less than 5 days, and was therefore not identified as a tropical system by the tracking methodology followed. Additionally, a fifth tropical system is identified from the tracking methodology applied to the NCEP data – this system lived for a period of more than five days but didn't attain the intensity needed to be classified as a tropical storm.

Tropical system	**Cyclogenesis date**	**Cyclolysis date**	**Category**
Celina	20071213	20071217	Tropical storm
Elnus	20080101	20080102	Tropical storm
Fame	20080125	20080201	Tropical cyclone
Gula	20080127	20080201	Tropical cyclone
Hondo	20080205	20080212	Tropical cyclone
Ivan	20080207	20080218	Tropical cyclone

Table 5.1. Tropical storms and cyclones that occurred west of 70˚E and south of 5°S over the Indian Ocean during DJF 2007/8, as identified by the NCDC. (http://www.ncdc.noaa.gov/oa/climate/research/2008/2008-south-indian-trop-cyclones.html).

The tropical system tracks identified from the NCEP reanalysis data and the 4 CCAM ensemble members are displayed in Figure 5.3. Note that the systems are tracked from the first moment in time that they developed closed circulation (a local geopotential minimum), rather from the moment that they have obtained tropical storm or tropical cyclone status). The systems are tracked until the last point in time that closed circulation existed at the given spatial resolution. The result is that the tracks shown are mostly longer (in time and space) than the tracks of the associated tropical storms and cyclones (this is true for tropical storm Elnus, in particular).

The four CCAM ensemble members predicted the occurrence of five (member 1, red lines), six (member 2, blue lines), seven (member 3, green lines) and five (member 4, gray lines) tropical systems during the DJF period under consideration, respectively. This compares well to the observed frequency of five systems as identified from the NCEP data. Of the observed tropical systems, four developed over the northeastern part of the area of interest, or entered the area from the northeast. Three of these systems followed southwesterly tracks over the Indian Ocean, with two systems (tropical cyclones Ivan and Celina) making landfall over Mozambique. A fifth system was observed to develop in the Mozambique Channel. The plume of tracks forecast by the CCAM system generally has more northerly locations than the observed tracks, and the forecast systems also generally developed to the north of the observed locations of development. However, ensemble member 2 shows two systems making landfall over Mozambique, whilst cyclogenesis within the Mozambique Channel is forecast by members 1 and 4 (representative of the observed tropical storm Elnus). Members 1 and 3 also forecast systems with a tendency to make landfall over northern Mozambique and Tanzania, a feature not present in the observed data.

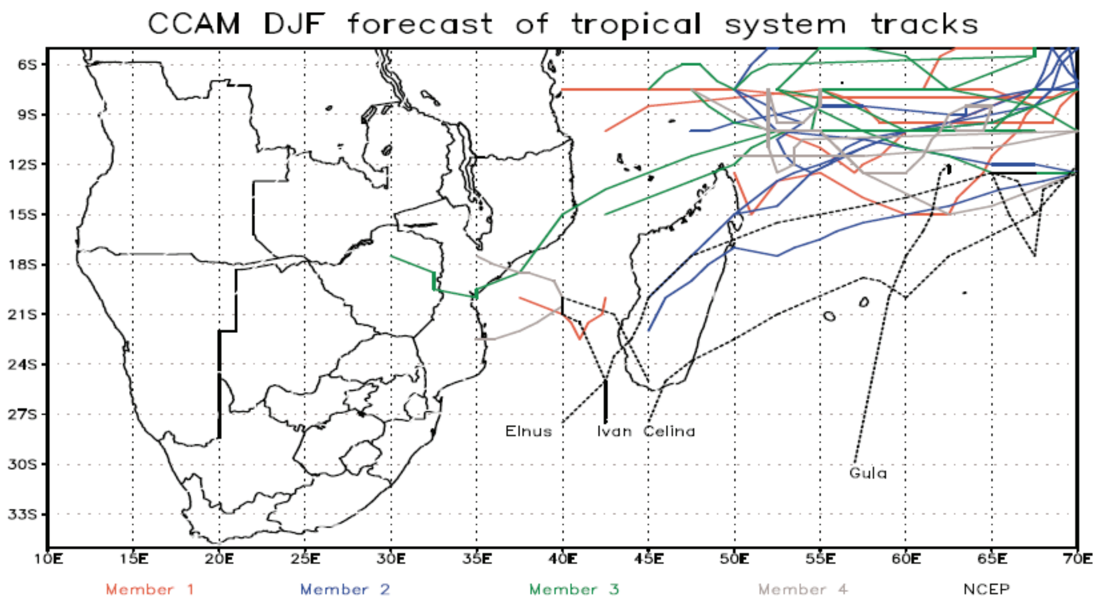

Figure 5.3. Tracks of observed (black lines) and predicted tropical systems during DJF 2007/08.

