Downscaling of Global Circulation Model Predictions to Daily Rainfall over the Upper Olifants River Catchment

by

Abraham Stephanus Steyn

Submitted in partial fulfilment of the requirements for the degree of

Magister Scientiae in Agricultural Meteorology

in the

Department of Soil, Crop and Climate Sciences

Faculty of Natural and Agricultural Sciences

University of the Free State

Supervisor: Prof. Sue Walker

Co-supervisor: Dr. Francois A. Engelbrecht

Bloemfontein

December 2008

CONTENTS

DECLARATION

I declare that this thesis hereby submitted for the degree of Magister Scientiae in Agricultural Meteorology at the University of the Free State is my own independent work and has not previously been submitted by me at another university or faculty.
I further more cede copyright of this thesis in favour of the University of the Free State.

_____

A.S. Steyn

Signed at the University of the Free State, Bloemfontein
Republic of South Africa
December 2008

ABSTRACT

Downscaling of Global Circulation Model Predictions to Daily Rainfall over the Upper Olifants River Catchment

Abraham Stephanus Steyn

M.Sc. in Agrometeorology at the University of the Free State
December 2008

Climate change could have far reaching consequences for all spheres of life. Continued greenhouse gas (GHG) emissions at or above current rates will cause further warming and induce further changes in the global climate system. This is particularly true for southern Africa where an ever-increasing population is already causing an increase in the demand for fresh water and much of the agricultural food production depends on rain.

Global Circulation Models (GCMs) are the main source of climate projections under varying GHG emission scenarios. The spatial resolution of GCMs is too coarse to resolve sub-grid processes such as convection and precipitation. However, agrohydrological application models often require information at a network of point locations, implying the need to downscale the GCM output. Downscaling approaches have subsequently emerged as a means of employing large-scale atmospheric predictor variables (such as the 500 hPa meridional velocity) to develop station-scale meteorological series. Variables such as daily rainfall, which are not always accurately represented by the GCMs, can be derived using statistical approaches to build relationships between the required forecast parameter and variables that are simulated more accurately.

Previous investigators have used the statistical downscaling model (SDSM) to downscale climate projections of daily rainfall over North America and Europe. A similar methodology was adopted to downscale daily rainfall projections under the A2 and B2 emission scenarios at five selected quaternary catchments (QCs) within

the Upper Olifants River catchment. The downscaling was performed for the summer months of December, January and February (DJF).

The set of generic predictors which were identified across all five QCs included surface airflow strength, vorticity, divergence and specific humidity, 850 hPa wind direction and relative humidity as well as 500 hPa relative humidity and meridional wind velocity. Generally, all the predictors exhibited a reasonably low explanatory power. The considerable variation in the resultant correlations between the large-scale predictors and the observed daily precipitation at the selected QCs may very well have stemmed from the convective nature of the rainfall patterns, being irregularly distributed in space and time. Generally, the downscaling model results were not very encouraging as the model failed to produce satisfactory results for four of the five QCs.

For one of the QCs, namely Groblersdal, the projected changes for the future climate were assessed by calculating several delta-statistics. Only a few of the indices revealed a clear change, while most indices exhibited inconsistent changes for DJF across three future periods centred on the 2020s, 2050s and 2080s. Similar changes in the characteristics of the daily rainfall series are projected for the early and mid 21st century under the A2 and B2 scenarios. Differences in the expected GHG forcing under the B2 scenario does not seem to affect any of the rainfall indices differently from the A2 scenario until the late 21st century. It should however be noted that the projected changes are often smaller than the model errors which implies that the downscaling model is simply not sensitive enough for the projected changes to be taken at face value. Therefore the results should only be used with caution. The fact that the downscaling procedure provides similar results for the A2 and B2 scenarios suggests that it is at least to some extent robust and stable.

OPSOMMING

Afskaling van Globale Sirkulasie Model Voorspellings na Daaglikse Reënval oor die Bo-Olifantsrivier Opvanggebied

Abraham Stephanus Steyn

M.Sc. in Landbouweerkunde aan die Universiteit van die Vrystaat
Desember 2008

Klimaatverandering kan verreikende gevolge inhou vir alle vlakke van die samelewing. Volgehoue kweekhuisgas (KHG) vrylatings teen vlakke wat die huidige tempo ewenaar of oorskry, sal verdere verwarming teweeg bring en verdere veranderinge in die globale klimaatstelsel veroorsaak. Dit is veral waar vir suider-Afrika waar ’n steeds groeiende bevolking reeds ’n toename in die vraag na vars water veroorsaak en ’n groot gedeelte van die landboukundige voedselproduksie van reënval afhanklik is.

Globale Sirkulasiemodelle (GSMs) is die hoofbron van klimaatvooruitskouings onder veranderende KHG vrystellingscenario’s. Die ruimtelike resolusie van GSMs is te grof om prosesse soos konveksie en reënval wat kleiner as die roosterveld is te hanteer. Landbou-hidrologiese toepassingsmodelle vereis dikwels inligting by ’n netwerk punte wat dan die behoefte om die GSM uitvoer af te skaal beklemtoon. Afskalingsbenaderings het gevolglik ontluik as ’n middel om groot-skaalse atmosferiese voorspellersvelde (soos die 500 hPa meridionale windspoed) in te span om stasievlak weerkundige reekse te ontwikkel. Veranderlikes soos die daaglikse reënval, wat nie altyd akkuraat deur GSMs voorgestel word nie, kan afgelei word deur middel van statistiese metodes wat verwantskappe vaslê tussen die vereiste parameter en veranderlikes wat meer akkuraat gesimuleer word.

Vorige navorsers het die statistiese afskalingsmodel (SDSM) ingespan om klimaatprojeksies van daaglikse reënval oor Noord-Amerika en Europa af te skaal. ’n Soortgelyke metodologie is aangeneem om daaglikse reënvalprojeksies onder die

A2 en B2 vrystellingscenario's by vyf gekose sub-opvanggebiede binne die Bo-Olifantsrivier af te skaal. Die afskaling is uitgevoer vir die somermaande Desember, Januarie en Februarie (DJF).

Die stel generiese voorspellers, wat oor al vyf sub-opvanggebiede geïdentifiseer is, sluit oppervlak windsterkte, vortisiteit, divergensie en spesifieke humiditeit, 850 hPa windrigting en relatiewe humiditeit asook 500 hPa relatiewe humiditeit en meridionale windspoed in. Oor die algemeen het al die voorspellers relatief lae verklarende vermoëns getoon. Die aansienlike variasie in die gevolglike korrelasies tussen die groot-skaalse voorspellers en die waargenome daaglikse reënval by die gekose sub-opvanggebiede mag teweeg gebring word deur die konvektiewe aard van die reënvalpatrone wat onreëlmatig in tyd en ruimte versprei is. In die algemeen was die afskalingsmodel se resultate nie baie bemoedigend nie aangesien dit gefaal het om aanvaarbare resultate vir vier uit die vyf sub-opvanggebiede te verskaf.

Vir een van die sub-opvanggebiede, naamlik Groblersdal, is die vooruitgeprojekteerde veranderinge vir die toekomstige klimaat geevalueer aan die hand van 'n aantal delta-statistieke. Slegs 'n paar van die indekse het 'n duidelike verandering getoon, terwyl meeste indekse vir DJF onkonsistente veranderings oor drie toekomstige periodes, wat op die 2020s, 2050s en 2080s fokus, getoon het. Soortgelyke veranderinge in die eienskappe van die daaglikse reënvalreeks word onder die A2 en B2 scenario's voorspel vir die vroeë- en mid-21ste eeu. Verskille in die verwagte KHG forserings tussen die A2 en B2 scenario's blyk nie 'n invloed op enige van die reënvalindekse te hê tot die laat 21ste eeu nie. Daar moet gelet word dat die geprojekteerde veranderinge dikwels kleiner is as die modelfoute wat dan impliseer dat die afskalingsmodel eenvoudig nie sensitief genoeg is om die geprojekteerde veranderinge blindelings te aanvaar nie. Die resultate moet gevolglik versigtig gebruik word. Die feit dat die afskalingsprosedure soortgelyke resultate vir die A2 en B2 scenario's lewer toon dat dit ten minste rigied en stabiel is.

ACKNOWLEGDEMENTS

Many thanks and appreciation to:

- My supervisors, Prof. Sue Walker and Dr. Francois Engelbrecht, for their assistance and advice.
- The WRC for funding project number K5/1646 on the “Applications of rainfall forecasts for agriculturally related decision making in selected catchments”.
- The developers of the SDSM and the staff of the Canadian Institute for Climate Studies for supplying the GCM and NCEP data.
- NCEP for reanalysis data provided by NOAA through their website at http://www.cdc.noaa.gov
- Trevor Lumsden from UKZN for supplying the observed QC data.
- My friends and family for their support – they know who they are.
- My daschounds for keeping me company at night.

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AOH	Atlantic Ocean High
ARC	Agricultural Research Council
CAB	Congo Air Boundary
C-CAM	Conformal-Cubic Atmospheric Model
CCCma	Canadian Centre for Climate Modelling and Analysis
CDF	Cumulative Distribution Function
CICS	Canadian Institute for Climate Studies
COADS	Comprehensive Ocean Atmosphere Data Set
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DAI	Data Access Integration
DARLAM	Division of Atmospheric Research Limited-Area Model
DEAT	Department of Environmental Affairs and Tourism
DJF	December, January, February
DREU	Daily Rainfall Extraction Utility
DWAF	Department of Water Affairs and Forestry
EOF	Empirical Orthogonal Function
GCM	General Circulation Model, Global Climate Model
GHG	Greenhouse gas
GIS	Geographical Information System
GMT	Greenwich Mean Time
GTS	Global Telecommunication System
GWR	Geographically Weighted Regression
HadCM3	Third Generation Hadley Centre Coupled Model
IOH	Indian Ocean High
IPCC	Intergovernmental Panel on Climate Change
ITCZ	Inter-tropical Convergence Zone
LAM	Limited-Area Model
LEPS	Linear Error in Probability Space
LSU	Large Stock Unit

MOS	Model Output Statistic
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
PDF	Probability Density Function
PP	Perfect Prognosis
QC	Quaternary Catchment
Q-Q	Quantile-quantile
R^2	Coefficient of determination
RAN	Reanalysis
RCM	Regional Climate Model
SASRI	South African Sugarcane Research Institute
SAWS	South African Weather Service
SE	Standard Error
SDSM	Statistical Downscaling Model
SRES	Special Report on Emissions Scenarios
SSE	Sum of squared errors (between the residuals and their means)
SSR	Regression sum of squares
SST	Total sum of squared deviations (between the residuals and their means)
STARDEX	Statistical and Regional dynamical Downscaling of Extremes for European Regions
UFS	University of the Free State
UP	University of Pretoria
WDC	World Data Centre for Climate
WRC	Water Research Commission

CHAPTER 1 INTRODUCTION

1.1 Background

According to Trenberth *et al*. (2007) global mean surface temperatures have risen by 0.74°C ± 0.18°C when estimated by a linear trend over the period spanning 1906 – 2005. Climate change could have far reaching consequences for all spheres of life as continued greenhouse gas (GHG) emissions at or above current rates will cause further warming and stimulate further changes in the global climate system. This is particularly true for southern Africa where an ever-increasing population is already causing an increase in the demand for fresh water and much of the agricultural food production depends on rain (Walker & Schulze, 2006).

Results from Global Circulation Models (GCMs) are the main source of climate forecasts of various time scales. These dynamical models represent the world as an array of grid-points. However, the spatial resolution of GCMs is too coarse to resolve regional scale effects (Hessami *et al*., 2008). Consequently, sub-grid processes, such as convection and precipitation, are particularly difficult to reproduce, necessitating the parameterisation of these important processes. This implies that locations and variables for which forecasts are required may not be represented explicitly by these models (Maini *et al.*, 2004). In addition, the GCMs have systematic errors and are deterministic. Non-linear responses and the intrinsically chaotic nature of the climate system make the job of climate forecasting that much more problematic (MacKellar *et al*., 2006). It is apparent that – complex and sophisticated as GCMs are – these models are by no means perfect representations of the climate system (MacKellar *et al*., 2006).

For some types of impact assessment (e.g. risk of drought or flooding in large catchments) aerially averaged quantities such as the grid-box variables output from a GCM may be sufficient. However, in many cases information are required at a network of point locations, implying the need to downscale the GCM output (Murphy, 1998). This

is particularly true when the model simulations are required to drive agrohydrological application models. Such models are frequently concerned with small, sub-catchment scale processes, occurring on spatial scales much smaller than those resolved in GCMs (Wilby & Wigley, 1997). The climate-sensitive agricultural sector can benefit from these forecasts by incorporating regional precipitation forecast information into agricultural planning and management strategies (Rossel & Garbrecht, 2001).

Downscaling approaches have subsequently emerged as a means of interpolating large-scale atmospheric predictor variables (such as mean sea-level pressure) to station-scale meteorological series (Wigley *et al*., 1990; Hay *et al*., 1991, cited in Wilby & Wigley, 1997). Variables such as rainfall, which are not always accurately represented by these models, can be derived using statistical approaches to build relationships between the required forecast parameter and variables that are simulated more accurately. Owing to model imperfections, systematic errors may occur. The statistical interpretation of numerical weather prediction forecasts possesses an inbuilt accounting capability for the local topographic and environmental conditions that control the precipitation and other surface weather parameters and can compensate for any model biases (Landman *et al.*, 2001, Maini *et al.*, 2004). Even if global models in future are run at high resolution the need will still remain to 'downscale' the results from such models to individual sites or localities for impact studies (Wilby & Wigley, 1997). Maini *et al.* (2004) found that even for the medium range statistically downscaled forecasts are a definite improvement over direct model output and even have an edge over man-machine mixed forecasts.

This study utilised the statistical downscaling model (SDSM) developed by Wilby *et al*. (2002). The model was calibrated for the summer months of December, January and February and tested with the use of observed datasets of daily rainfall as the predictand and normalised NCEP variables as the predictors. The calibrated model was tested against an independent set of observed daily rainfall data. The model was then used to construct downscaled daily rainfall projections under the A2 and B2 emission scenarios at the quaternary catchment level.

1.2 Objectives of the Research

1.2.1 Problem statement and research question

There is a gap between the spatial resolution at which contemporary GCMs provide their output variables and the resolution required by agrohydrological application models. This implies the need to downscale the GCM output to smaller spatial scales. The research question thus arises: “Is it possible to use statistical methods to effectively downscale GCM data to produce realistic daily rainfall simulations over the Upper Olifants River catchment?”

1.2.2 Objectives

Though the general objective of this study is to develop a method to statistically downscale GCM data to produce daily rainfall simulations over the Upper Olifants River catchment, the following specific objectives were identified:

- To identify quaternary catchments for which the downscaling will be performed;
- To obtain climatological and model data and prepare the data for manipulation;
- To develop a statistical model that will produce downscaled daily rainfall over selected quaternary catchments;
- To compare the GCM projected rainfall with the daily rainfall series of the current climatic period; and
- To compare the GCM projected rainfall under different GHG emission scenarios.

1.3 Organisation of the Report

A taxonomy of downscaling methods are provided in Chapter 2 accompanied by a general review of each downscaling method. A description of the downscaling method used in this study is also furnished. In Chapter 3 the reader is introduced to the study area. This section mainly focuses on the geographical and climatological aspects that are relevant to the study. All the climatological data that were used in this study are described in Chapter 4. The source of the data as well as subsequent manipulations are

discussed. The methodology, which draws from that used by other climate change scenario impact researchers, are described in Chapter 5, followed by a discussion of the downscaling results in Chapter 6. Conclusions regarding the statistical downscaling technique and projected changes in the daily summer rainfall are furnished in Chapter 7. The thesis concludes with a discussion of the proposed future research.

CHAPTER 2 REVIEW OF DOWNSCALING TECHNIQUES

2.1 Introduction

Downscaling activities are normally either spatial or temporal in nature. This spatial or temporal nature usually stems directly from the application of the downscaling procedure. Certain studies require the use of either high resolution gridded data or the use of site-specific data, while other studies may require the use of hourly or daily data, neither of which is catered for by large-scale GCMs.

According to the scientists at the Canadian Institute for Climate Studies (CICS, 2007) spatial downscaling refers to “the techniques used to derive finer resolution climate information from coarser GCM output”. The foundation of spatial downscaling is the assumption that it will be possible to establish significant relationships between the local and large-scale climate (thus allowing important site-scale information to be determined from large-scale information alone) and that these relationships will remain valid under future climate conditions. By integrating some of these regional climate controls, spatial downscaling may be able to add value to coarse-scale GCM output in some areas, although its effectiveness will be very much dependent on the region and climate data available. Each case will be different and may necessitate the investigation of different downscaling techniques before a suitable methodology is identified – and in some cases it may not be possible to improve upon the coarse-scale simulations by downscaling with currently available methods.

Adhering to the following general recommendations should facilitate the spatial downscaling process (CICS, 2007):

- The GCM being used for spatial downscaling should be able to simulate the atmospheric features which will influence the specific area’s climate quite well e.g. positions of large anticyclones, jet streams and storm tracks.

- The downscaling technique should be based on a climate variable which does not exhibit large sub-grid variations in space i.e. it is better to use a variable such as mean sea level pressure rather than one such as precipitation.
- The variables used in the downscaling process should also ideally be direct model output (e.g. sea level pressure) and not be based on parameterisations involving other model variables, as is the case with precipitation.

According to Murphy (1998) any viable downscaling technique must also consider regional forcings (arising from orography, coastlines, lakes, land surface characteristics, etc.) known to influence local climate.

Temporal downscaling refers to "the derivation of finer-scale temporal data from coarser-scale temporal information e.g. daily data from monthly or seasonal information" CICS (2007). Its main application is in scenario impact studies, particularly for the derivation of daily scenario data from monthly or seasonal scenario information. Monthly model output is available from many GCM runs, whilst only a small number of these have archived daily model output. Daily output is also not considered to be as robust as model output at the monthly or seasonal time scales and so is not generally recommended for use in scenario impact studies. The most straightforward method for obtaining daily data for a particular climate change scenario is to relate the monthly or seasonal changes to a historical daily weather record from a particular station. In this way the current observed climate variability and matching sequences of wet and dry days can be emulated, thus assuming that the wet and dry day sequencing does not change.

2.2 Classification of Techniques

Drawing from reviews by Hewitson and Crane (1996), Wilby and Wigley (1997), Murphy (1998), Wilby *et al*. (2002), the Canadian Institute for Climate Studies (2007), Wilby and Dawson (2007) and Hessami *et al*. (2008), downscaling methods may be grouped into the categories presented in Table 2.1. In reality, some downscaling approaches

embrace the attributes of more than one of these techniques and therefore tend to be hybrid in nature (Wilby & Wigley, 1997).

Table 2.1: Classification of downscaling methods

Statistical Downscaling	Dynamical Downscaling
Empirical methods	Limited-area modelling
Weather pattern-based approaches	Stretched-grid modelling
Stochastic weather generators	
Regression-based methods	

2.3 Statistical Downscaling

As a nonlinear dynamical system, the atmosphere is not perfectly predictable in a deterministic sense. A large portion of weather forecasting has a statistical basis and, therefore, statistical methods are useful, and indeed necessary parts of the forecasting endeavour (Wilks, 1995). Statistical downscaling is based on the fundamental assumption that regional climate is conditioned by both the local physiographic features as well as the large scale atmospheric state (Hessami *et al.*, 2008). On this basis, large scale atmospheric fields are related to local variables through a statistical model in which GCM simulations are used as input for the large scale atmospheric variables (or "predictors") to downscale the local climate variables (or "predictands") with the use of observed climatic data. Most statistical downscaling work has concentrated on predicting the rainfall and temperature at a single site as these are the most important input variables for many natural systems models (Wilby *et al*., 2004). The choice of downscaling method is governed by the application and to some extent the nature of the local predictand. According to Wilby *et al*. (2004) issues that need to be considered when attempting statistical downscaling are the choice of downscaling method, the choice of predictors, whether or not extremes should be modelled, whether or not tropical areas are included and possible feedbacks from other climate subsystems.