The CCAM seasonal forecasting system managed to capture some of the most important properties of the observed 2007/08 DJF observed tropical system tracks (a tendency for systems to develop in the Mozambique Channel and for systems to make landfall over Madagascar). However, most forecast tracks have locations further to the north than is observed. In this study the CCAM forecasts consisted of only four ensemble members. Recently, the operational system was extended to having eight members, with plans to increase the number of members to twenty-four during 2009. This would increase the potential of the system to capture more completely the range of uncertainty associated with observed tropical system tracks, and it would also make feasible assigning forecast probabilities to different categories of tracks. Model skill in predicting tropical system track behavior can also be quantified, for example by statistically comparing simulated tracks in AMIP-type runs to the corresponding observed tracks.

5.4. Synopsis

Tropical vortex detection and tracking applied to the output from two different physical model configurations have been investigated. The tracked paths of these vortices seem realistic so that these tracking techniques can be employed in an operational environment to generate probabilistic estimates of tropical cyclone occurrence over the south-western Indian Ocean and the likelihood of their landfall.

Much more work needs to be dedicated to the development and testing of a forecast system for tropical storm occurrence on a seasonal time scale, but the results presented here provide the evidence that such a study is feasible.

CHAPTER 6
MULTI-MODEL SEA-SURFACE TEMPERATURE PREDICTION

6.1. Background

A number of models exist for making long-range forecasts of SST anomalies and most notably forecasts of the equatorial Pacific SST because of its importance to inter-seasonal climate variability worldwide. Notwithstanding, there is a need to predict the seasonal to interannual variability of ocean areas other than the equatorial Pacific Ocean since global SST forecasts are required to provide boundary conditions of atmospheric GCMs. Moreover, in many areas of Africa, for example, climate variability is significantly affected by ocean areas other than the equatorial Pacific Ocean.

The SAWS and UP operationally run only atmospheric GCMs at present, so there is a need to predict the SST anomalies to force these models. Currently the CCAM produces 8 ensemble members when forced with SST anomalies that are obtained by persisting the most recent month's anomalies, and the ECHAM4.5 only 6 members. An additional 6 ECHAM4.5 forecasts are produced by forcing the model with predicted SST anomalies obtained every month from the IRI. These predicted SST fields are downloaded from an anonymous ftp site (crunch.ldeo.columbia.edu/pub/xgong/IRI_SSTa_fcst/individual_model). Three sets of forecasts from three different forecasting systems are available, but the average forecast is used to force the ECHAM4.5 run at the SAWS. The three forecast sets are generated by the NCEP CFS (Saha *et al*., 2006), the LDEO5 (Chen *et al*., 2004) and through a constructed analogue statistical model (Van den Dool, 1994).

6.2. A combination experiment

The SST archived forecasts from three coupled GCMs (UKMO, ECMWF and Météo-France) are combined here using CCA and equal weighting. Output from the three models is once again taken from the DEMETER archive (Palmer *et al*., 2004), but forecasts produced at only two initialization dates (August and November) are considered. The purpose of this chapter is therefore not to fully explore the predictability of SST anomalies using the DEMETER data, but rather to demonstrate the potential in combining forecasts of SSTs into one combined global field. DEMETER model forecasts, evaluated over the period 1959 to 2001 and at a 2.5°x2.5° horizontal resolution, are projected linearly to the 2°x2° resolution of the extended reconstructed SST data set (Smith and Reynolds, 2004). The skill of the models is expressed as Spearman rank correlations and is shown in Figure 6.1. High forecast skill is found over the tropical oceans, and in particular the equatorial Pacific Ocean. Combining the three models using CCA produces similar skill patterns (Figure 6.2).

Most of the multi-model forecasts for rainfall produced by simply averaging the forecasts from individual systems lead to higher skill than combining the forecasts using CCA (Chapter 3). This notion is also tested here on single and multi-model forecasts of equatorial SSTs. The areas to be tested are the Nino3.4 region, the equatorial Indian Ocean (48°E-95°E; 5°N-5°S) and the equatorial Atlantic (35°W-Greenwich; 5°N-5°S). An SST index is calculated for each equatorial domain by simply averaging the grid-point values for each domain and then standardizing these averaged values. Figure 6.3 shows the mean-squared-error skill scores (MSESS; Wilks, 2006) of the equatorial SST indices by using two multi-model combination schemes (CCA and equal weighting) and for the individual models. For Nino3.4 and Indian Ocean SST indices the multi-model using equal weighting produced the best results. For the equatorial Atlantic the UKMO fared the best, and for all three cases the equal weighting outscores combinations with CCA.