2.3.1 Empirical methods

In this method the local variable in question (e.g. surface air temperature or precipitation) can be predicted from values of a corresponding variable simulated at nearby GCM grid-points, with empirical adjustments to allow for systematic simulation errors and unresolved subgrid-scale effects (Murphy, 1998). This implies that a linear or non-linear factor can be applied to the GCM simulated predictand in order to derive a "post-processed" predictand. It should be noted that this technique does not comply with the general recommendations as laid out by the Canadian Institute for Climate Studies (2007) since the corresponding variable is bound to exhibit marked sub-grid variations in space. This does not, however, mean that this technique cannot be used in conjunction with another downscaling method such as high resolution modelling as part of a more sophisticated hybrid approach.

Empirical downscaling has successfully been applied to multi-model ensembles consisting of different GCM scenarios in order to explore inter-model similarities and differences (Benestad, 2004). Empirical downscaling requires an adequate record of past observations for the local predictand, which limits the downscaling to locations where there are observations.

2.3.2 Weather pattern-based approaches

These approaches (also referred to as weather typing or the use of analogues) typically involve grouping local, meteorological data in relation to prevailing patterns of atmospheric circulation (Wilby & Dawson, 2007). The weather classification scheme may either be objectively or subjectively derived (Wilby & Wigley, 1997). The circulation-to-environment approach, as put forward by Yarnal (1993) finds the investigator assessing specific environmental variables relative to synoptic classes. The investigator designs a fairly general synoptic classification to relate to a particular region. The classification typically represents the entire period for which data is available and is independent of the environmental response.

Synoptic classifications can either employ ‘synoptic types’ which classify similar weather properties (e.g. distinct combinations of weather elements) or ‘map-pattern classifications’ which classify the relationships between objects (e.g. pressure patterns). Hewitson and Crane (2002; 2006) employed self-organising maps as a mechanism for climate classification. Yarnal (1993) identified the following synoptic classification methods:

- Manual synoptic types;
- Correlation-based map patterns;
- Eigenvector-based synoptic types;
- Eigenvector-based map patterns;
- Eigenvector-based regionalisations;
- Compositing;
- Circulation indices; and
- Specification.

After selecting a classification scheme it is then necessary to simulate the local surface variables, such as precipitation, from the corresponding (daily) weather patterns (Wilby & Wigley, 1997). This is accomplished by deriving conditional probability distributions for observed data. The precipitation series may be further disaggregated by month or season, or by the dominant precipitation mechanism (Wilby *et al*., 1995, cited in Wilby & Wigley, 1997). The ‘forcing’ weather pattern series are typically generated using Monte Carlo techniques or from the pressure fields of GCMs (Wilby & Wigley, 1997). According to Díez *et al*. (2005), when applied to an ensemble forecast system, the method of analogues can be used in probabilistic mode (considering the joint empirical Probability Density Function (PDF) obtained by combining the analogue sets for each of the ensemble members), or in numeric mode (considering the 75^{th} percentile estimation of the set of analogues for each of the ensemble members).

Wilby *et al*. (2004) and CICS (2007) list the following advantages and disadvantages common to weather pattern-based approaches:

Advantages:

- This technique may provide more realistic scenarios of climate change at individual sites than the direct application of GCM-derived scenarios;
- This technique is much less computationally demanding than dynamical downscaling using numerical models;
- This approach is based on sensible physical linkages between climate on the large scale and weather on the local scale;
- This technique is quite versatile as it can be applied to a wide variety of studies e.g. surface climate, air quality, flooding, etc.; and
- Overlaying (compositing) can be employed for the analysis of extreme events.

Disadvantages:

- This technique requires the additional task of weather classification;
- Large amounts of observational data may be required to establish statistical relationships for the current climate;
- Specialist knowledge may be required to apply the technique correctly;
- The relationships may not be valid under future climate forcing;
- It may not capture intra-type variations (i.e. variations that occur within a specific synoptic type) in surface climate; and
- Different relationships between the weather types and local climate may have occurred at some sites during the observed record.

Regardless of the means of classifying and/or generating new weather pattern series, the circulation-based approach to downscaling remains particularly appealing because it is founded on sensible physical linkages between climate on a large scale and weather on the local scale (Wilby & Wigley, 1997). In their review of downscaling methods, Wilby and Wigley (1997) found that circulation-based approaches perform better than some of the other statistical downscaling methods.

2.3.3 Stochastic weather generators

Stochastic weather generators can be regarded as “statistical characterisations of the local climate, or as elaborate random number generators whose output resembles real weather data” (Wilks, 1999). Their application in climate change studies involves perturbing the stochastic model parameters to reflect a changed climate, and then generating synthetic weather series consistent with this new climate for use with impact models (Wilks, 1999). At the heart of all stochastic weather generators are first- or multiple-order Markov renewal processes in which, for each successive day, the precipitation occurrence (and possibly amount) is governed by outcomes on previous days (Wilby & Wigley, 1997). Although stochastic weather generators are more widely used in temporal downscaling, they may also be used for spatial downscaling which requires a large amount of observed station data that may not be readily available (CICS, 2007). Daly *et al*. (1994, cited in Wilby & Wigley, 1997) present a method of spatially distributing stochastic weather generator parameters across landscapes, even in complex terrain, by combining interpolation techniques with digital elevation models. Semenov and Brooks (1999) describe a method to produce daily rainfall and temperature data for the gaps between observed sites with the aid of spatial interpolation of stochastic weather generator output.

Alternatively, disaggregating of monthly precipitation totals obtained from GCMs can be done by means of a stochastic weather generator. Such a weather generator consists of a model of weather variables as stochastic processes and it must be calibrated with daily meteorological observations. The estimation of precipitation involves first using a Markov procedure to model the occurrence of wet and dry days, where after the amount of precipitation falling on wet days is modelled using a functional estimate of the precipitation frequency distribution. Remaining variables are then computed based on their correlations with each other and with the wet or dry status of each day. After calibrating the weather generator, a parameter file is produced which contains a statistical description of the characteristics of the climate at the site under examination. The stochastic component within a weather generator is controlled by the selection of a random number. By varying this random number completely different weather

sequences can be generated (CICS, 2007). This means that it is possible to generate many sequences of daily weather for a particular scenario. However, the statistical characteristics (e.g. mean and variance) of each sequence should be very similar, if not identical, but the day-to-day values will vary thus representing the natural variability. Weather generators have been used with success in a range of applications in agriculture and environmental management. Wallis and Griffiths (1995) used a weather generator to derive daily values for precipitation, wind speed and wind direction, while Oelschlägel (1995) employed a statistical weather generator to derive daily values for precipitation, temperature and radiation.

Wilby *et al*. (2004) and the Canadian Institute for Climate Studies (2007) lists the following advantages and disadvantages associated with the use of stochastic weather generators:

Advantages:

- The ability to produce large ensembles for uncertainty analysis or time series of unlimited length for extremes;
- The opportunity to obtain representative weather time series in regions of sparse data, by interpolating observed data; and
- The ability to alter the weather generator's parameters in accordance with scenarios of future climate change – changes in variability can be incorporated as well as changes in mean values.

Disadvantages:

- Seldom able to describe all aspects of climate accurately, especially persistent events, rare events and decadal- or century-scale variations;
- Designed for use independently at individual locations and few weather generators can account for the spatial correlation of climate (e.g. changing precipitation parameters may have unanticipated effects on secondary variables like temperature); and

- Assume similar wet and dry day sequencing under future climate forcing (personal addition).

2.3.4 Regression-based methods

These approaches generally involve establishing linear or nonlinear relationships between sub-grid scale parameters and coarser resolution (grid scale) predictor variables (Wilby & Wigley, 1997). These methods are also referred to as “statistical interpretation” or “statistical postprocessing” in the literature (Maini *et al.*, 2004; Marzban *et al.*, 2005). Two of the more popular approaches that improve over climate simulation and numerical weather prediction (NWP) and are used in most operational centres the world over, are model output statistics (MOS) and perfect prognosis (PP) (Maini *et al.*, 2004; Marzban *et al.*, 2005). Both of these methods utilise the idea of relating model forecasts to observations through linear regression (Marzban *et al.*, 2005). More sophisticated techniques, such as ‘expanded downscaling’ (Burger, 1996), can model the mean and short-term variability by linking in a bilinear way the covariance of the global circulation with the covariance between local weather variables. Marzban (2003, cited in Marzban *et al.*, 2005) also allows for non-linear relationship among the variables. Since the internal weights of an artificial neural network (ANN) model imitate nonlinear regression coefficients, is seems reasonable to group ANN approaches under regression methods as well (Hewitson & Crane, 1996).

Having derived a regression equation or trained an ANN to relate the observed local and regional climates, the equations may then be ‘forced’ using regional scale climate data obtained from a GCM operating in either a ‘control’ or ‘perturbed’ state (Wilby & Wigley, 1997). An alternative approach, relating to the empirical method, involves regressing the same parameter from a regional to local scale, or across several scales (e.g. Carbone & Bramante, 1995, cited in Wilby & Wigley, 1997).

The Canadian Institute for Climate Studies (2007) lists the following advantages and disadvantages that apply to regression-based approaches:

Advantages:

- Relatively straightforward to apply and computationally less demanding than dynamical downscaling;
- Provides more realistic scenarios of climate change at individual sites than the straight application of GCM-derived scenarios to an observed climate data set;
- Ensembles of high resolution climate scenarios may be produced relatively easily.

Disadvantages:

- Large amounts of observational data may be required to establish statistical relationships for the current climate;
- Specialist knowledge may be required to apply the technique correctly;
- It may not be possible to derive significant relationships for some variables;
- Provides a poor representation of the observed variance and extreme events;
- The relationships are only valid within the range of the data used for calibration and so should not be extrapolated as future projections for some variables may lie outside of this range; and
- A predictor variable which may not appear as the most significant when developing the transfer functions under the present climate may be critical under future climate conditions.

2.3.4.1 Model output statistics (MOS)

The MOS approach uses quantities from climate simulations or NWP output as predictor variables, whereas the PP approach only uses the climate simulation or NWP forecast predictors when making forecasts. As depicted in Figure 2.1 the MOS approach uses these predictors in both the development and implementation of the statistical equations (Wilks, 1995). This gives MOS the capacity to include the influences of specific characteristics of different GCM or NWP models at different projections into the future directly in the regression equations (Wilks, 1995). The regression equations are developed for a future predictand (e.g. tomorrow's temperature) using GCM or NWP forecasts for values of the predictors at that future time (e.g. tomorrow's forecasted 1000 – 850 hPa thickness). Therefore, to develop

MOS forecast equations it is necessary to have a developmental data set composed of historical records of the predictand, together with archived records of the forecasts produced by the climate simulations or NWP model for the same days on which the predictand was observed (Wilks, 1995). The time lag in MOS forecasts is therefore incorporated through using the GCM or NWP forecast.

Although MOS is known to remove the bias from climate simulations or NWP forecasts, its development generally requires large datasets involving both observations and model variables that are not always readily available (Marzban *et al.*, 2005). Furthermore, GCM or NWP models are not static and regularly undergo changes aimed at improving their performance. The MOS method therefore requires that during the archival period the model configuration should have been kept unchanged. Today's rapidly changing model environment prevents the widespread use of the MOS technique because every time a significant change in the numerical model is made, the MOS equations have to be redeveloped (Maini *et al.*, 2004). According to Marzban *et al.* (2005), MOS is known to maintain reliability but loses sharpness and converges to climatology for longer time-period forecast projections. In order to achieve greater stability, a larger developmental sample is required for both MOS and PP (Maini *et al.*, 2004).

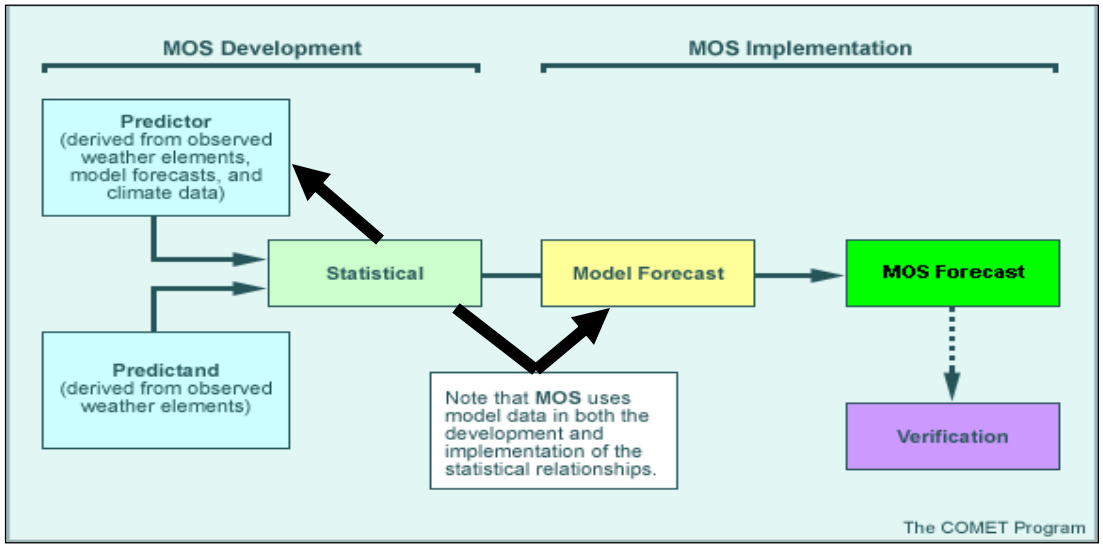

Figure 2.1: Development of a MOS forecast (COMET, 2008)

2.3.4.2 Perfect prognosis (PP)

As the term “perfect prognosis” implies, this technique makes no attempt to correct for possible climate simulation or NWP model errors or biases, but takes their forecasts for future atmospheric variables at face value, thus assuming they are perfect (Wilks, 1995). The assumption is that the model predictor (e.g. model forecast of 700 hPa geopotential height) is equal to the observed predictor (e.g. observed 700 hPa geopotential height) for all times. Here it is sufficient to produce the regression equations from simultaneous values of the observed predictors and observed predictand (Marzban *et al.*, 2005). Thus, only historical climatological data are used in the development of a PP forecasting equation as depicted in Figure 2.2.

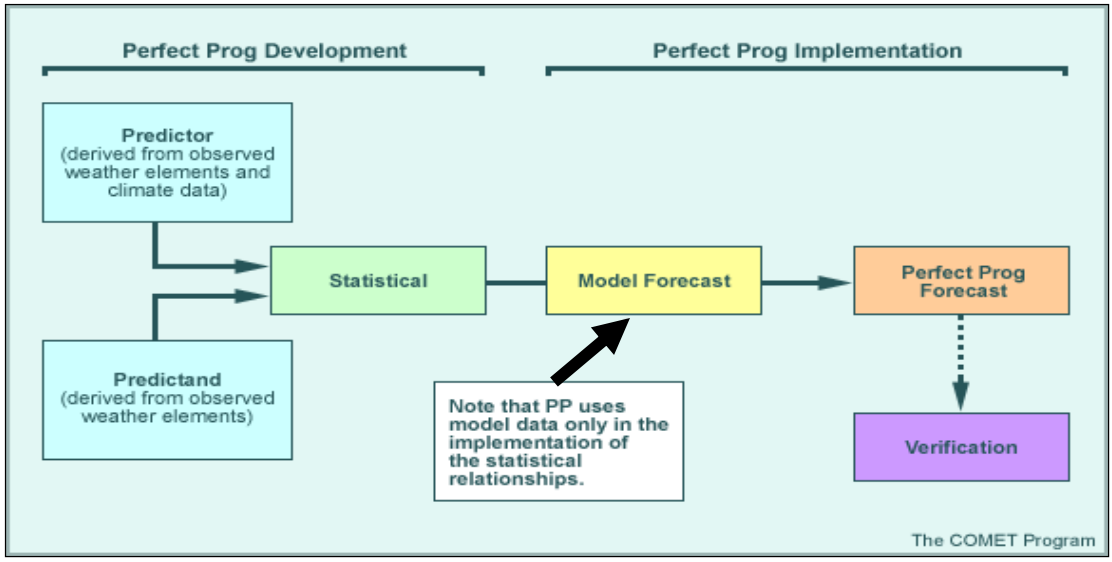

Figure 2.2: Development of a PP forecasting system (COMET, 2008)

PP equations do not incorporate any time lag. Simultaneous values of observed predictors and predictands are used to fit the regression equations i.e. the equations specifying “tomorrow’s” predictand are developed using “tomorrow’s” predictor values (Wilks, 1995). However, in applying the regression equation, it is the GCM forecasts of the predictors that are substituted into the regression equation. Therefore, the forecast time lag in the PP approach is contained entirely within the GCM time steps (Wilks, 1995). This, however, implies that quantities not forecast by the GCM or NWP model cannot readily be included as potential predictors. If the GCM or NWP forecasts for

tomorrow's predictors really are perfect, the PP regression equations should theoretically provide very good forecasts (Wilks, 1995).

However, if the climate simulations or NWP model is flawed, information is lost due to model deficiency. It then follows that a generalisation of PP where the predictor and predictand are taken at different times (e.g. 700 hPa geopotential height at analysis hour against future rainfall) may actually outperform the conventional PP (Marzban *et al.*, 2005). This stems from the fact that even for a deficient model the model analysis should be more accurate than the model forecasts. Marzban *et al.* (2005) noted that PP is less restrictive because its development is not limited by the availability of model data, but concluded that its forecasts are biased and have higher error variance than MOS forecasts. Maini *et al.* (2004) followed a PP approach for the statistical interpretation of NWP products. The resultant medium range precipitation forecasts showed increased skill when compared with that from the direct model output.

It has been well established that MOS provides better forecasts than PP due to its ability to account for some of the systematic errors in GCMs (Maini *et al.*, 2004) but in the case of short-term forecasts over Canada, Brunet *et al.* (1988) have shown that PP outperforms MOS. Although PP forecasts are not bias free, their development is much simpler as it requires only observations for both predictor and predictand (Marzban *et al.*, 2005). PP forecasts also do not deteriorate when significant changes are made to the numerical model and the same equation will remain valid as they were not developed using GCM output (Maini *et al.*, 2004).

2.3.4.3 Reanalysis (RAN)

Kalnay (2003, cited in Marzban *et al*., 2005) proposed the utilization of reanalysis data to develop a postprocessor with the advantages of both MOS and PP, but without the weaknesses due to limited training data. This method, referred to as RAN by Marzban *et al*. (2005) also has the added quality of separating the loss of information between predictor and predictand into its components – one due to the inadequacies of the numerical model, and the other due to chaos in the atmosphere itself. As a first step,

one may develop a regression equation that translates the numerical model predictor to the observed one. This regression model would capture only model deficiencies (Marzban *et al*., 2005). The second step would then involve developing a regression equation that maps the observed predictor to the observed predictand. Since this regression does not involve the model at all, it captures the loss of information due to chaos in the atmosphere (Marzban *et al*., 2005). This two-step approach may be employed in practice to produce a forecast for the predictand, and so, in a way, this method can be considered as a hybrid of MOS and PP, since both the observed and numerical model predictors are engaged in forecasting the predictand (Marzban *et al*., 2005). Since this approach does not allow for the predictor and predictand to be the same physical quantity (as the second step would then involve mapping a variable onto itself), Marzban *et al*. (2005) suggested replacing the observed predictor with a "reanalysis" value. Here the numerical model is used to provide the best estimate of the reanalysis, followed by a regression model to provide the best estimate of the observed predictand. Marzban *et al*. (2005) concluded that MOS may be expected to outperform PP and RAN in terms of bias, error variance, and mean squared error, but that the uncertainty of MOS forecasts may be hindered by the limited size of available model data. This may be due to the fact that the calibration period in MOS is limited by the period of archived GCM forecasts which is sometimes too short to capture the full climate variability. RAN forecasts have lower uncertainty than MOS if its sample size is larger than MOS's sample size (Marzban *et al*., 2005).

2.4 Dynamical Downscaling

2.4.1 Limited-area modelling

The resolution of contemporary GCMs is still not fine enough to resolve small-scale atmospheric circulations, for example those affected by complex topographical features and land cover inhomogeneity (McGregor, 1997). Dynamical downscaling involves the nesting of a higher resolution Limited-Area Model (LAM) within a coarser resolution GCM (Wilby *et al*., 2002). LAMs are similar to GCMs, but operations are performed at a

higher resolution and therefore contain a better representation of, among other things, the underlying topography within the model domain (CICS, 2007). Depending on the model resolution, LAMs may also be able to resolve some of the atmospheric processes which are parameterised in a GCM (CICS, 2007). A high resolution model thus simulates the climate features and physical processes in much greater detail for a limited area of the globe.

The general approach is to embed a higher-resolution LAM within the 'driving' GCM, using the GCM to define the initial and (time-varying) boundary conditions (Wilby & Wigley, 1997). This procedure is commonly referred to as 'nesting'. Most nesting techniques are one-way i.e. there is no feedback from the LAM simulation to the driving GCM. The global model simulates the response of the global circulation to large scale forcing, whilst the LAM accounts for sub-GCM grid scale forcing, such as complex parameterisations, orography or details of the land surface, in a physically-based way and thus enhances the simulations of atmospheric and climatic variables at finer spatial scales (CICS, 2007). LAMs may be computationally demanding, depending on the domain size and resolution, and are as expensive to run as a global GCM (Wilby & Wigley, 1997; CICS, 2007). This has limited the length of many experiments. They are also somewhat inflexible in the sense that the computational demands apply each time that the model domain is shifted to another region. Moreover, the LAM is completely dependent on the accuracy of the GCM grid-point data that are used to force the boundary conditions of the region – a problem that also applies to circulation-driven downscaling methods (Wilby & Wigley, 1997). Any errors in the GCM fields may be aggravated in the LAM thus resulting in poor simulation of the regional climate (CICS, 2007).

Kanamaru and Kanamitsu (2007) used a Regional Climate Model (RCM) to successfully perform a dynamical downscaling of the NCEP–NCAR reanalysis over the Northern Hemisphere. They claim their success was due to the use of the scale-selective bias-correction scheme, which maintains the large-scale analysis of the driving global reanalysis in the centre of the domain where lateral boundary forcing has little control

(Kanamaru & Kanamitsu, 2007). With the aim of producing higher-resolution global reanalysis datasets from coarse-resolution reanalysis, Yoshimura and Kanamitsu (2008) developed a global version of the dynamical downscaling using a global spectral model. In their study a variant of spectral nudging, the modified form of scale-selective bias correction, was adopted for regional models. Spectral nudging implies that the forcing technique is stipulated not only at the lateral boundaries but also in the model interior (Von Storch *et al*., 2000).

2.4.2 Stretched-grid modelling

An alternative method of dynamical downscaling is presented in the form of "variable resolution modelling" as employed in the Conformal-Cubic Atmospheric Model (C-CAM). This GCM has the capacity to run in a variable resolution stretched-grid mode to function as a RCM (Engelbrecht *et al*., 2009). It thus provides high resolution over the area of interest i.e. shrinking the grid intervals over the area of interest, whilst gradually decreasing the resolution as one moves away from the area of interest (Engelbrecht *et al*., 2009).

Variable resolution modelling provides great flexibility for dynamic downscaling from any GCM as compared to the more customary nested limited-area modelling approach (Engelbrecht *et al*., 2009). It basically requires only sea-surface temperatures and far-field winds from the host model (McGregor and Dix, 2001). Variable resolution modelling also circumvents other problems that may arise with limited-area models, such as reflections at lateral boundaries (McGregor and Dix, 2001).

Since different downscaling methods have different strengths and weaknesses, this has prompted some commentators to advocate closer integration of statistical (i.e. stochastic and empirical) and dynamical downscaling methods (Hostetler, 1994; Bass, 1996, cited in Wilby & Wigley, 1997). Wilby & Wigley (1997) recommend that rigorous testing and comparison of statistical downscaling approaches with RCMs be undertaken and claim that much can be learnt from applying a number of different approaches in combination and from evaluations of the relative merits of regression, weather pattern,

stochastic and dynamic models. In time a framework may then be set up to assist climate change impact researchers to select a combination of downscaling techniques that should provide the best results for their particular application.

2.5 Choice of Downscaling Method

For this study, the decision was made to explore the suitability of the statistical downscaling model (SDSM) developed by Wilby *et al*. (2002) to downscale GCM projections of future climate. The software was downloaded from the SDSM website (https://co-public.lboro.ac.uk/cocwd/SDSM). Within the classification of downscaling techniques, SDSM can be viewed as a hybrid of the stochastic weather generator and regression-based methods (Wilby & Dawson, 2007). This is because large–scale predictor variables are used to condition local–scale weather generator parameters such as precipitation occurrence and intensity. In addition, stochastic techniques are used to synthetically increase the variance of the downscaled daily time series to better agreement with observations (Wilby & Dawson, 2007).

The SDSM software reduces the task of statistically downscaling daily rainfall to the following discrete steps (Wilby *et al*., 2002):

a) quality control and data transformation;
b) screening of predictor variables;
c) model calibration;
d) weather generation (using observed predictors);
e) generation of climate change scenarios (using climate model predictors);
f) statistical analysis.

More detail on every step is provided in Chapter 5. Figure 2.3 provides a diagrammatical depiction of the SDSM scenario generation process.

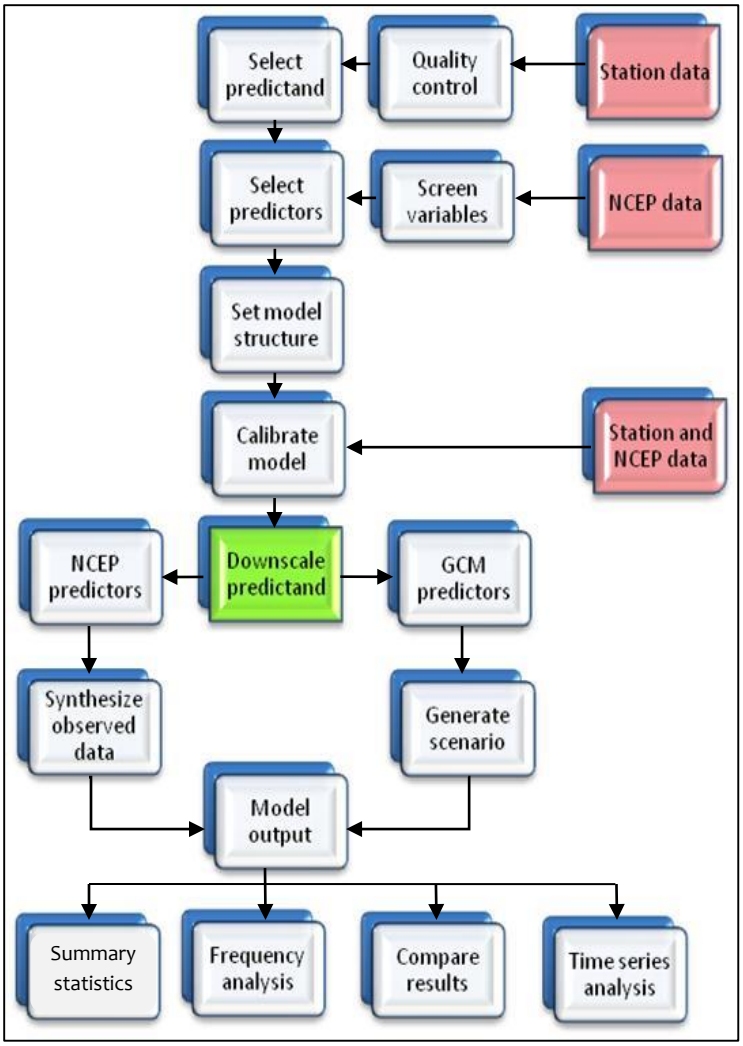

Figure 2.3: SDSM climate scenario generation (after Wilby & Dawson, 2007)

The downscaling will be performed within a PP milieu i.e. only observed large-scale predictors and observed site-specific predictands will be used in the training of the transfer equations. This also implies that the same downscaling model can be used with different model experiments (scenarios).

To date, SDSM has been applied to several meteorological, hydrological and environmental assessments (e.g. Lines *et al*., 2005; Wilby *et al*., 2006, Hessami *et al*., 2008). In particular, Lines *et al*. (2005) used SDSM to downscale the expected climate change impacts with respect to daily mean, maximum and minimum temperature as well as precipitation for 14 sites across Atlantic Canada. According to Wilby and

Dawson (2007) SDSM has also been applied to a range of geographical contexts including Africa, Europe, North America and Asia. Work done on several statistical downscaling models by Goldstein *et al*. (2004) revealed that SDSM produced optimal results for producing station-scale daily meteorological series of temperature and precipitation over Canada.

CHAPTER 3 STUDY AREA AND CLIMATOLOGY

3.1 Physical and Geographical Description

The location of the Olifants River as one of the primary catchments in the north-eastern part of South Africa is shown in Figure 3.1.

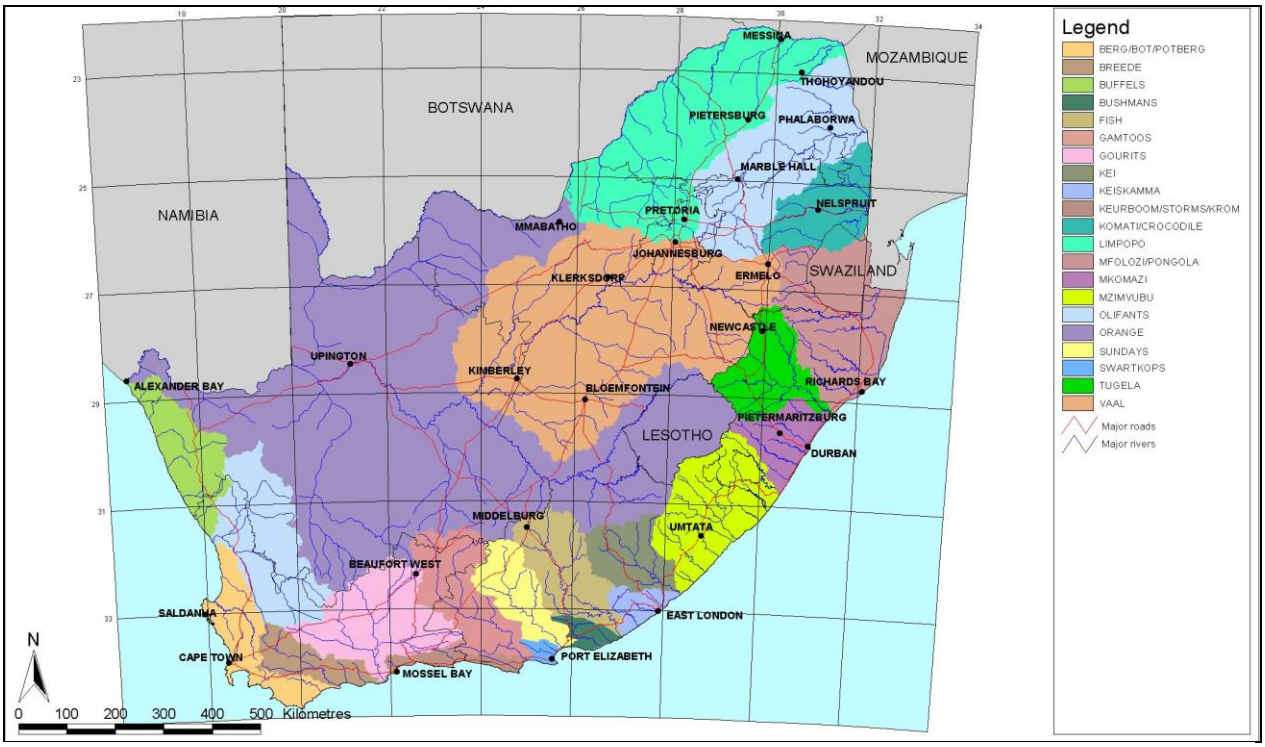

Figure 3.1: Primary catchments of South Africa (DEAT, 2000)

The Olifants River Catchment covers about 54 570 km^2 and is subdivided into 9 secondary catchments (Institute for Water Quality Studies, 2001) and has a total mean annual runoff of approximately 2400 million cubic metres per year. The Olifants River and some of its tributaries, notably the Klein Olifants River, Elands River, Wilge River and Bronkhorstspruit, rise in the Highveld grasslands. The Olifants River flows north through Loskop Dam, meanders past the foot of the Strydpoort Mountains and is forced east by the Transvaal Drakensberg, descending over the escarpment. The Steelpoort and Blyde tributaries, among others, join the Olifants River before it enters the Kruger National Park and neighbouring private game reserves. It then flows east to join with the

Letaba River, crosses into Mozambique where it is named the Rio dos Elefantes which eventually joins the Limpopo River before entering the Indian Ocean at Xai-Xai north of Maputo.

This study focuses on the Upper Olifants River Catchment, which is primarily situated in the Highveld spanning the eastern part of Gauteng and western Mpumalanga. For the purposes of this study the Upper Olifants River is defined as that part of the catchment that is located on the Highveld, upstream (south) of the confluence with the Eland's River near Marble Hall. Figure 3.2 indicates all the quaternary catchments (QCs) within the Upper Olifants River Catchment, with some large towns included for orientation purposes - the Olifants River Catchment is marked by the light yellow colouring. QCs constitute the most detailed level of operational catchment in the Department of Water Affairs and Forestry (DWAF) used for general planning purposes (Midgley *et al*., 1994).

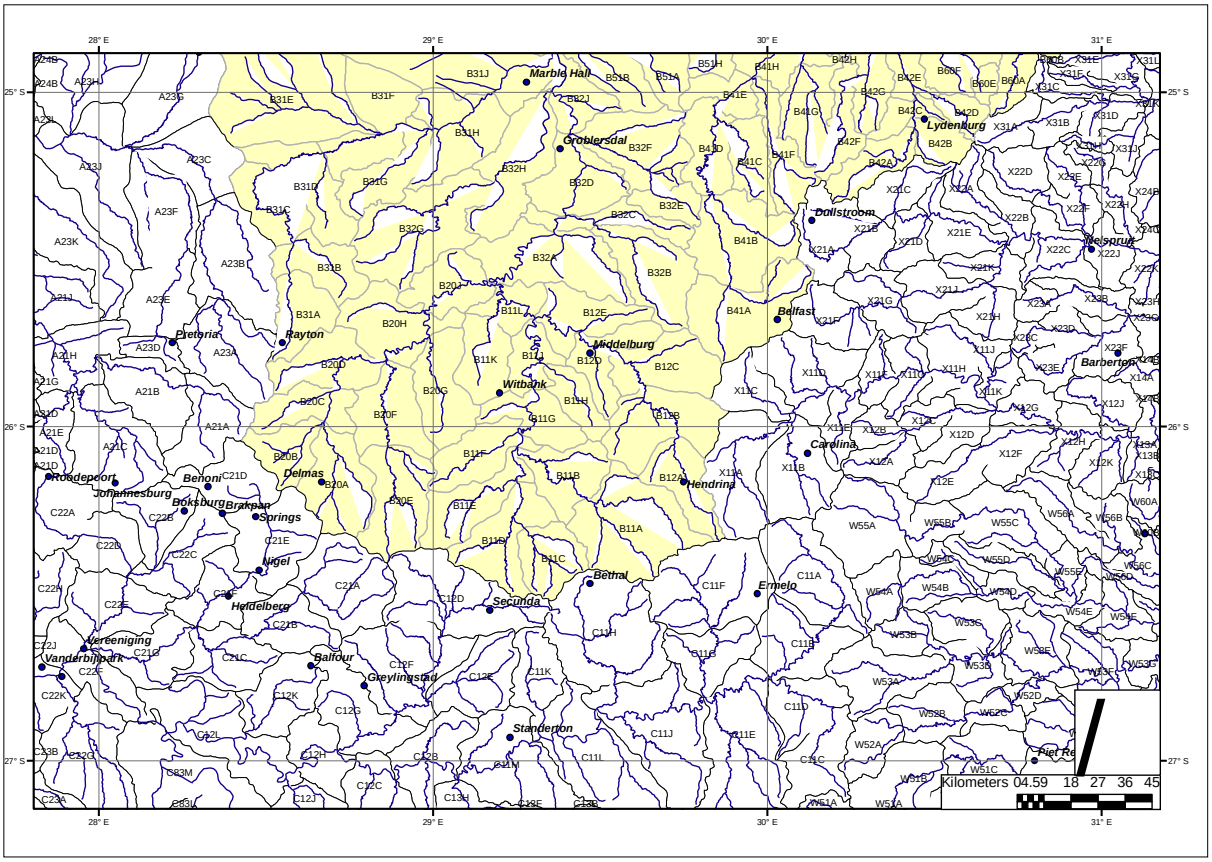

Figure 3.2: Quaternary catchments of the upper-Olifants River (Schulze, 2006)

From Figure 3.2 it can be seen that this area stretches from Rayton and Delmas to the west, Belfast and Dullstroom to the east, Bethal and Secunda to the south and Marble Hall to the north. When it comes to selecting GCM grid boxes during a later phase of the investigation, it is important to note that this covers the area between 24º 50' and 26º 30' S, spanning 28º 30' E to 30º 05' E.

The Highveld is part of the interior plateau of the southern African subcontinent and ranges in altitude from 900 m to 1900 m above sea level. Figure 3.3 provides the terrain morphology as developed by Kruger (1983) for the Highveld region. From this figure it can be seen that the region is dominated by undulating plains with high mountains in the extreme east and a section of lowland and hills in the extreme north. An area of pans and plains are also situated in the extreme southwest.