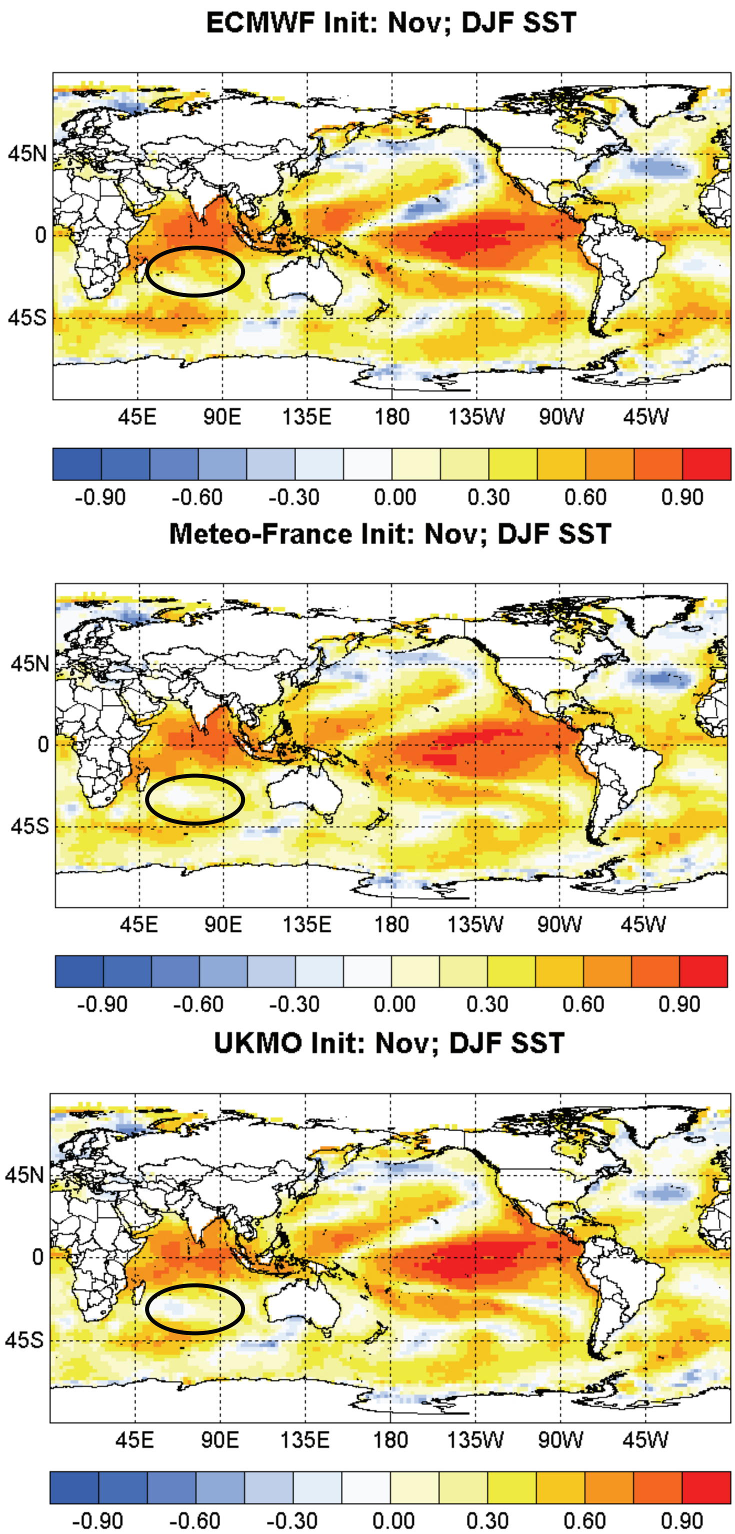

Figure 6.1. Forecast skill (Spearman rank correlations) of the ECMWF (top panel), Météo-France (middle panel) and UKMO (bottom panel) for model initialized in November to produce DJF SST forecasts. The black ellipse identifies the Indian Ocean south-east of Madagascar.