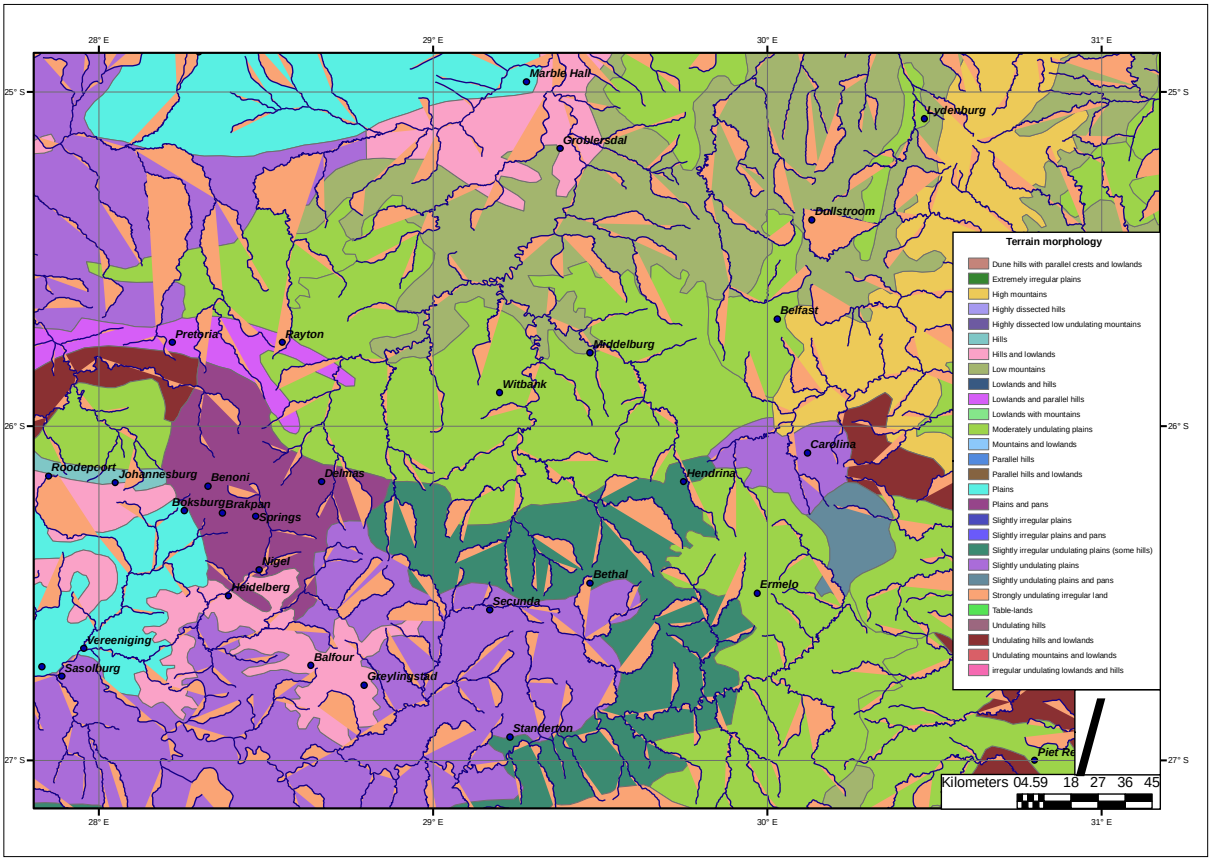

Figure 3.3: Terrain morphology of the Highveld (Schulze, 2006, after Kruger, 1983)

Activities in the upper reaches of the Olifants River Catchment are characterised primarily by mining (mainly coal), electricity generation, manufacturing (mainly steel in the vicinity of Middelburg), agriculture (mainly commercial dryland but irrigated in the north) and conservation. The potential for arable agriculture is high in the south and low to marginal in the north of the study area, while the long-term grazing capacity also drops from about 40 large stock units (LSU) in the higher rainfall regions in the south and east to about 15 LSU in the drier northern parts (DEAT, 2007).

3.2 Quaternary Catchment Selection

Figure 3.2 indicates all the quaternary catchments (QCs) within the Upper Olifants River Catchment. For the purpose of this study five quaternary catchments were selected for which statistical downscaling will be performed. The selection was carried out in such a way as to obtain a reasonable spatial separation between the sites and to include a range of terrain morphology types (Figure 3.3), a range of altitudes, mean annual precipitation totals (Figure 3.4) as well as well-known towns. Table 3.1 summarises some of the physical and climatological characteristics of the selected quaternary catchments. For the purpose of geographical identification the five QCs will be named after the towns located in them, viz. Witbank, Middelburg, Delmas, Groblersdal and Belfast.

Table 3.1: Summary of selected quaternary catchments (after Schulze, 2006)

QC name	Coordinates	Town in QC	Altitude	Annual Rainfall	Terrain Morphology
B11K	25º 50' S 29º 15' E	Witbank	1490 – 1600 m	~ 640 mm	Moderately undulating plains
B12D	25º 45' S 29º 30' E	Middelburg	1400 – 1580 m	~ 650 mm	Low mountains & Moderately undulating plains
B20A	26º 10' S 28º 40' E	Delmas	1600 – 1640 m	~ 660 mm	Plains and pans
B32D	25º 10' S 29º 20' E	Groblersdal	800 – 1000 m	~ 610 mm	Hills and lowlands
B41A	25º 40' S 30º 00' E	Belfast	1800 – 2000 m	~ 880 mm	High mountains & Moderately undulating plains

3.3 Climatological Description

3.3.1 General climatological description

According to Köppen's climate classification (Schulze, 2006; Ahrens, 2003) the lower lying northern extremities are classed as Cwa (humid mesothermal with dry winters and long, hot summers) while the remainder of the area is classed as Cwb (humid mesothermal with dry winters and long, cool summers). Here "long" implies that at least 4 months of the year has an average temperature of more than 10 ºC, while "hot" implies that the average temperature of the warmest month exceeds 22 °C (Ahrens, 2003; Kruger, 2004).

The average annual precipitation in this Highveld region varies from about 900 mm on its eastern border to approximately 650 mm in the west and 550 mm in the lower lying northern extremities (Figure 3.4). The rainfall is almost exclusively due to showers and thunderstorms and occurs mainly in summer, from October to March, with the maximum falls occurring in December and January (Kruger, 2004). There is a gradual shift from early summer rain (December maximum) in the east to mid-summer rain (January maximum) in the west. A mid-summer dry spell of between 13 to 21 days occurs on the Highveld in about 9 out of 10 years (Walker & Schulze, 2008). The winter months are normally dry and about 85% of the annual rainfall falls in the summer months; heavy falls of 125 to 150 mm occasionally fall in a single day (Schulze, 1994). An average of about 75 thunderstorms occur over this area per year. These storms are often violent with severe lightning and strong (but short-lived) gusty south-westerly winds and are sometimes accompanied by hail. This region has the highest hail frequency in South Africa; in the order of 4 to 7 occurrences (depending mainly on altitude) expected annually at any one spot, whilst occasionally hailstones may grow to the size of hen's eggs or tennis balls and can cause tremendous damage (Schulze, 1994). Snow may occur about once or twice a year, but is still considered exceptional this far north.

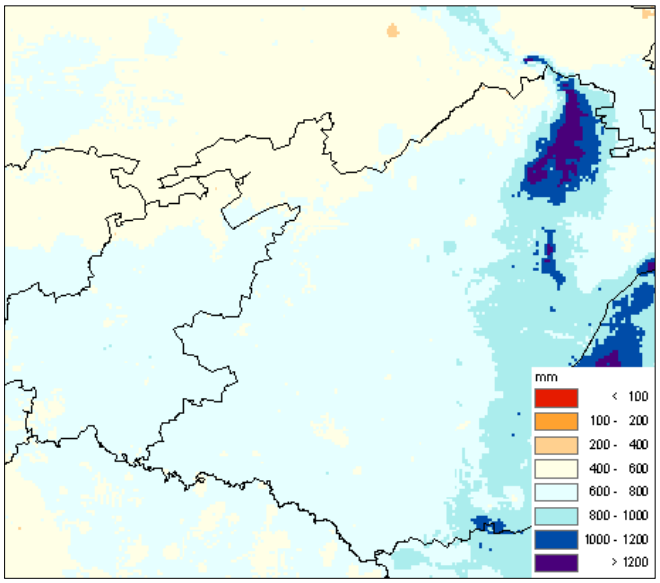

Figure 3.4: Mean annual precipitation in mm (Schulze, 2006, after Lynch 2004)

Average daily maximum temperature is roughly 27 ºC in January and 17 ºC in July but in extreme cases these may rise to 38 ºC and 26 ºC respectively (Schulze, 1994). Average daily minima range from about 13 ºC in January to 0 ºC in July, whereas extremes can drop to 1 ºC and -13 ºC respectively. The period during which frost is likely to form lasts on average for about 120 days from May to September (Schulze, 1994). In Middelburg 166 days exhibit a 10% probability of recording subzero temperatures, while 94 days adhere to an 80% probability (Kotze, 1980). In Belfast these figures are 230 days (10% probability) and 131 days (80% probability) respectively (Kotze, 1980), mainly due to the higher altitude.

Winds are highly variable but easterly and westerly winds are more prevalent (Kruger, 2004). On the whole, winds are light except for short periods during storms. Very occasionally tornadoes do occur and cause tremendous damage if they happen to strike a populated area.

Sunshine duration in summer is about 60% and in winter about 80% of the possible total (Schulze, 1994). Solar radiation is higher in the western than the eastern parts of the Highveld. The mid-summer solar radiation ranges from 21 MJ m^{-2} day^{-1} in the east to

about 23 MJ m^{-2} day^{-1} in the west and reaching a maximum of 25 MJ m^{-2} day^{-1} in the north (Schulze, 2006). Mid-winter solar radiation is considerably lower, ranging from 14 to 16 MJ m^{-2} day^{-1} (Schulze, 2006).

3.3.2 Typical near-surface synoptic scale weather patterns over southern Africa

Although pressure, wind and rainfall patterns are constantly changing, there are certain basic patterns in the pressure and wind fields that occur regularly. Seasonal variations exist in the location of the southern Atlantic Ocean High (AOH) and southern Indian Ocean High (IOH) pressure cells. In the Indian Ocean the pressure cell is subjected to a half-annual variation in its latitudinal (N-S) movement, but an annual variation in its longitudinal (E-W) movement. In the Atlantic Ocean both the latitudinal and longitudinal variations are half-annually (Taljaard, 1996). On average the AOH is located 3º further north than the IOH, but both cells shift 5 - 6º northwards in the winter. The annual longitudinal shift of the AOH is 7 - 13º, while the IOH is subject to a considerably larger longitudinal shift of 24 - 30º (Taljaard, 1996). The latter, therefore, has a much larger effect on the weather and climate of South Africa.

3.3.2.1 The basic summer weather patterns

During the summer months the following weather conditions prevail as depicted in Figure 3.5 (Taljaard, 1996):

i) The Indian and Atlantic high pressure systems move further southwards, causing westerly winds to occur well to the south of the country.
ii) The Indian Ocean High pressure system (IOH) is centred further out to sea. Wind blowing from this high pressure cell then has a longer sea track over the warm Indian Ocean – where lots of moisture is accumulated – before it moves in over the eastern parts of the subcontinent. This often results in cloudy conditions over these eastern areas with drizzle along the eastern escarpment and adjacent Lowveld.
iii) Moisture laden south-easterly trade winds invade the eastern (and especially north-eastern) parts of the subcontinent. These winds sometimes recurve

southwards, influencing the northern provinces; on occasion they move further northwards and influence Zimbabwe and Zambia.

iv) When moist air is in circulation (as imported by the south-easterly trades), uplift thereof will result in condensation, cloud formation and precipitation.

v) The Atlantic Ocean High pressure system (AOH) is a source of subsiding air, having its centre fairly near the West Coast of South Africa. The winds blowing from it have a short sea track over a cold ocean and thus carry little moisture.

vi) When the air from the AOH meets that from the IOH, a moisture boundary (also called a moisture front or dry-line) forms. Uplift occurs along this moisture front due to the undercutting effect of the colder/dryer air (from AOH), often affecting the rainfall distribution over the entire region.

vii) Sometimes the AOH lies further south and then the south-westerly winds have a longer sea track and contain more moisture. General rain may then occur over the south-eastern parts.

viii) Due to strong surface heating, a heat low normally develops over the north-western interior with convergence in and to the east of it.

ix) South-easterly trades blow to the north of the AOH. These winds accumulate moisture and recurve clockwise around the tropical low – which develops over northern Angola or the Congo – and invades Angola and the Congo from the southwest. These winds are then known as the south-westerly monsoon winds. A convergence zone, known as the “Congo Air Boundary” (CAB), develops where these winds meet the south-easterly trades from the Indian Ocean.

x) The north-easterly trades blow across the equator (monsoon) and where they meet the south-easterly trades a convergence zone is formed, which is known as the Inter-Tropical Convergence Zone (ITCZ).

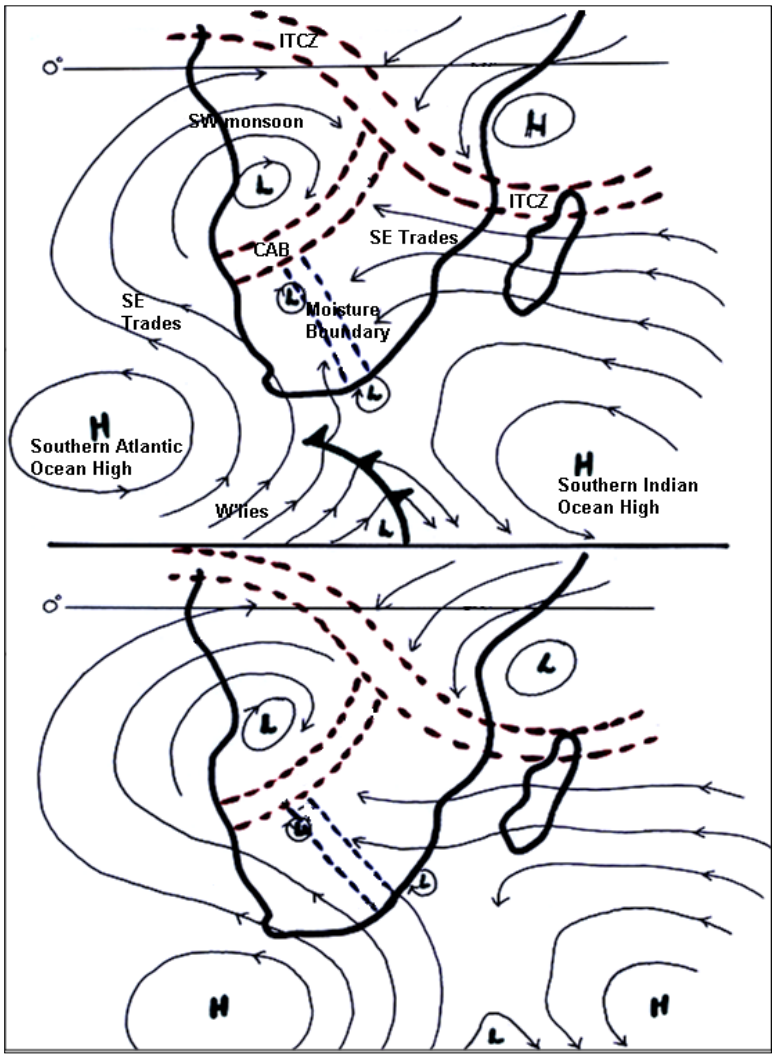

Figure 3.5: Basic summer weather patterns over Southern Africa (Adapted from Taljaard, 1996) - Bottom figure depicts the situation when the AOH ridges to the south of the subcontinent

3.3.2.2 The basic winter weather patterns

During the winter months the following weather conditions prevail as depicted in Figure 3.6 (Taljaard, 1996):

i) The IOH and AOH move further northwards, bringing westerly winds to the southern and south-western coastal regions of South Africa.

ii) Generally cool to cold conditions prevail and in the absence of heat lows over the interior, the AOH and IOH are linked across the land.

iii) A separate high-pressure cell usually forms over the north-eastern interior. The associated subsiding air results in clear skies and calm conditions over large parts of the interior.

iv) With the northwards movement of the AOH and IOH pressure systems, mid-latitude cyclones, linked with cold fronts develop over the Atlantic Ocean, invade the southern regions of South Africa. On occasions these cold fronts have influenced regions as far north as Zambia when a strong high-pressure system ridges closely behind the cold front.

v) Cloudy conditions associated with rain influence the southern, south-western and eastern coasts. Uplift is mainly due to cyclonic or frontal action, but orographic uplift does occur over the escarpment where snowfalls may occur. Interior snowfalls usually occur with the presence of a strong high, following closely behind a cold front, thus pushing the cold air into the interior. Strong upper air flow aids with the uplift, and thus enables condensation and cloud formation to take place.

vi) The north-easterly monsoon disappears completely and the ITCZ shifts far north of the equator (between 5ºN and 20ºN). The south-easterly trades still blow to the north of the high-pressure belt and move across the equator and recurve north-eastwards, becoming part of the great summer Monsoon of India.

vii) The south-westerly monsoon also moves further north, crossing the coast well north of Angola. This results in the “Congo Air Boundary” moving much further north.

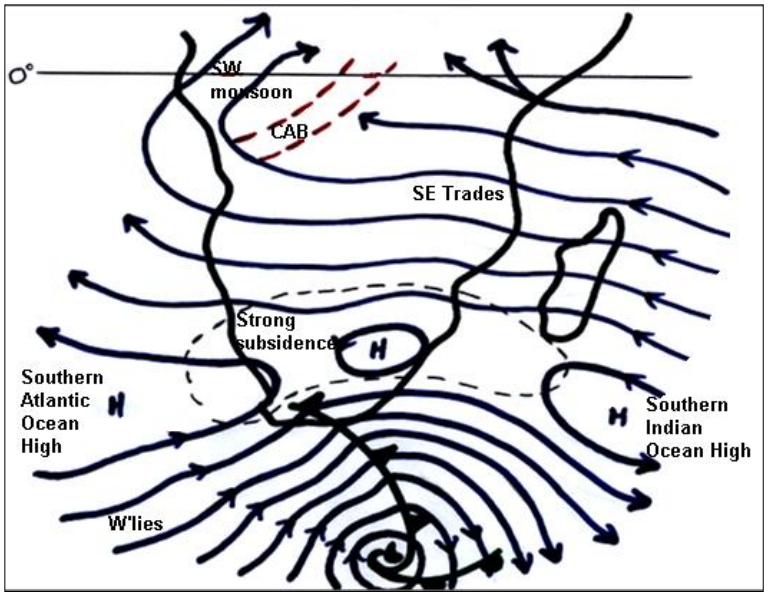

Figure 3.6: Basic winter weather patterns over Southern Africa (Adapted from Taljaard, 1996)

3.3.3 Description of common climatic elements

3.3.3.1 Geopotential heights

Analyses of the geopotential heights at 1000 hPa can be taken as a carbon copy of the mean sea-level pressure analyses. Southern Africa falls within the subtropical high pressure belt (Figure 3.13) and on most days synoptic charts will reveal the position of the AOH with its centre close to the west coast and the IOH to the east of the subcontinent. During the summer months a low pressure trough develops over the western interior of the subcontinent while the IOH extends a ridge over the north-eastern interior (Figure 3.7 top). In the winter a separate high pressure cell develops over the north-eastern interior (Figure 3.7 bottom) so that the subtropical anticyclones actually link up to form a continuous belt of high pressure over southern Africa. Such regions of high-pressure are characterized by subsidence throughout the larger portion of the troposphere, divergence at the surface and mainly stable and dry conditions.

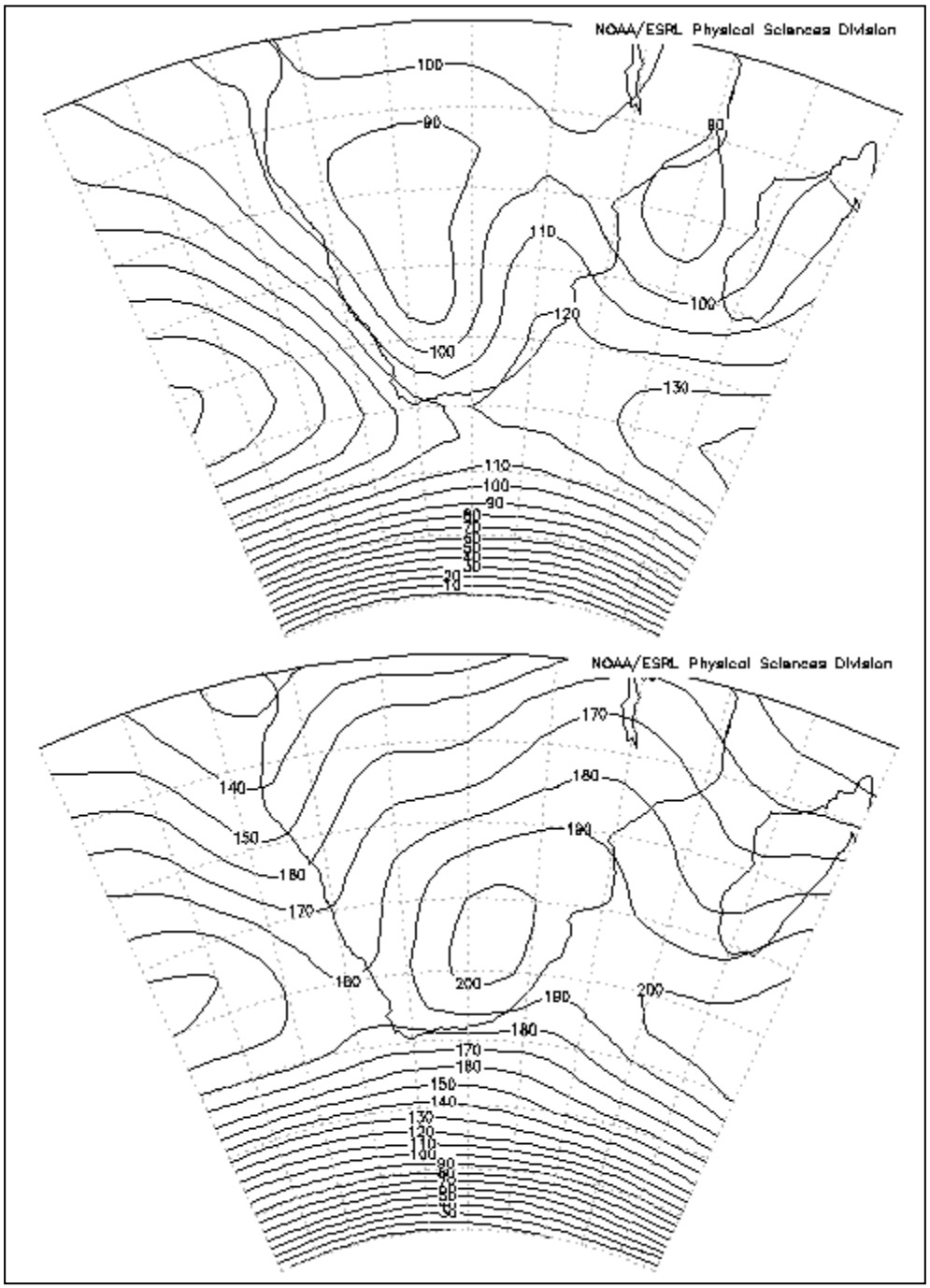

Figure 3.7: Mean 1000 hPa geopotential heights (in gpm) for January (top) and July (bottom) calculated for the 1968-1996 period (NOAA, 2008)

The 850 hPa geopotential heights indicate similar conditions, with a north and westward displacement of the 1000 hPa high pressure cells (Figure 3.8). The invasion of the

south-western parts by westerly troughs during the winter (Figure 3.8 bottom) is also evident. The north-westward displacement of the high-pressure cell with increasing height is clearly evident at 500 hPa, where it is centred over the north-western parts during summer (Figure 3.9 top).

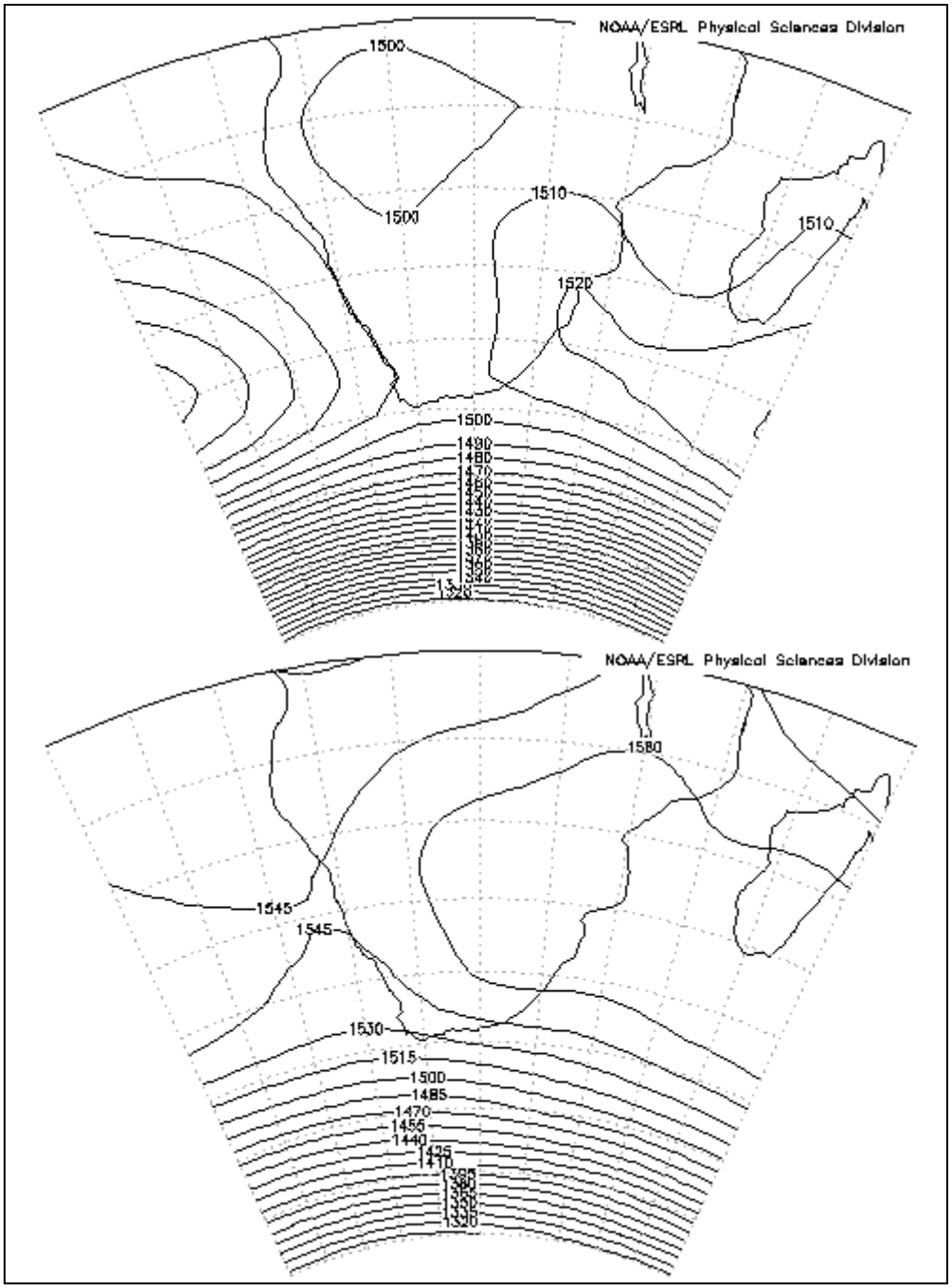

Figure 3.8: Mean 850 hPa geopotential heights (in gpm) for January (top) and July (bottom) calculated for the 1968-1996 period (NOAA, 2008)

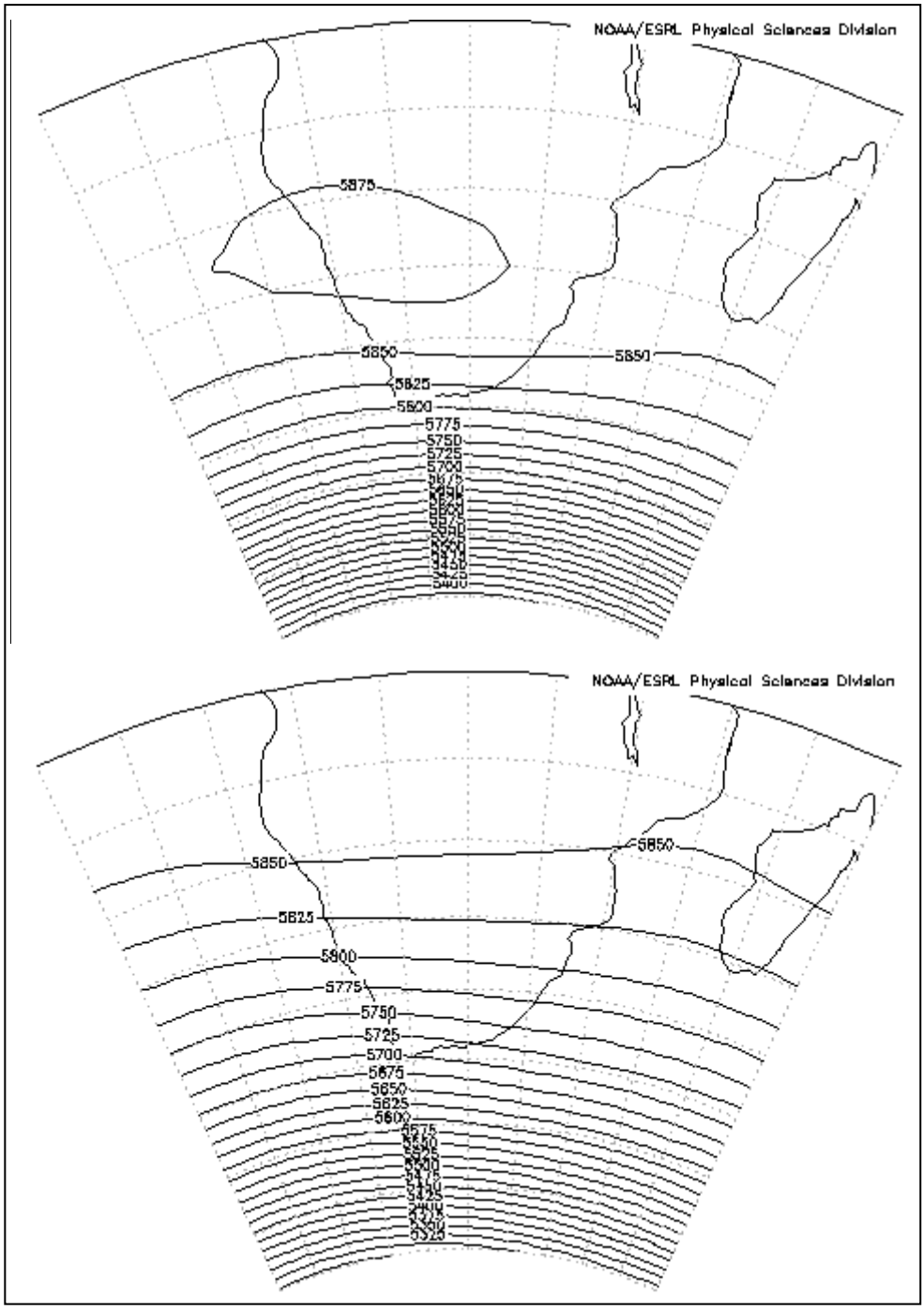

Figure 3.9: Mean 500 hPa geopotential heights (in gpm) for January (top) and July (bottom) calculated for the 1968-1996 period (NOAA, 2008)

3.3.3.2 Zonal wind components

Wind is a vector quantity having both direction and speed. For the purpose of this discussion the wind will be split into a zonal (west to east) and meridional (south to north) component. The distribution of the predominant zonal components of the wind for 850 hPa and 500 hPa are representative of the circulation in the lower and upper troposphere. Generally speaking, the northern parts of southern Africa experience surface easterly component winds (the trade wind regime) while the southern parts are exposed to the westerlies of the mid-latitudes. In summer (Figure 3.10 top), when the subtropical anticyclones have shifted further south, a zone of easterly component winds dominate the north-eastern parts of the subcontinent. The observed easterly winds over the south-western Cape also stem from this southward shift of the subtropical high pressure systems. Westerlies are located over the remainder of the subcontinent with a maximum over the western interior. During winter (Figure 3.10 bottom) the westerlies intensify and shift further north in accordance with the northward shift of the subtropical anticyclones.

At 500 hPa the upper-air high pressure system is situated over the north-western part of the subcontinent in summer (Figure 3.9 top) with westerlies predominating over the areas to the south of it (Figure 3.11 top). Weaker easterly component winds are experienced over the north-eastern parts of the subcontinent and to the north of the upper-air high. During wintertime (Figure 3.11 bottom) the westerlies intensify (due to increasing pressure gradients as indicated by Figure 3.9 bottom) and invade the whole subcontinent so that easterly components are only to be located in the extreme north.

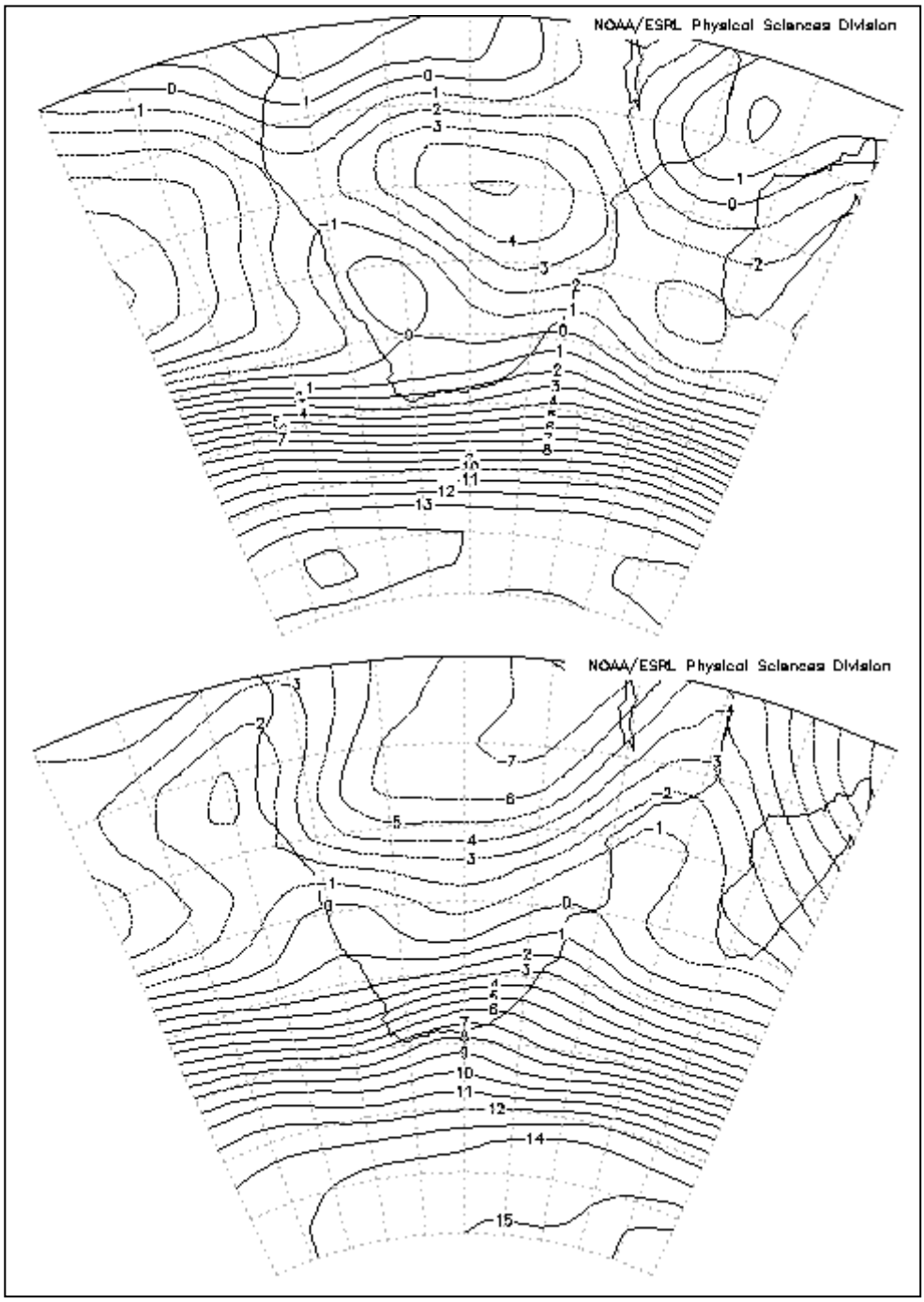

Figure 3.10: Mean 850 hPa zonal flow (in ms^{-1}) for January (top) and July (bottom) calculated for the 1968-1996 period. Positive values indicate westerly winds (NOAA, 2008)

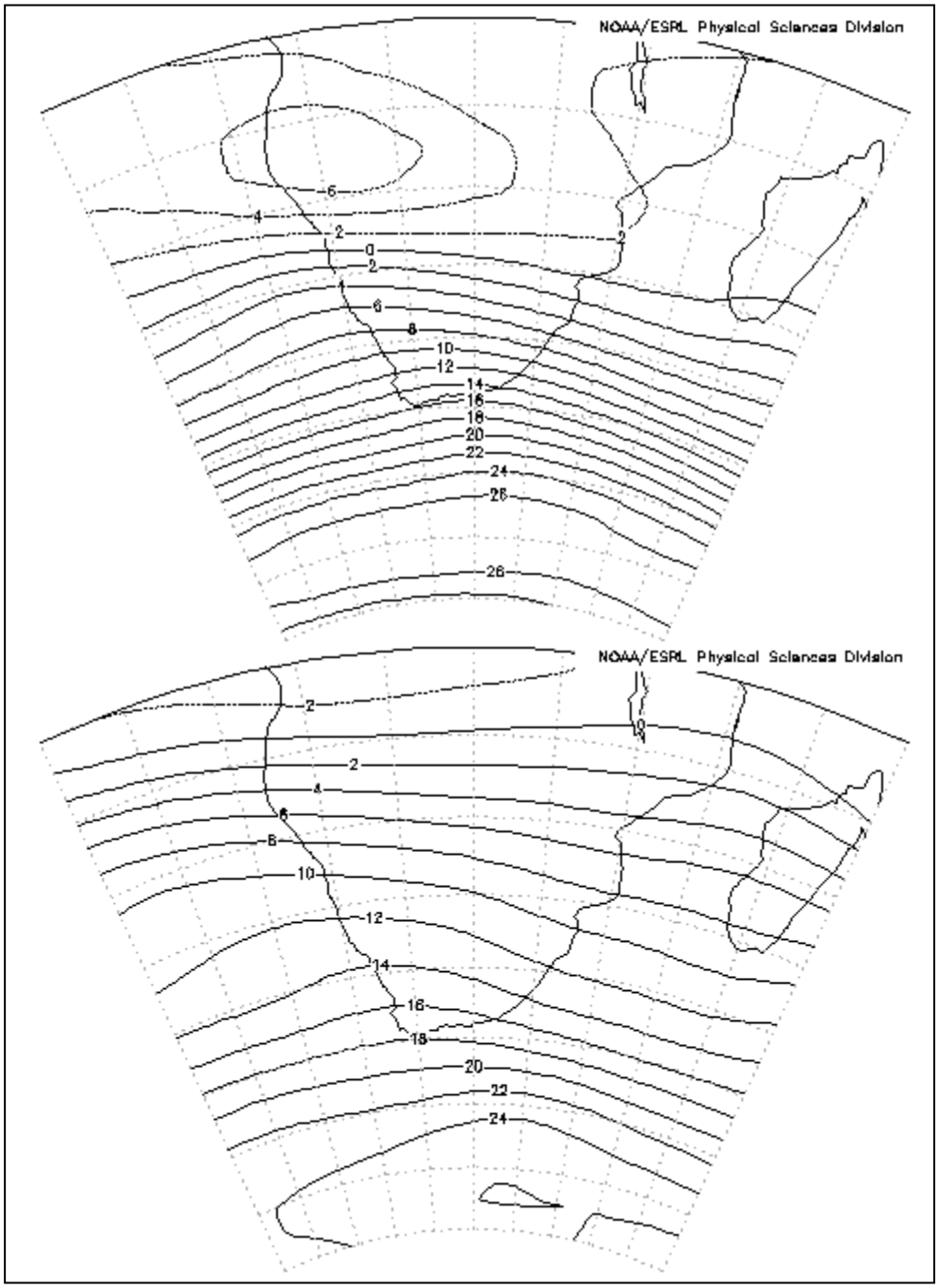

Figure 3.11: Mean 500 hPa zonal flow (in ms^{-1}) for January (top) and July (bottom) calculated for the 1968-1996 period. Positive values indicate westerly winds (NOAA, 2008)

Figure 3.12 is a Hovmoller diagram of the monthly long-term mean zonal wind components at a point over the Upper Olifants catchment, viz. 25 ºS 30 ºE. From this diagram it can be seen that this area is dominated by easterly component winds at the lower levels, with westerlies in the upper-air. The westerlies intensify during the winter

and penetrate further down to reach the surface (situated at approximately 850 hPa in this region).