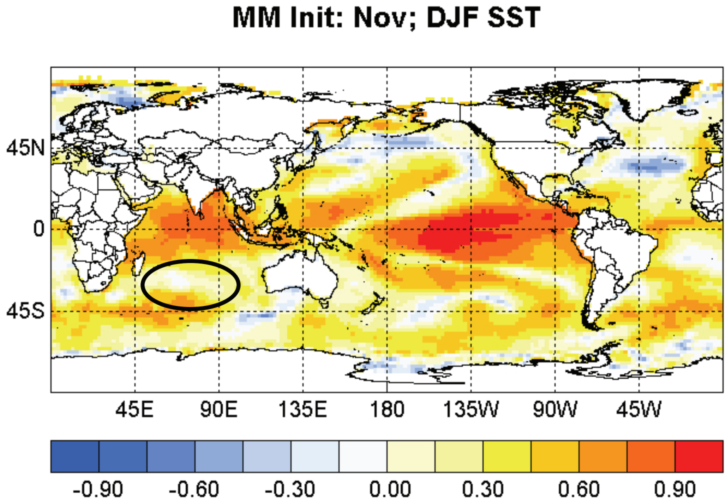

Figure 6.2. Forecast skill (Spearman rank correlations) of the multi-model system. Forecasts of the ECMWF, Météo-France and UKMO coupled models are combined with CCA. Model initialized is in November to produce DJF SST forecasts. The black ellipse identifies the Indian Ocean south-east of Madagascar.

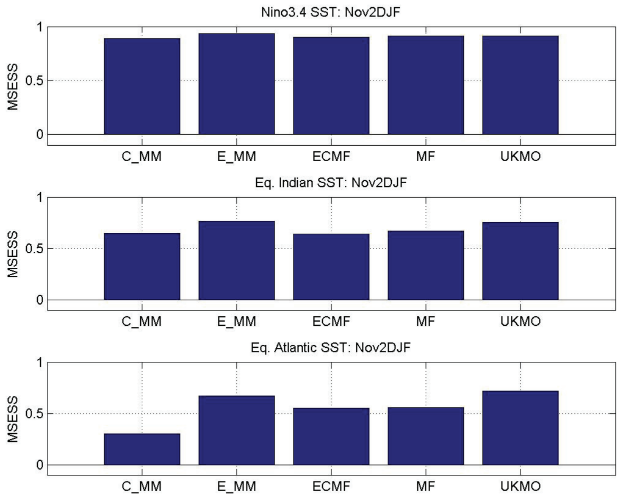

Figure 6.3. Forecast skill (MSESS) associated with two multi-model combination schemes and three individual models for the Nino3.4 region and the equatorial domains of the Indian and Atlantic Oceans. Model initialized in November to produce DJF SST forecasts. C_MM: multi-model combination using CCA; E_MM: multi-model combination using equal weighting.

Similar results are also found predicting January SST from August and November initial states (not shown). However, January SST forecasts initialized in August produces slightly lower skill than forecasts initialized in November. In addition to the lower skill found at longer forecast lead-times, the predictability of equatorial Indian Ocean SST is lower than the predictability of Nino3.4 SST, but is in turn more predictable than the equatorial SST of the Atlantic Ocean. These results are not uncommon (Landman and Mason, 2001).