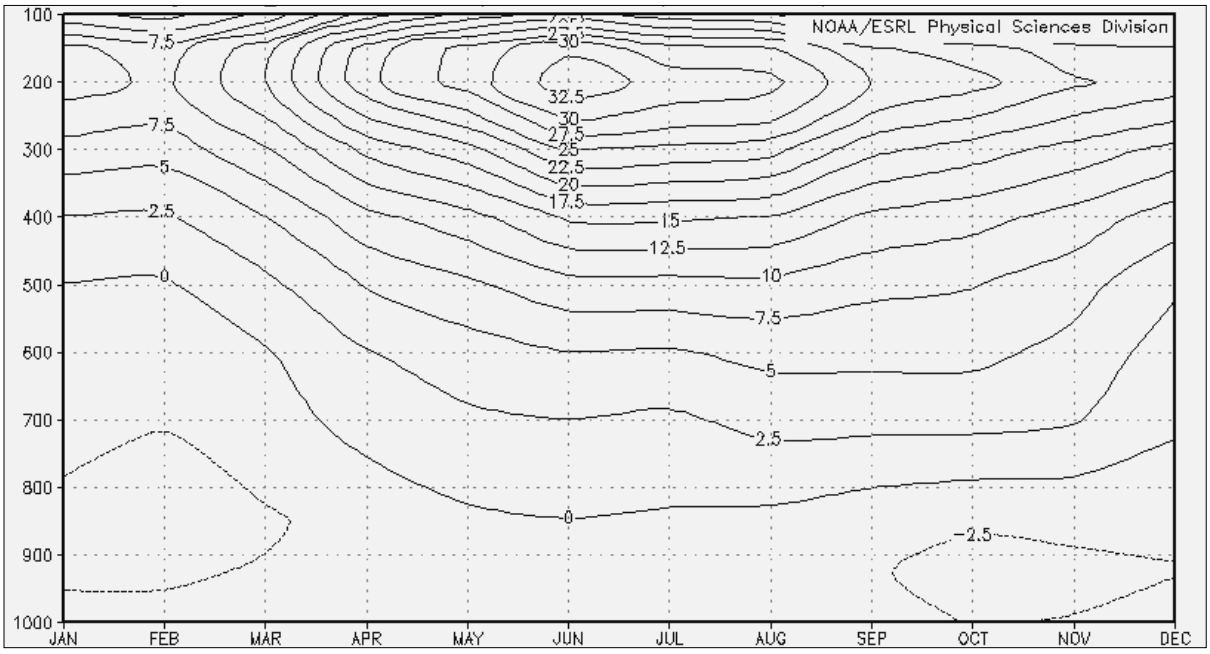

Figure 3.12: Time section of monthly mean zonal flow (in ms^{-1}) at 25ºS 30ºE calculated for the 1968-1996 period. Positive values indicate westerly winds (NOAA, 2008)

3.3.3.3 Meridional wind components

The maintenance of the atmospheric general circulation requires meridional exchanges of energy and angular momentum between the poles and the tropics (Taljaard, 1994). For meridional flow, the general case is sketched by the Idealised Global Circulation (Figure 3.13). According to this 3-cell model the Hadley cell is located in the zone between the equator and roughly 30º S. Here the surface flow is equatorward while the flow aloft is poleward. As the upper flow in the cell moves poleward, it begins to subside in a zone between 20º and 35º latitude. Near the centre of this zone of descending air the winds are generally weak and variable (*horse latitudes*). From the centre of the horse latitudes the surface flow spills into a poleward branch and an equatorward branch.

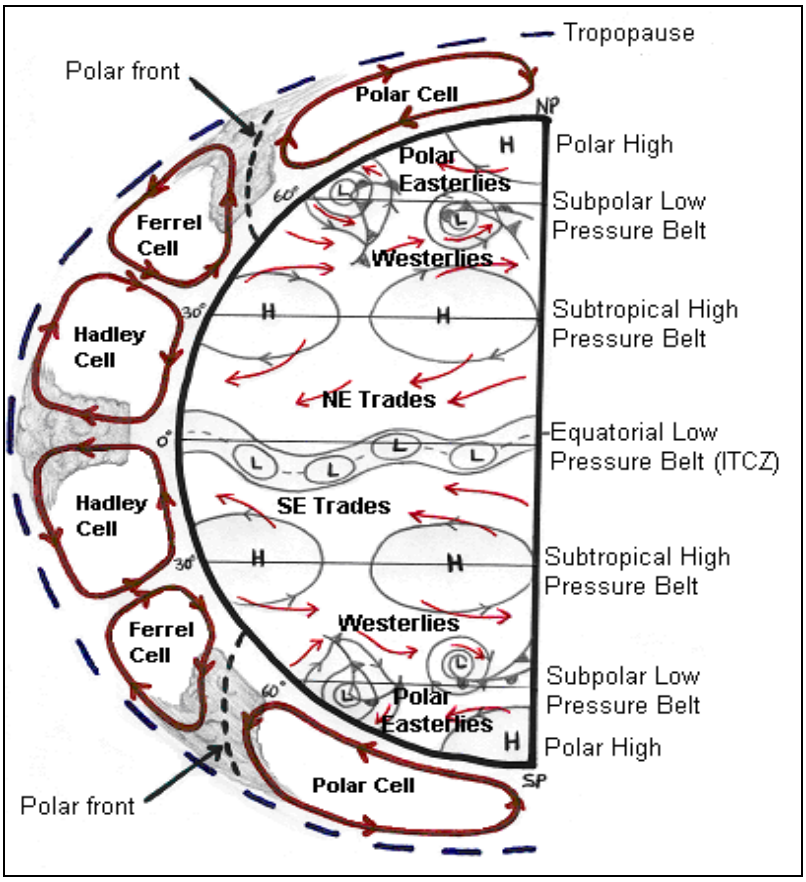

Figure 3.13: Idealised Global Circulation (Three-Cell Model)

In summer (Figure 3.14 top), when a surface trough frequents the central interior, the 850 hPa flow obtains a northerly component over the central and eastern interior (to the east of the surface trough) with southerly components dominating over the western parts (to the west of the surface trough and north of the southern Atlantic Ocean High) and over the north-eastern parts (south-easterly trades to the north of the southern Indian Ocean High). During winter the high pressure cell over the north-eastern parts builds while the surface trough shifts to the west coast. As a result northerly components, which also tend to be stronger, spread across the whole western and central interior (Figure 3.14 bottom).

As mentioned previously the 500 hPa flow is dominated by the upper-air high pressure system. Southerlies predominate over the eastern and northern parts of the

subcontinent (Figure 3.15) while the northerly winds tend to intensify and invade the south-western parts during winter.

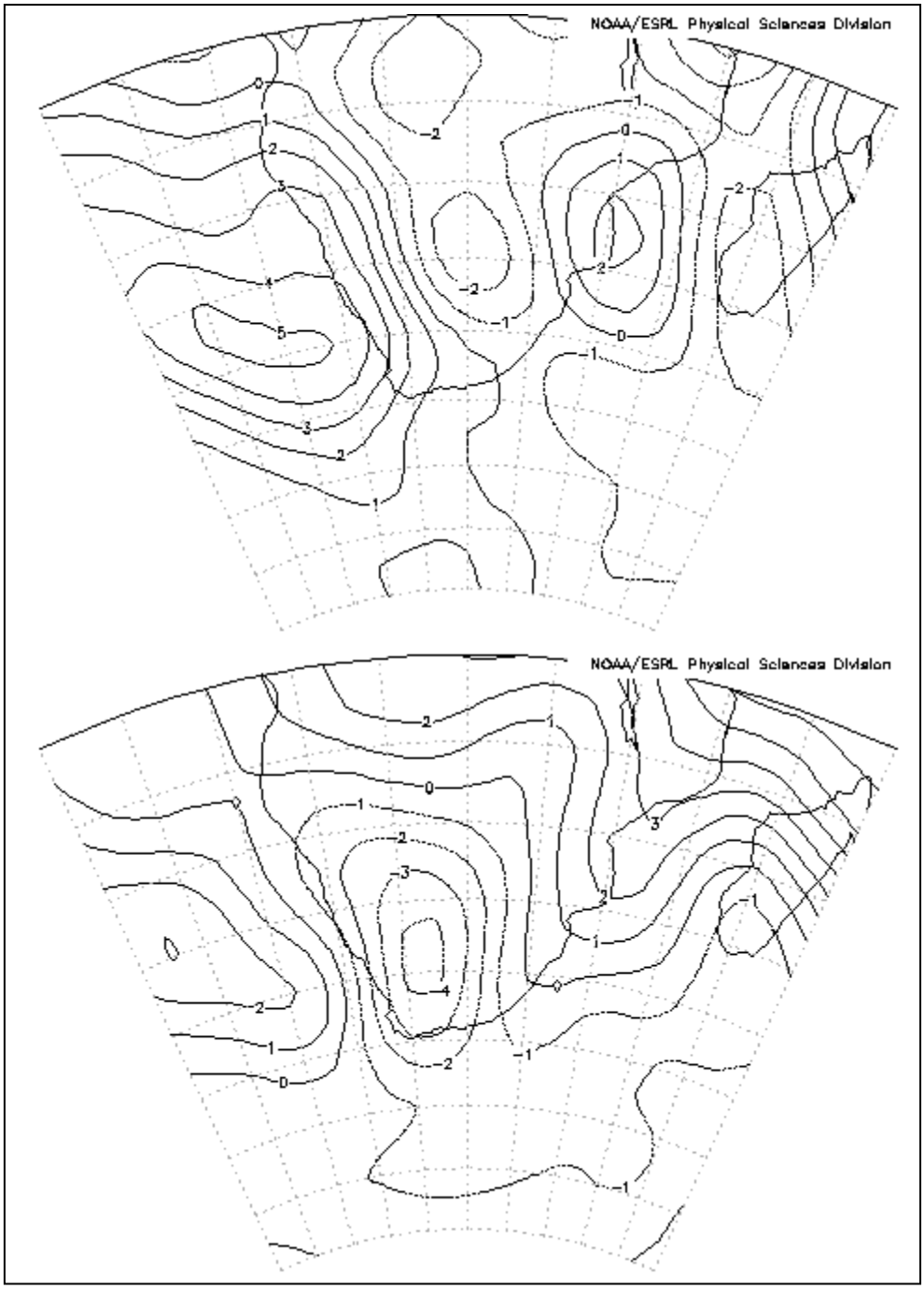

Figure 3.14: Mean 850 hPa meridional flow (in ms^{-1}) for January (top) and July (bottom) calculated for the 1968-1996 period. Positive values indicate southerly winds (NOAA, 2008)

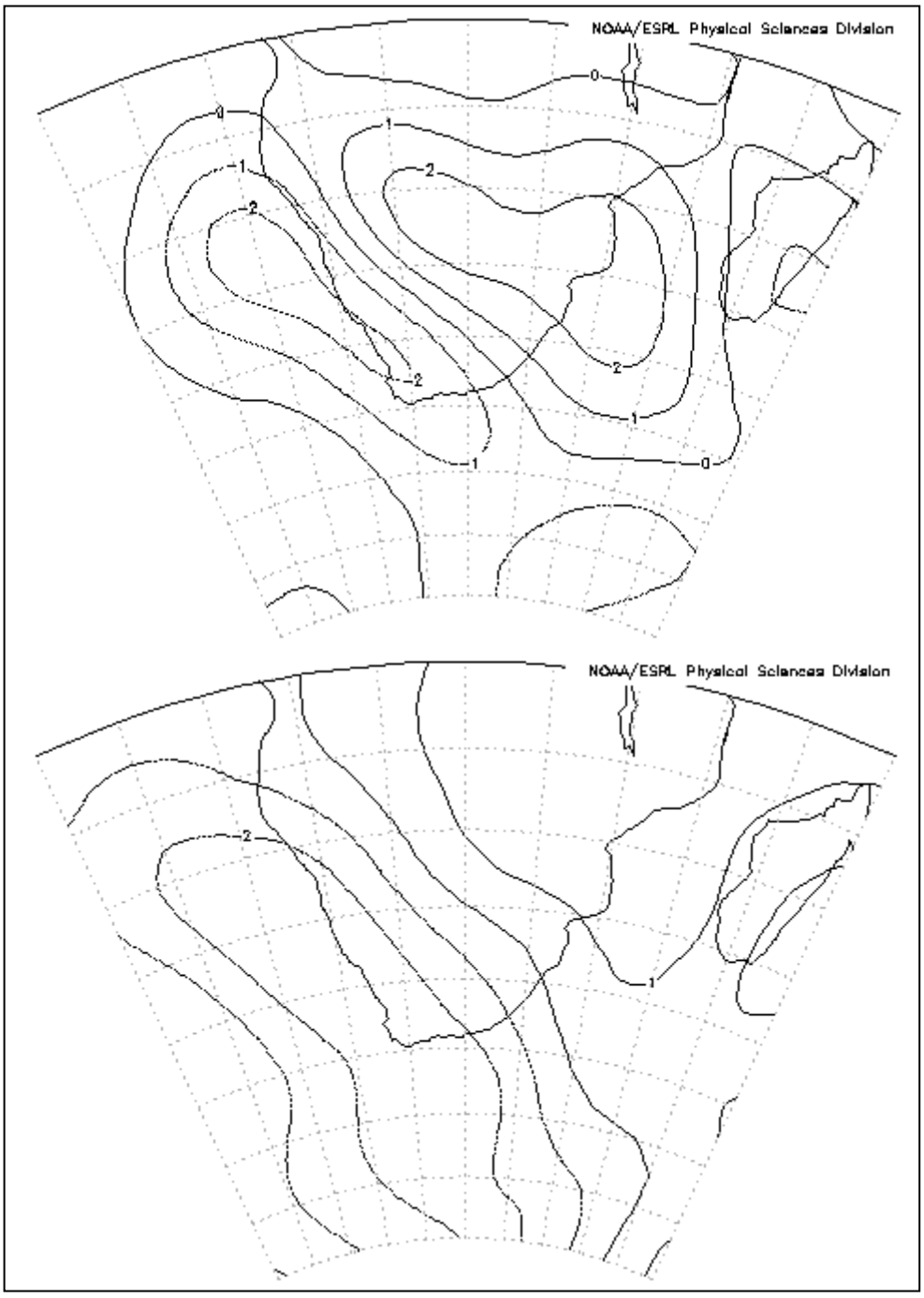

Figure 3.15: Mean 500 hPa meridional flow (in ms^{-1}) for January (top) and July (bottom) calculated for the 1968-1996 period. Positive values indicate southerly winds (NOAA, 2008)

Figure 3.16 is a Hovmoller diagram of the monthly long-term mean meridional wind components at a point over the Upper Olifants catchment, viz. 25 ºS 30 ºE. From this diagram it can be seen that the upper-air flow over this area is dominated by southerly-component winds which stretch all the way to the surface from mid January to June.

Low-level northerly component winds invade the region from mid July to December. It is the northerly component winds that advect high-energy tropical air into the region during the warm season when instability and uplift may lead to cloud development and precipitation.

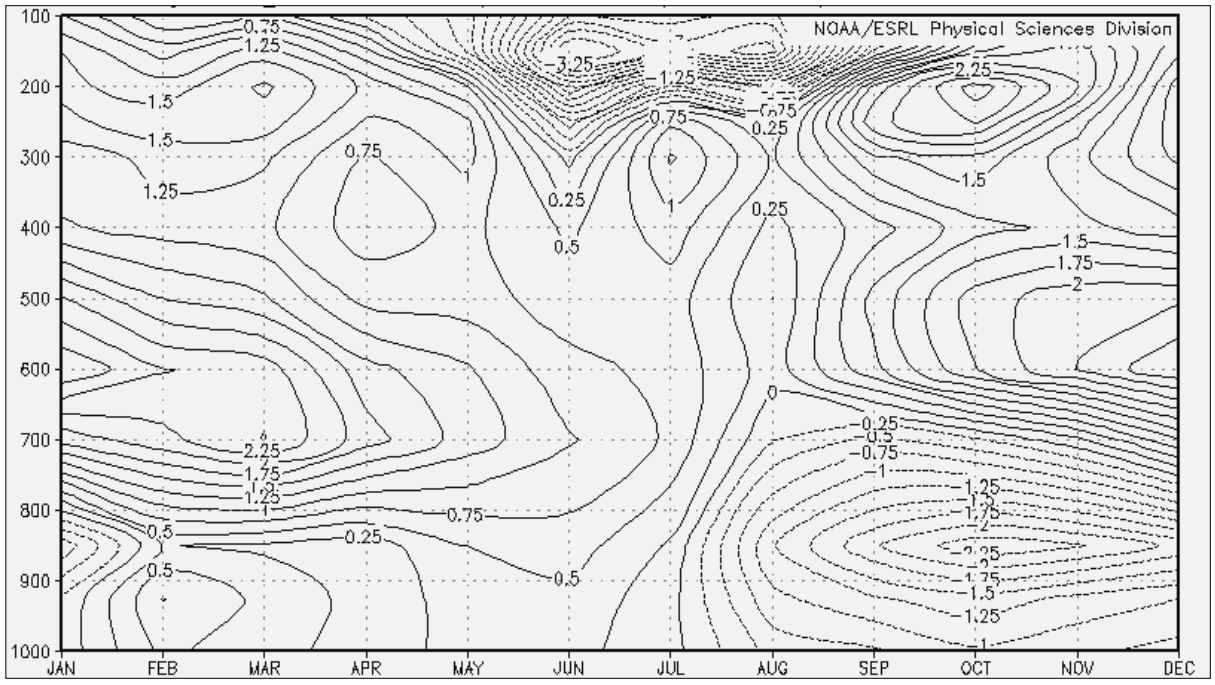

Figure 3.16: Time section of monthly mean meridional flow (in ms^{-1}) at 25ºS 30ºE calculated for the 1968-1996 period. Positive values indicate southerly winds (NOAA, 2008)

3.3.3.4 Relative humidities

Due to the proximity of the AOH near the west coast of South Africa (Figure 3.7), dry conditions are experienced over the western parts due in part to the shorter fetch of the southerly component winds (Figure 3.14) over a cold ocean and drying of the air due to subsidence. In contrast, the trade winds invading the eastern parts of the subcontinent are moisture laden. During the summertime moist air from the tropical latitudes are also allowed to be advected southwards over the central and eastern interior (Figure 3.17 top). In the winter the source of tropical moisture is far removed and very dry conditions are experienced over the interior (Figure 3.17 bottom). At 500 hPa the main moisture sources are the westerly waves of the mid-latitudes in the winter and a combination of

these westerly waves and tropical easterly waves in the summer. This is reflected in the long-term mean relative humidities depicted in Figure 3.18.