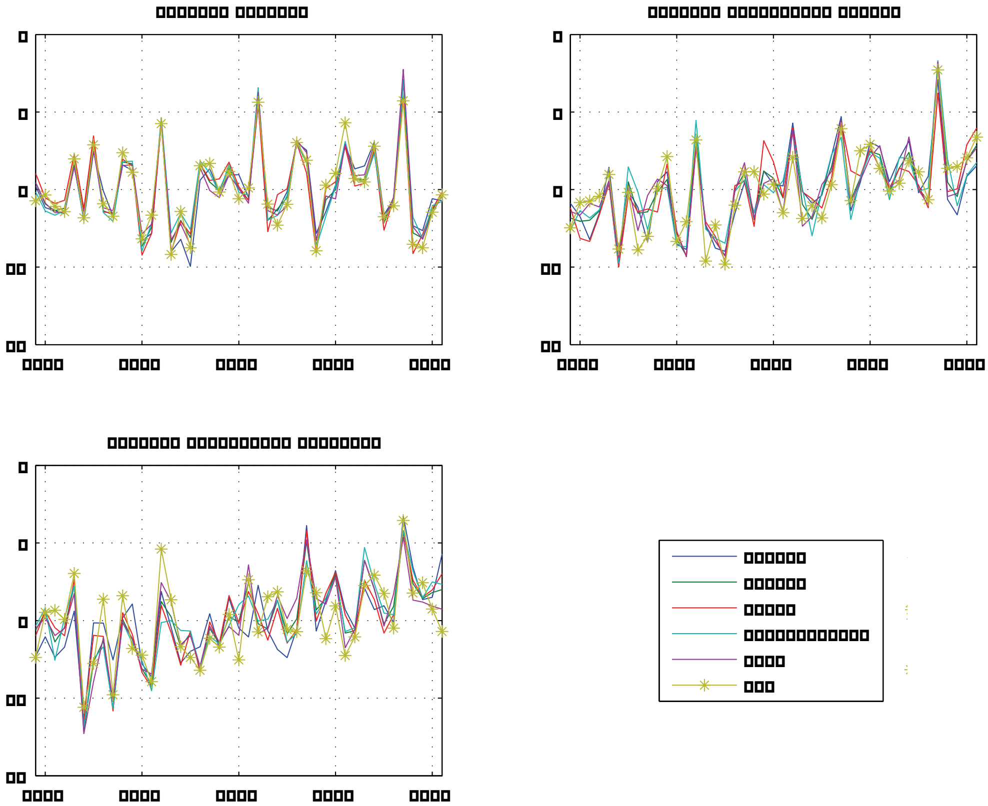

Figure 6.4. Forecast and observed index time series of the Nino3.4 and the equatorial domains of the Indian and Atlantic Oceans. The forecast time series comprise of forecasts produced by a CCA-multi-model, equal-weights-multi-model, ECMWF, Météo-France and UKMO coupled models.

The time series (indices) of the predicted equatorial SST indices are shown in Figure 6.4, and the correlation between forecast and observed indices in Figure 6.5. All the forecasts of the single models and multi-models show good agreement with the observed, and correlations confirm what has already been shown above, that the highest predictability is found over the equatorial Pacific Ocean, followed by the equatorial Indian and Atlantic Ocean domains. A linear upward trend (warming) is seen for both the Indian and Atlantic Ocean basins (Figure 6.4). Detrending the forecast and observed time series by removing the best straight-line fit linear trend from the data and then recalculating the correlations produced only small differences between the two (Figure 6.5). Even though the largest differences are found for the equatorial Indian Ocean domain, the skill seen over these basins is therefore not a result of the trends in the data.

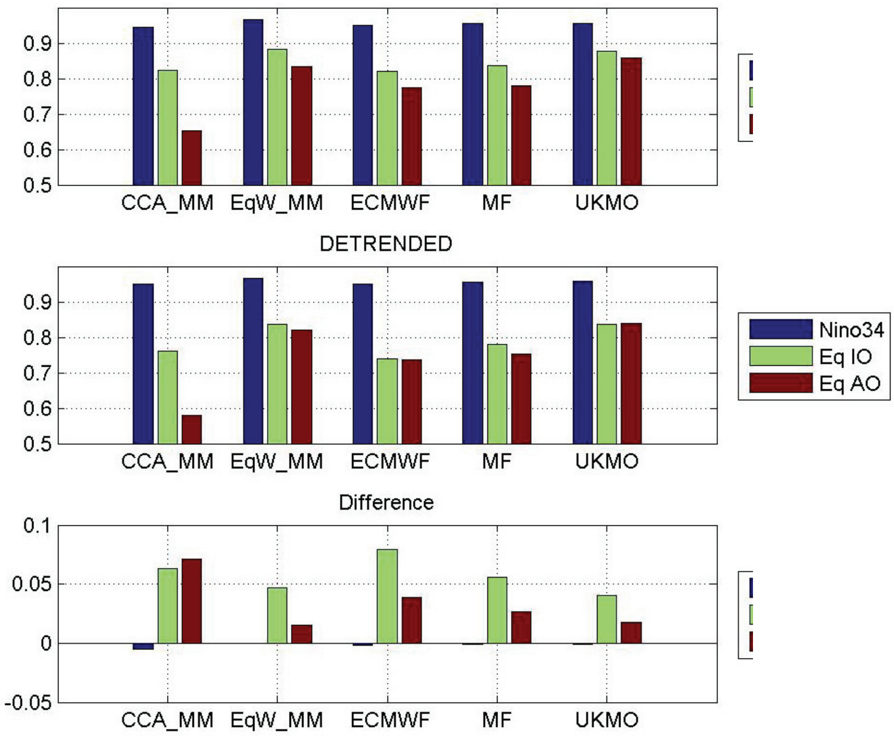

Figure 6.5. Correlations (top and middle panel) between forecast and observed SST indices of the three equatorial domains. The middle panel shows correlations after detrending the data, and the bottom panel shows the correlation differences between the values of the top and middle panel. Forecasts are for DJF initialized in November. CCA_MM: multi-model combination using CCA; EqW_MM: multi-model combination using equal weighting.

6.3. A possible problem identified: consequence for rainfall prediction

The ocean area south-east of Madagascar is given special attention in this section. According to Figures 6.1 and 6.2 very little predictability of SST anomalies resides over this area. Low skill is also found there using statistical models (Landman and Mason, 2001). However, this area is understood to be of significant importance in the understanding and predictability of southern African summer rainfall variability. In fact, in exploring the role of Indian Ocean SST in contributing to extremely wet years over southern Africa, it was found that the optimal atmospheric GCM response to Indian Ocean SST is obtained when the model is forced with anomalously warm SSTs over the Indian Ocean south-east of Madagascar and cold anomalies further north (Washington and Preston, 2006). Since this area has been identified as important to predict rainfall extremes over the region and even sophisticated forecast models such as fully coupled GCM are not skilful in predicting the SST anomalies there, modelling efforts may have to be directed towards improvement in forecast skill over this part of the Indian Ocean. This effort may be jeopardised by the lack of sustainable surface and sub-surface observational systems there.

6.4. Synopsis

In order to optimally force atmospheric GCMs to make skilful seasonal predictions at lead-times, the forcing SST fields need to be skillfully predicted. This notion is of particular importance to centres who have the infrastructure to run global models operationally, but who do not have the capabilities to run fully coupled models. The SAWS and UP are such centres. Moreover, the SAWS has an operational mandate to issue forecasts and therefore needs to constantly improve on existing forecasting systems.

In this chapter ensemble mean forecasts of DJF SSTs are combined using CCA and equal weighting. Single and multi-models produce skilful forecasts over the equatorial regions. Although statistical models and persistence as a forecast do equally well over these regions (Landman and Mason, 2001), and in particular for the lead-time of 1-month presented here, the goal is to illustrate the ability to combine such forecasts and to test if the use of the most simple combination scheme has merit. In fact it has, and the results show that simply averaging the forecasts outscores combination using CCA. Moreover, predictability of the SST of equatorial Indian and Atlantic Ocean domains is real and not a mere consequence of the observed and forecast warming trends in the data.

The ability to predict SSTs have been demonstrated for a large part of the global oceans. However, areas over the south-western Indian Ocean shown to be of importance to the understanding and predictability of southern African rainfall variability may not be as predictable as areas such as the tropical oceans. Further modelling research has to be undertaken and focused on these areas in order to further improve on forecast skill.

CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

This project has been of significant importance to the operational structures of the SAWS and has lead the way for further development of objective seasonal rainfall (and possibly temperature) forecasting systems for southern Africa. The performances of a number of physical models and also an empirical model have been investigated and tested against their combined simulations and forecasts. Some of the results presented are not new, for example that most of the seasonal predictability is found during the season when tropical circulation dominates the atmosphere, but, very importantly, the conclusion can be made that even simple model combination improves on the forecast skill associated with individual models. However, this is only true under certain circumstances, which includes the use of long training periods and that only combining the best forecast models will improve on forecast skill. For example, including an empirical model that relates SSTs to rainfall does not necessarily improve on skill since this model has relatively low skill compared to the GCM-MOS models.

The geographical distribution of skill shows similarities with earlier work that showed that during the season of highest rainfall predictability most of the skill is found over the north-east and central-western regions (e.g. WRC Report No. 1334/1/06). Such a skill pattern is in close agreement with the areas over South Africa that show a strong association with mid-summer equatorial Pacific SSTs. This agreement highlights the difficulty even state-of-the-art models have in predicting rainfall over southern Africa when there is no significant forcing evident from the equatorial Pacific Ocean. Simply stated, no ENSO, no skill!

Southern Africa is affected by tropical cyclones, especially during landfall. The report has shown the potential in predicting tracks of tropical vortices through the application of detection and tracking algorithms applied to physical model output. Operational tropical cyclone prediction on a seasonal time scale is non-existent in South Africa, but needs to be addressed. Tropical cyclone landfall significantly influences streamflow, and the report has shown that streamflow measured in rivers is potentially predictable. However, large runoff caused by tropical cyclone-associated rainfall is not captured by the multi-model presented here.

For atmospheric GCMs to make skilful seasonal predictions at lead-times, the forcing SST field needs to be described properly. It has been demonstrated here that high prediction skill for SSTs is found over large ocean areas, albeit at short lead-times (most certainly at longer lead-times too). The best skill is found when the SST forecasts from the different models are simply averaged. Notwithstanding, the prediction skill over certain ocean areas needs to be further improved in order to improve on GCM forecasts of rainfall over the region. This study did not address the best configuration of a two-tiered system comprising of predicted SSTs and a GCM. Modellers need to investigate whether to either use the best single SST global forecast (e.g. a multi-model forecast), or to use the individual SST forecasts to force a GCM and therefore include into the forecasting system uncertainties associated with the boundary conditions.

As an operational forecasting centre the SAWS has to constantly improve on existing forecasting systems. The report is evidence of the work that has gone into the development of new state-of-the-art forecasting techniques for southern Africa's seasonal climate variability. Moreover, this work has shown the way to develop objective multi-model forecasting techniques that not only address seasonal rainfall and streamflow prediction, but also the prediction of additional parameters such as minimum and maximum temperatures, and the number of rain days exceeding certain predetermined thresholds.