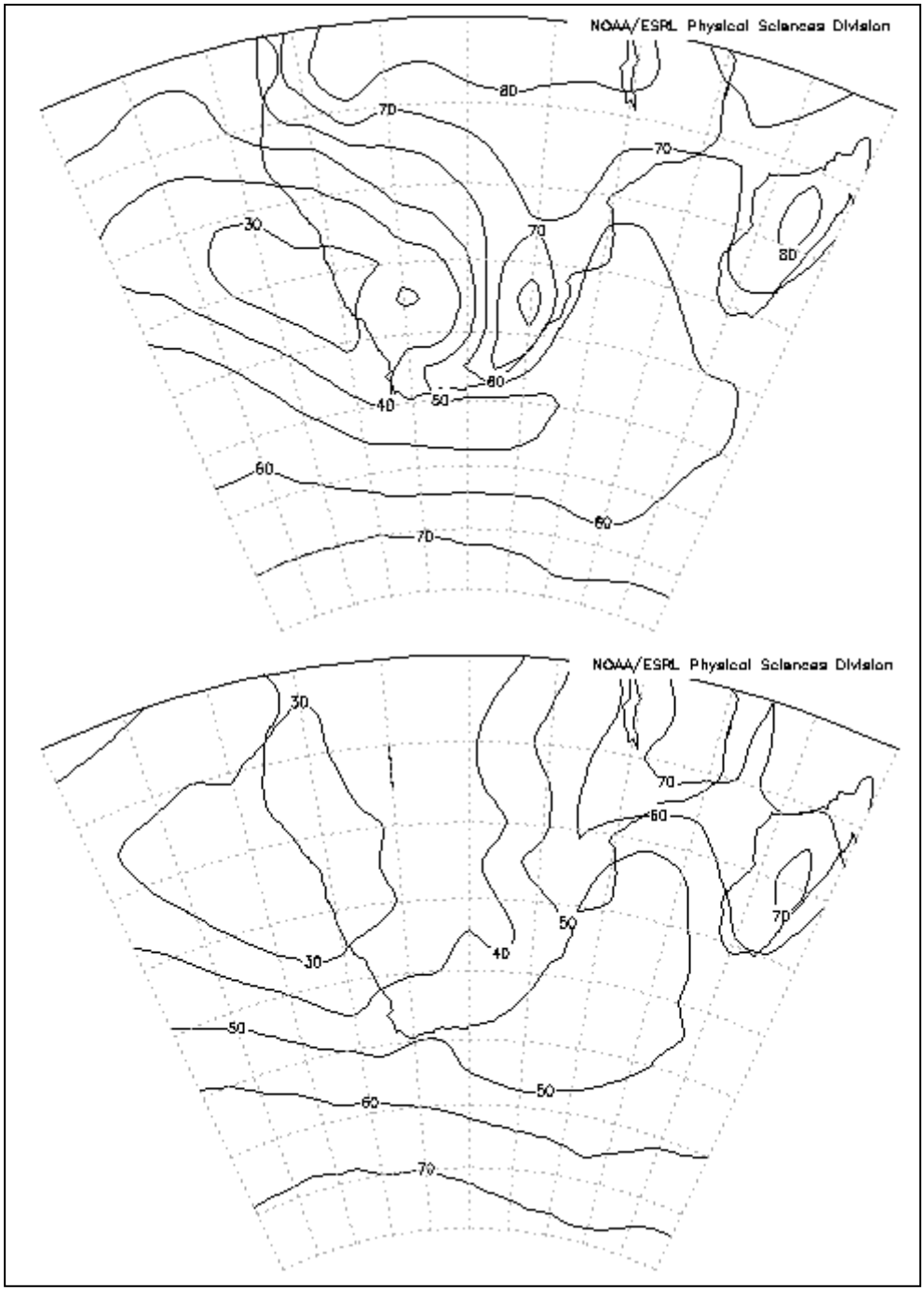

Figure 3.17: Mean 850 hPa relative humidities (as a %) for January (top) and July (bottom) calculated for the 1968-1996 period (NOAA, 2008)

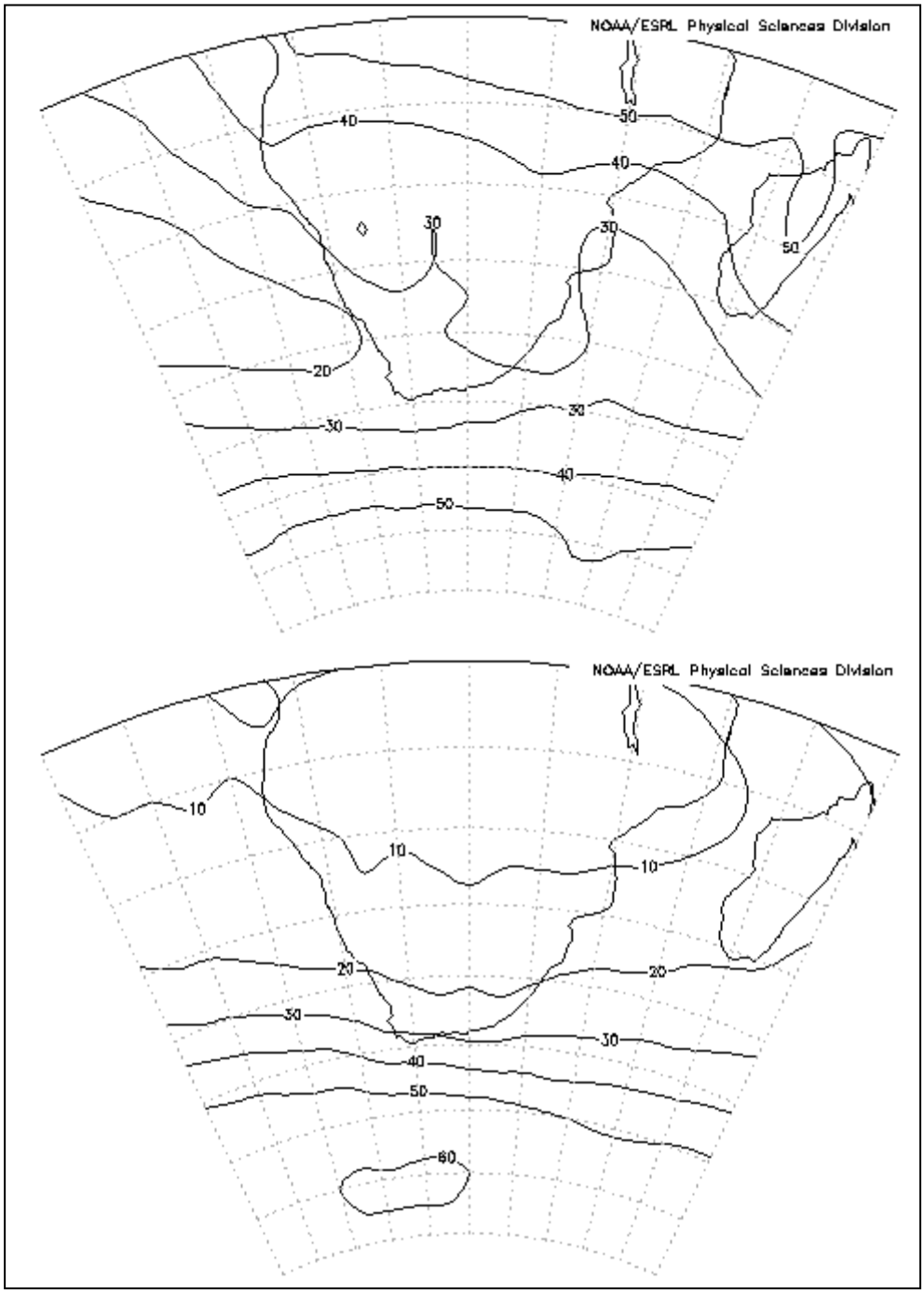

Figure 3.18: Mean 500 hPa relative humidities (as a %) for January (top) and July (bottom) calculated for the 1968-1996 period (NOAA, 2008)

In summary, the study area is strongly influenced by the south-easterly trade winds in summer (which sometimes curve north-easterly to the east of the surface trough). These moisture-laden winds are to a large degree responsible for advecting low-level

moisture into the area which is required for thunderstorm development. The unique topography with mountains to the east has a strong influence on low-level moisture convergence and resultant precipitation – albeit in the form of drizzle along the escarpment or thundershowers over the Highveld. A dry, anticyclonic circulation dominates the region during the winter while westerly-component surface winds sporadically invade the region with the passage of each frontal trough.

CHAPTER 4 CLIMATE DATA

4.1 Observed Predictand Data

The predictand in this study is daily rainfall over the selected QCs within the larger Upper Olifants River Catchment. In order to develop a statistical downscaling model for deriving daily rainfall at a specific point location (in this case the specified QC) from course-grid GCM data, records of historically observed daily rainfall are required for each location.

The record of quality controlled daily rainfall data for each quaternary catchment, developed by Lynch (2004) as part of a project sponsored by the Water Research Commission (WRC), was used as the observed predictand data. Lynch (2004) compiled a comprehensive 50-year database (1950 - 1999) of infilled rainfall station data. This dataset was kindly provided by Trevor Lumsden of the School of Bioresources Engineering and Environmental Hydrology, University of KwaZulu-Natal.

The initial daily and monthly rainfall datasets used in the study by Lynch (2004) were acquired from datasets that had been developed for another WRC project by Dent *et al*. (1989) and had been updated annually till 2000. That information was obtained from an assortment of organisations and individuals that include, among others, the South African Weather Service (SAWS) which also supplied the data for Lesotho and Swaziland (8 281 stations), the Agricultural Research Council (ARC) (2 661 stations), the South African Sugarcane Research Institute (SASRI) (161 stations) as well as a number of municipalities, private companies and individuals (1 050 stations). All of these stations have not collected data concurrently (Lynch and Schulze, 2006). Figure 4.1 indicates the distribution of SAWS observation stations in 2007 alone.

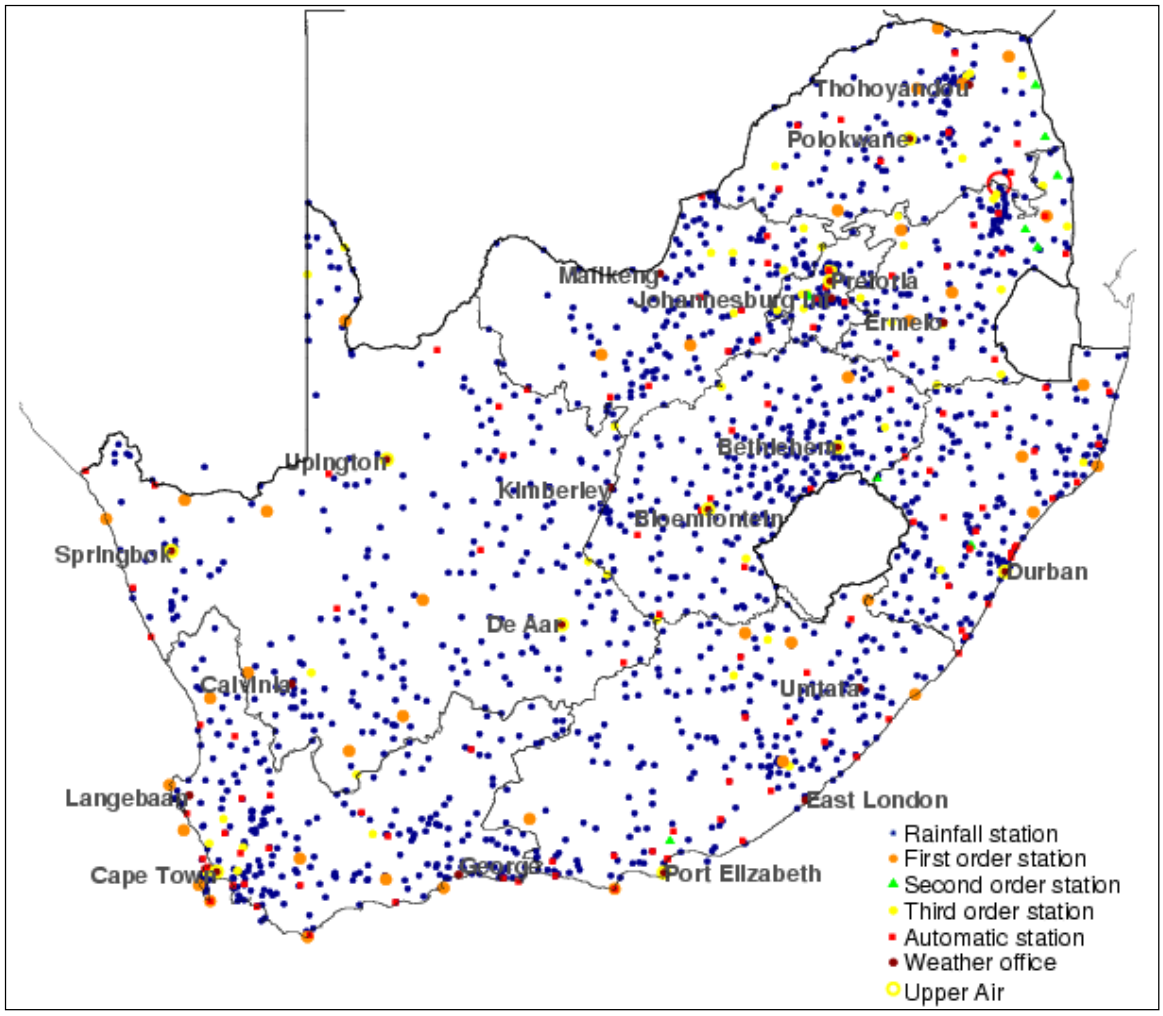

Figure 4.1: SAWS observational station network in 2007 (provided by Colleen de Villiers of SAWS, 2007)

It is important to note that the amount of rainfall reported at a station on a particular day is the accumulated rainfall in a 24-hour period ending at 06:00 GMT. For the period in question (1950 -1999) the SAWS and ARC automatically allocated the 24-hour rainfall measured at 06:00 GMT to the previous day during which the bulk of it may have fell (personal communication, de Villiers (2007)).

The data were checked by Lynch and Schulze (2006) for incorrect recording of the time and date at which the gauge was read and suspect extreme daily rainfall events. In the case of missing records, four infilling techniques were employed, viz. an expectation maximisation algorithm, a median ratio method, an inverse distance weighting and a

monthly infilling technique for rainfall less than 2 mm (Lynch, 2004). Infilling of missing values more than doubled the size of the daily rainfall database from 105 753 218 daily observed values to a total size of 341 908 152 values (Lynch & Schulze, 2006). According to Lynch and Schulze (2006) the infilling process also augmented the size of the annual database significantly from an initial 5 118 stations with more than 15 years with a complete record to a staggering 9 641 stations that have more than 15 years of measurements.

It is important to note that precipitation is intermittent and highly variable in space and time, while local topographic and environmental conditions play a key role in the distribution thereof (Maini *et al.*, 2004). After missing rainfall values had been filled in and the station records extended by the different infilling techniques, Lynch (2004) employed a Geographically Weighted Regression (GWR) approach in order to estimate rainfall values at those points on a raster grid where no stations with observed data or infilled values existed.

For each QC, Lynch (2004) determined a centroid using ArcView GIS. The Daily Rainfall Extraction Utility (DREU) developed by Kunz (2004, cited in Lynch & Schulze, 2006) was used to extract the ten nearest rainfall stations to each pair of the centroid's co-ordinates. These 10 stations are ranked by the DREU using a number of criteria including the distance from the rainfall station to the point of interest, station recording period and reliability (i.e. the percentage of actual data vs. infilled values). The best ranked station was selected as the so-called "driver" rainfall station, with that station's data considered representative of the daily rainfall of that QC (Lynch & Schulze, 2006).

4.2 Observed Predictor Data

Large-scale observed predictor data were derived from reanalysis data supplied by the National Centers for Environmental Prediction (NCEP). NCEP is an arm of the United States of America's National Weather Service (NWS) and is comprised of nine distinct Centres which provide a wide variety of national and international weather guidance products to NWS field offices, government agencies, emergency managers, private sector meteorologists, and meteorological organizations and societies throughout the world (NCEP, 2008).

Observational data are continuously gathered by NCEP from a wide range of sources (Kalnay *et al.*, 1996), including global rawinsonde data (since 1948), the Comprehensive Ocean Atmosphere Data Set (COADS) which includes ships, buoys and near surface data from ocean station reports (since 1947), aircraft data from the Global Telecommunication System (GTS) and a number of research experiments (since 1962), surface land synoptic data from the GTS and the U.S. air force (since 1949), satellite sounder data (since 1979), special sensing microwave imager data (since 1987) and satellite cloud drift winds from geostationary meteorological satellites (since 1978).

Basically, the NCEP reanalysis system comprises the following three major modules (Kalnay *et al.*, 1996):

- Data decoder and a quality control pre-processor;
- Data assimilation module with an automatic monitoring system;
- Archive module.

The data input is pre-processed. Optimal interpolation quality control of all data is done in order to identify and withhold data containing gross errors produced by instrumental, human, or communication-related mistakes that may occur during the process of making or transmitting observations (Kalnay *et al.*, 1996). Observations that are accurate but with large errors of representativeness whose measurements represent

spatial and temporal scales impossible to resolve properly in the analysis forecast system, will also be withheld.

The central module is the data assimilation. The NCEP reanalysis system uses a frozen T62 (equivalent to a horizontal resolution of 210 km) global spectral model with a state-of-the-art spectral statistical interpolation as the analysis module (Kalnay *et al.*, 1996). The model has 28 levels in the atmosphere, 5 of which are in the boundary layer and about 7 are above 100 hPa (Kalnay *et al.*, 1996). The model contains parameterisations of main physical processes such as convection, large-scale precipitation, gravity wave drag, radiation with diurnal cycle, interaction with clouds, boundary layer physics, surface hydrology and diffusion processes (Kalnay *et al.*, 1996). Further details of the NCEP model dynamics and physics can be found in Kanamitsu (1989) and Kanamitsu *et al*. (1991).