REFERENCES

Barnett, T. P., and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States air temperature determined by canonical correlation analysis. *Monthly Weather Review*, **115**, 1825-1850.

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *Journal of Climate*, **7**, 1513-1564.

Bengtsson, L., M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus*, **47A**, 175-196.

Blender, R., and M. Schubert, 2000: Cyclone tracking in different spatial and temporal resolution. *Mon. Wea. Rev.,* **128,** 377-384.

Blender, R., K. Fraedrich, and F. Lunkeit, 1997: Identification of cyclone-track regimes in the North Atlantic. *Quarterly Journal of the Royal Meteorological Society,* **123,** 727-741.

Branković, Č., and T. N. Palmer, 2000: Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2035-2067.

Brown, B. H., and A. H. Murphy, 1996: Improving forecasting performance by combining forecasts: the example of road-surface temperature forecasts. *Meteorological Applications*, **3**, 257-265.

Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Nino in the past 148 years. *Nature*, **428** (6984), 733-736.

Coelho, C. A. S., D. B. Stephenson, M. Balmaseda, F. J. Doblas-Reyes, G. J. van Oldenborgh, 2006: Toward an integrated seasonal forecasting system for South America. *Journal of Climate*, **19**, 3704-3721.

Dirmeyer, P. A., M. J. Fennessy, and L. Marx, 2003: Low skill in dynamical prediction of boreal summer climate: Grounds for looking beyond sea surface temperature. *Journal of Climate*, **16**, 995-1002.

Doblas-Reyes, F. J., M. Déqué, J. –P. Piedelieve, 2000: Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2035-2067.

Evans, R. E., M. S. J. Harrison, R. J. Graham, and K. R. Mylne, 2000: Joint medium-range ensembles from the Met. Office and ECMWF systems. *Monthly Weather Review*, **128**, 3104-3127.

Graham, R. J., A. D. L. Evans, K. R. Mylne, M. S. J. Harrison, and K. B. Robertson, 2000: An assessment of seasonal predictability using atmospheric general circulation models. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2211-2240.

Goddard, L. and S. J. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Climate Dynamics*, **19**, 619-631.

Jackson, J. E., 1991: *A User's Guide to Principal Components*, Wiley, New York, pp. 569.

Joubert, A. M., and B. C. Hewitson, 1997: Simulating present and future climates of southern Africa using general circulation models. *Progress in Physical Geography*, **21**, 51-78.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77,** 437-472.

Kgatuke, M.M., Landman W.A. and Mbedzi, M.P., 2006: Predicting South African DJF 2006/07 rainfall with the RegCM3. *COLA Experimental Long-Lead Forecast Bulletin*, **15**, http://grads.iges.org/ellfb/Dec06/kgatuke/kgatuke.pdf.

König W., R. Sausen, and F. Sielmann, 1993: Objective identification of cyclones in GCM simulations. *Journal of Climate,* **6,** 2217-2231.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate*, **13**, 4196-4216.

Landman, W. A., and L. Goddard, 2002: Statistical recalibration of GCM forecasts over southern Africa using model output statistics. *Journal of Climate*, **15**, 2038-2055.

Landman, W. A., and L. Goddard, 2003: Model output statistics applied to multi-model ensemble forecasts for southern Africa. *Proceedings of the Seventh International Conference on Southern Hemisphere Meteorology and Oceanography*, 24 – 28 March 2003, Wellington, New Zeeland, pp. 249-250.

Landman, W. A., and L. Goddard, 2005: Predicting southern African summer rainfall using a combination of MOS and perfect prognosis. *Geophysical Research Letters*, **32**, L15809, doi:10.1029/2005GL022910.

Landman, W. A., and S. J. Mason, 1999: Operational long-lead prediction of South African rainfall using canonical correlation analysis. *International Journal of Climatology*, **19**, 1073-1090.

Landman, W. A., and S. J. Mason, 2001: Forecasts of near-global sea-surface temperatures using canonical correlation analysis. *Journal of Climate*, **14**, 3819-3833.

Landman, W. A., A. Seth, and S. J. Camargo, 2005: The effect of regional climate model domain choice on the simulation of tropical cyclone-like vortices in the southwestern Indian Ocean. *Journal of Climate*, **18**, 1263-1274.

Landman, W. A., S. J. Mason, P. D. Tyson, and W. J. Tennant, 2001: Retro-active skill of multi-tiered forecasts of summer rainfall over southern Africa. *International Journal of Climatology*, **21**, 1-19.

Mason, S. J., 1995: Sea-surface temperature – South African rainfall associations, 1910-1989. *International Journal of Climatology*, **15**, 119-135.

Mason, S. J., A. M. Joubert, 1997: Simulated changes in extreme rainfall over southern Africa. *International Journal of Climatology*, **17**, 291-301.

Mason, S. J., and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *Journal of Climate*, **15**, 8-29.

Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, L. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bulletin of the American Meteorological Society*, **80**, 1853-1873.

McGregor, J. L., 1996: Semi-Lagrangian advection on conformal-cubic grids. Monthly Weather Review, 124, 1311-1322.

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

Palmer, T. N., and D. L. T. Anderson, 1994: The prospects of seasonal forecasting – a review paper. *Quarterly Journal of the Royal Meteorological Society*, **120**, 755-793.

Palmer, T. N., and Others, 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-85-6-853, 853-872.

Reason, C. J. C., and M. Rouault, 2005: Links between the Antarctic Oscillation and winter rainfall over western South Africa. *Geophysical Research Letters*, **32**, L07705, doi:10.1029/2005GL022419.

Reason, C. J. C., W. Landman, and W. Tennant, 2006: Seasonal to decadal prediction of southern African climate and its links with variability of the Atlantic Ocean. *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-87-7-941, 941-955.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate*, **7**, 929-948.

Roeckner, E., K. Arpe, L. Bengtsson, S. Brinkop, L. Dumenil, M. Esch, E. Kirk, F. Lunkeit, M. Ponater, B. Rockel, R. Sausen, U. Schlese, S. Schubert, and M. Windelband, 1996: Simulation of present-day climate with the ECHAM4 model: Impact of model physics and resolution. Report No. 93, Max-Planck-Institut fur Meteorologie, Hamburg, Germany, 171 pp.

Saha, S., and Others, 2006: The NCEP climate forecasting system. *Journal of Climate*, **19**, 3483-3517.

Schulze, R.E., Warburton, M., Lumsden, T.G. and Horan, M.J.C. 2005. The Southern African Quaternary Catchments Database: Refinements to, and Links with, the ACRU System as a Framework for Modelling Impacts of Climate Change on Water Resources. In: Schulze, R.E. (Ed) Climate Change and Water Resources in Southern Africa: Studies on Scenarios, Impacts, Vulnerabilities and Adaptation. Water Research Commission, Pretoria, RSA, WRC Report 1430/1/05. Chapter 8, 111 - 139.

Schulze, R.E., Lumsden, T.G., Horan, M.J.C., Warburton, M. and Maharaj, M. 2005. An Assessment of Impacts of Climate Change on Agrohydrological Responses Over Southern Africa. In: Schulze, R.E. (Ed) Climate Change and Water Resources in Southern Africa: Studies on Scenarios, Impacts, Vulnerabilities and Adaptation. Water Research Commission, Pretoria, RSA, WRC Report 1430/1/05. Chapter 9, 141 - 189.

Shongwe, M. E., W. A. Landman, and S. J. Mason, 2006: Performance of recalibration systems for GCM forecasts for southern Africa. *International Journal of Climatology*, **26**, 1567-1585.

Sinclair, M.R., 1994: An objective cyclone climatology of the Southern Hemisphere. *Mon. Wea. Rev.,* **123,** 1601-1619.

Smith, T. M., and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854-1997). *Journal of Climate*, **17**, 2466-2477.

Smith, T. M., R. W. Reynolds, R. E. Livezey, D. C. Stokes, 1996: Reconstruction of historical sea surface temperature using empirical orthogonal functions. *Journal of Climate*, **9**, 1403-1420.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Weather and Forecasting*, **8**, 379-398.

Tyson PD, Preston-Whyte RA., 2000: *The Weather and Climate of Southern Africa*. Oxford University Press.

Van den Dool, H.M., 1994: Searching for analogues, how long must we wait? Tellus, **46A**, 314-324.

Vijaya Kumar, T. S. V., T. N. Krishnamurti, M. Fiorino, and M. Nagata, 2003: Multimodel superensemble forecasting of tropical cyclones in the Pacific. *Monthly Weather Review*, **131**, 574-583.

Vitart, F., and J. L. Anderson, and W. F. Stern, 1997: Simulation of inter-annual variability of tropical storm frequency in an ensemble of GCM integrations. *Journal of Climate*, **10**, 745-760.

Walsh, K., 1997: Objective detection of tropical cyclones in high-resolution analyses. *Monthly Weather Review*, **125**, 1767-1779.

Washington, R., and A. Preston, 2006: Extreme wet years over southern Africa: Role of Indian Ocean sea surface temperatures. *Journal of Geophysical Research*, **111**, D15104, doi:10.1029/2005JD006724.

Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, **21**, 530-548.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences. Second Edition*. Academic Press, San Diego, pp. 627.