The NCEP reanalysis system includes not only the computation of grid point values, but also temporal and spatial averages over prescribed areas (Kalnay *et al.*, 1996). Optimal averages are computed for temperature, specific humidity, zonal and meridional wind components and wind speeds at the 1 000, 850, 700, 500, 300, 200 and 100 hPa levels. The reanalysis output is available for 00:00, 06:00, 12:00 and 18:00 GMT on a regular 2.5º × 2.5º grid. All analysis output fields are monitored with a complex quality control system, wherein the data statistics and time tendencies are compared to climatological statistics in order to detect errors (Kalnay *et al.*, 1996).

The NCEP reanalysis products have been interpolated onto the same grid as the GCM to be described in section 4.4, i.e. 2.5º × 3.75º. The reanalysis products are available for the calibration procedure of the Statistical Downscaling Model (SDSM) over the current climate period (1961 - 2001). All NCEP data have been averaged on a daily basis from 6 hourly data, before being linearly interpolated to match the GCM data. Where variables are derived, they are computed on the resident 2.5° x 2.5° regular grid, and then interpolated (Gachon *et al*., 2008). The list of predictors has been chosen according to the data availability and to correspond to the same physical variables

issued from the GCM predictors listed in section 4.4. All the data included are of quality ‘A’ or ‘B’, which means that they are influenced directly (to some extent) by observational data (Kalnay *et al.*, 1996; Gachon *et al*., 2008).

The mean and standard deviations for the observed predictor variables were calculated for the base period 1961 - 1990. The means were subtracted from each daily value before dividing by the standard deviation. In this way the predictor variables were standardised (normalised) with respect to the set base period. Similar operations were carried out on the observed predictand and GCM-derived predictor variables. Since GCMs do not always perform well at simulating the climate of a particular region this means that there may be large differences between observed and GCM-simulated conditions (i.e. GCM bias or error) which could potentially violate the statistical assumptions associated with SDSM and give poor results if the predictor data were not normalised (CICS, 2008). The normalisation process guarantees that the distributions of observed and GCM-derived predictors are in closer agreement than those of the raw observed and raw GCM data (CICS, 2008).

NCEP predictor data were downloaded for selected grid boxes from the Data Access Integration (DAI) website (CICS, 2008). This DAI portal provides 41 years of daily observed predictor data covering the period 1961 – 2001, derived from the NCEP reanalysis, interpolated to the same grid as HadCM3 before normalisation was implemented by the CICS scientists.

4.3 IPCC SRES Scenarios

Greenhouse gas (GHG) emissions are the product of extremely complex dynamic systems. Future emissions are highly uncertain and determined by driving forces such as demographic development, socio-economic development, and technological change (Nakićenović *et al*., 2000). By the end of the century the world will have changed in ways that are difficult to imagine. The Intergovernmental Panel on Climate Change

(IPCC) hence developed long-term emissions scenarios for use in climate change analysis, including climate modelling and the assessment of impacts, adaptation and mitigation. These scenarios are described in detail in the IPCC Special Report on Emission Scenarios (SRES) (Nakićenović *et al*., 2000). A brief description of these scenarios is included here as the description of the GCM data, which follows in section 4.4, refers to them.

According to Nakićenović *et al*. (2000) the scenarios are alternative descriptions of how the future might unfold and are a suitable tool with which to analyse how driving forces may affect future emission outcomes and to gauge the related uncertainties. The prospect that any single emissions scenario will transpire as described is highly uncertain. Four divergent storylines were developed to describe the relationships between emission driving forces and their evolution. These storylines add background for the scenario quantification (Nakićenović *et al*., 2000).

Each storyline embodies varying demographic, social, economic, technological, and environmental developments. All the scenarios based on the same storyline make up a “scenario family” (Nakićenović *et al*., 2000). The scenarios do not include further climate initiatives i.e. no scenarios are included that explicitly assume implementation of the United Nations Framework Convention on Climate Change or the emissions targets set by the Kyoto Protocol. Table 4.1 summarises the main characteristics of the four SRES storylines and scenario families as described by Nakićenović *et al*. (2000), while Figure 4.2 illustrates the range of GHG emissions in the SRES scenarios.

Table 4.1: Main characteristics of the four SRES storylines and scenario families (adapted from Nakićenović *et al*., 2000)

SRES Scenario	A1	A2	B1	B2
Economy	rapid growth, substantial reduction in regional differences in per capita income	development is primarily regionally oriented, per capita economic growth are more fragmented and slower	rapid changes in economic structures toward a service and information economy, with reductions in material intensity	intermediate levels of development
Global Population	peaks in mid-century and declines thereafter	continuously increasing	peaks in mid-century and declines thereafter	continuously increasing (slower than A2)
Technology	rapid introduction of new and more efficient technologies	technological changes are more fragmented and slower than in other storylines	introduction of clean and resource-efficient technologies	more diverse technological change than in the B1 and A1
Socio-Political Aspects	increased cultural and social interactions, capacity building	self-reliance and preservation of local identities	improved equity, global solutions to economic, social, and environmental sustainability	local solutions to economic, social, and environmental sustainability
Underlying Themes	convergence among regions	heterogeneous world	convergent world	focuses on local and regional levels

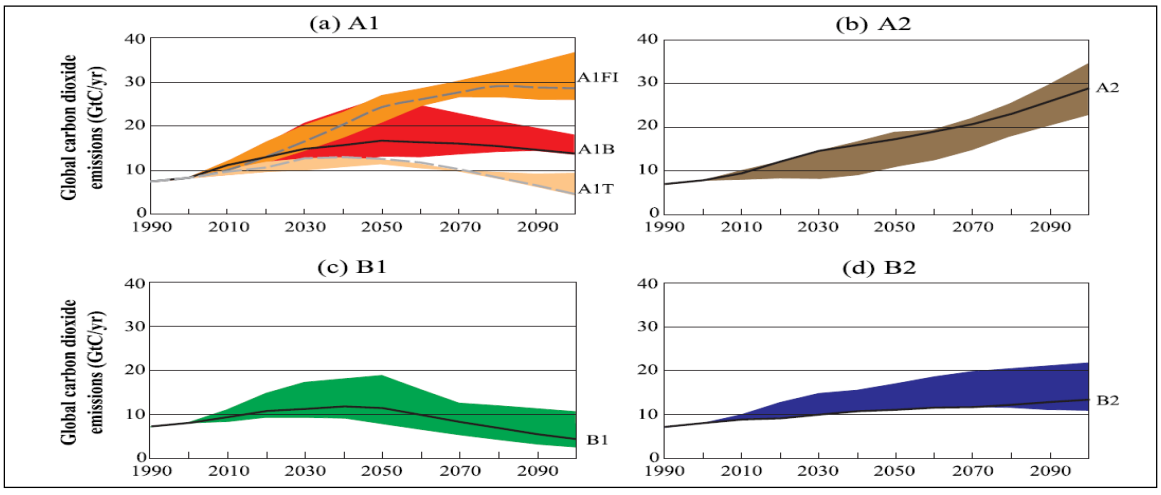

Figure 4.2: Total global annual CO_2 emissions from 1990 to 2100 (in gigatonnes of carbon (GtC/yr)) for the various scenario groups. Capital letters refer to the four scenarios described in Table 4.1. Each coloured emission band shows the range of projections within each group (Nakićenović *et al*., 2000)

4.4 Global Circulation Model Data

Coupled ocean-atmosphere GCMs have become valuable instruments in attempting to comprehend and predict climate change (Houghton *et al*., 1996). Downscaling scenarios using more than one GCM running different experiments (i.e. SRES scenarios) may produce somewhat different but equally plausible results. Impact researchers working at site specific scales would ultimately benefit by comparing downscaled projections from two or more scenarios (Lines *et al*., 2005). For this reason, and due to availability, the choice of GCM data fell on the HadCM3 model running the SRES A2 and B2 scenarios. A concise description of the model is provided below.

The Third Generation Hadley Centre Coupled Model (HadCM3) is a coupled atmosphere-ocean GCM developed at the Hadley Centre for Climate Prediction and Research in the UK. It was one of the major models used in the IPCC Third Assessment Report in 2001. It has stable control climatology and does not require flux adjustments to prevent large climate drifts in the simulation (Gordon *et al*., 2000).

The atmospheric component of the model has a horizontal resolution of 2.5º of latitude by 3.75º of longitude, which produces a global grid of 96 x 73 grid cells (Pope *et al*., 2000). This is equivalent to a surface resolution of about 417 km x 278 km at the Equator, reducing to 361 km x 278 km at 25º S (over the study area). HadCM3 has 19 vertical levels using a hybrid vertical co-ordinate and employs a 30 minute time step (WDC, 2008). HadCM3 has model years consisting of 360 days in each year (Wilby & Dawson, 2007), i.e. each month has 30 days and leap years are not included.

Some of the key features that are new to the third generation model include the introduction of a new radiation scheme with 6 shortwave and 8 longwave spectral bands, a new land surface scheme that includes a representation of the freezing and melting of soil moisture as well as surface runoff and soil deep drainage, modifications to the penetrative convective scheme in the atmosphere to include an explicit downdraught and the direct impact of convection on momentum (WDC, 2008).

Revisions were also made to the parameterisations of orographic and gravity wave drag. The atmosphere component of the model also optionally allows the transport, oxidation and removal by physical deposition and precipitation scavenging of anthropogenic sulphur emissions (WDC, 2008). The atmospheric component is described in more detail in Pope *et al*. (2000) and Gordon *et al*. (2000).

The atmosphere and ocean exchange information once per day. The oceanic component of the model has 20 levels with a horizontal resolution of 1.25 x 1.25 degrees (WDC, 2008). At this resolution it is possible to represent important details in oceanic current structures. The sea ice model uses a simple thermodynamic scheme while ice is advected by the surface ocean current. The ocean component is described in more detail in Gordon *et al*., (2000).

HadCM3 data were first downloaded for the grid box overlying the study area (i.e. 25ºS 30ºE) from the DAI website (CICS, 2008). At a later stage HadCM3 data were also downloaded for all eight grid boxes surrounding the selected grid box in order to explore the effects of expanding the predictor domain. The DAI portal provides the following data subsets:

- 139 years of daily GCM predictor data covering the period 1961 – 2099, derived using the SRES A2(a) scenario; and
- 139 years of daily GCM predictor data covering the period 1961 – 2099, derived using the SRES B2(a) scenario.

All the data (except in the case of wind directions) have been normalised with respect to the 1961 – 1990 period. The value in each grid box or cell (defined over an area of 3.75º longitude and 2.5º latitude) corresponds to the value over the centre of the cell (CICS, 2008). Table 4.2 provides a summary of the 26 HadCM3 predictors which were available to this study. The units refer to the units of the climatic element before normalisation.

Table 4.2: Description of HadCM3 predictors (adapted from Wilby *et al*., 2002; Wilby & Dawson, 2007)

Levels	Climatic Element	Unit	Calculation Method
2 m	Temperature	ºC	Interpolated at 2 m from the lowest model level
MSL	Pressure	hPa	Calculated from surface pressure, temperature and surface geopotential Averaged over the 6-h daily values
850 hPa 500 hPa	Geopotential height	m^2s^{-2}	Calculated from surface pressure, temperature and specific humidity Averaged over the 6-h daily values
Near surface	Specific humidity	g kg^{-1}	Calculated from the model's humidity variable Averaged over the 6-h daily values
Near surface 850 hPa 500 hPa	Relative humidity	%	Calculated from the model's humidity variable Averaged over the 6-h daily values
Near surface 850 hPa 500 hPa	Wind direction	º from N	Calculated from averaged 6-h daily geostrophic winds
Near surface 850 hPa 500 hPa	Geostrophic airflow	hPa	Calculated from zonal and meridional geostrophic wind components Averaged over the 6-h daily values
Near surface 850 hPa 500 hPa	Zonal airflow component	hPa	Geostrophic – calculated from pressure gradients Averaged over the 6-h daily values
Near surface 850 hPa 500 hPa	Meridional airflow component	hPa	Geostrophic – calculated from pressure gradients Averaged over the 6-h daily values
Near surface 850 hPa 500 hPa	Divergence	s^{-1}	Calculated from averaged 6-h daily geostrophic winds
Near surface 850 hPa 500 hPa	Vorticity	s^{-1}	Calculated from averaged 6-h daily geostrophic winds

CHAPTER 5 METHODOLOGY

5.1 Process Description

The methodology used in this study follows the procedure outlined by Lines and Barrow (2002), Wilby *et al*. (2002) and Lines *et al*. (2005). The study makes use of the Statistical Downscaling Model (SDSM) developed by Wilby *et al*. (2002) who also used SDSM to develop single-site ensemble scenarios of daily rainfall under current and future regional climate forcing for Toronto, Canada. Lines *et al*. (2005) used SDSM to downscale the expected climate change impacts with respect to daily mean, maximum and minimum temperature as well as precipitation for 14 sites across Atlantic Canada. In a more recent study Wilby *et al*. (2006) also used SDSM to downscale daily temperature, precipitation and potential evaporation for the River Kennet in the UK.

Within the nomenclature of downscaling techniques SDSM is best described as a hybrid of the stochastic weather generator and regression-based methods (Wilby *et al*., 2002). The SDSM software reduces the task of statistically downscaling daily rainfall into the following discrete steps (Wilby *et al*., 2002):

a) quality control and data transformation;
b) screening of predictor variables;
c) model calibration;
d) weather generation (using observed predictors);
e) generation of climate change scenarios (using climate model predictors); and
f) statistical analysis.

A perfect prognosis (PP) approach is followed, where the forecast or simulated predictor variables are taken at face value – assuming them to be perfect. In model calibration, observed predictors (in the form of NCEP reanalysis data) are used to describe the observed predictand (in the form of daily rainfall data for the selected QCs). In implementation, it is the GCM simulation of the predictors that are substituted into the

regression equation. For more information on the PP approach the reader is referred to Section 2.3.4.2.

5.2 Fundamental Assumptions

From the preceding process description it is apparent that the assumption has to be made that it is indeed possible to derive significant relationships between the potential predictor variables and the predictand over the selected study area. The derived relationship must be valid at the synoptic scale as that is the spatial scale used to condition the downscaled response (Hewitson & Crane, 1996; Wilby *et al*., 2004; Hessami *et al*., 2008).

Although SDSM has a built in function for variance inflation and bias correction, the underlying assumption of PP is that the GCM simulations of predictor variables are accurate. This assumption is necessary in order to apply the same regression equations (that are based on historical observations) to simulated predictor values. It also stands to reason that the large-scale potential predictors must be variables that are well simulated by the GCMs. In principle, the same PP regression equations can be used with different GCMs, scenarios, or projections (Wilks, 1995). However, new equations would have to be trained for different study areas in order to accommodate regional forcings.

Another major assumption is time invariance. It must be assumed that the relationships that have been established between the predictors and predictand under the current climatic period will also remain valid during future climatic periods outside the fitting period. Many climate change studies inherently make this assumption (Hewitson & Crane, 1996; Wilby & Wigley, 1997; Wilby *et al*., 2004) as it is unlikely that the physical links between weather systems and the predictand will change. However, it remains a point of concern as the established relationships are only valid within the long-term variability of the observational data used for calibration while future projections for some

variables may lie outside of this range (Wilks, 1995; Wilby *et al*., 2004). Furthermore, one has to assume that the predictor set sufficiently incorporates the future climate change signal (Wilby *et al*., 2004).

5.3 Stratification and Transformation of Data

The physical and/or statistical relationships between predictors and predictand may change with seasons (Wilks, 1995). One approach to addressing seasonal differences is to include predictors that are functions of the day of the year. In Section 3.3 it was indicated that the rainfall in the selected QCs across the study area is highly seasonal with the bulk of the rainfall occurring in the summer months of December, January and February (DJF). For this reason the decision was made to stratify the observed data according to the time of year, and to produce forecast equations specifically for this three month summer rainfall season.

Rainfall data usually follows a positively skewed distribution. Power transformations can be useful for converting atmospheric data to conform to the assumptions of regression analysis (Wilks, 1995). In an attempt to downscale precipitation over India, Maini *et al.* (2004) applied the square root to their rainfall data in order to make the distribution nearly symmetrical. According to Wilby and Dawson (2007) fourth root, natural log and inverse normal transformations are used whenever data are skewed. After comparing transformations involving the natural log and powers of two, three, four as well as square, third and fourth roots, the decision was made to apply a fourth root transformation to the predictand data prior to model calibration. It is hoped that a symmetry-producing transformation will allow the application of multivariate statistical methods that may assume normal distributions (Wilks, 1995).

Large-scale observed predictor data were derived from NCEP reanalysis data. In most spatial downscaling studies, the predictor data used are first normalised with respect to the period mean and standard deviation, rather than using the actual data themselves

(CICS, 2007). The idea behind the normalisation (or standardised anomalies) is to attempt to remove the influences of location and spread from a batch of data (Wilks, 1995). The physical units of the original data cancel, so normalised values are dimensionless quantities with a mean of zero and a standard deviation of 1. An alternative way to view normalisation is as a measure of distance, in standard deviation units, between a data value and its mean (Wilks, 1995). The NCEP predictor data were downloaded for selected grid boxes from the DAI website. The downloaded data has already been interpolated to the same grid as the HadCM3 model before normalisation with respect to the base period of 1961 – 1990.

5.4 Predictor Selection

The selection of predictor variables is one of the most important steps in the development of a statistical downscaling scheme because the choice largely determines the character of the downscaling results (Hessami *et al*., 2008). According to Wilby *et al.* (2002; 2004) the identified large scale climate predictor variables should be:

- physically and conceptually sensible with respect to the site variable (the predictand);
- strongly and consistently correlated with the predictand;
- readily available from archives of observed data and GCM output; and
- accurately modelled by GCMs.

According to Wilby *et al*. (2004) the selection process is complicated by the fact that the explanatory power of individual predictor variables may be low – especially for daily rainfall. This explanatory power may also vary from one month to the next and from one location to the next.

In accordance with the study done by Lines *et al*. (2005), candidate predictors from the GCM grid-box overlying the study area will be used. There are almost always more potential predictors available than can be used in a regression, and finding good

subsets of these in particular cases is not as easy as one might first imagine. The process is definitely not as simple as adding members of the list of potential predictors to the regression until an apparently good relationship is achieved (Wilks, 1995). Surprisingly, there are dangers associated with including too many predictor variables in a forecast equation. Wilks (1995) provides an example where any $K = n - 1$ predictors will produce a perfect regression fit to any predictand for which there are n observations. This is easiest to see for the case of $n = 2$, where a straight line can be fit using $K = 1$ predictor (simple linear regression), since a line can be found that will pass through any two points in the plane, and only an intercept and a slope are necessary to define a line. This problem, referred to as “overfitting”, generalises to any sample size. Overfitting usually manifests itself as an apparent excellent fit on the training data, while the fitted relationship falls apart when used with independent data not used in the development of the equation.

Wilks (1995) provides several important lessons that can be drawn from this example:

- Begin development of a regression equation by logically choosing only physically sensible or meaningful potential predictors.
- A tentative regression equation needs to be tested on a sample of data not involved in its development. One way to approach this important step is to simply reserve a portion of the available data as the independent verification set, and fit the regression using the remainder as the training set. The performance of the resulting regression equation will nearly always be better for the dependent than the independent data, since the coefficients have been chosen specifically to minimise the squared residuals in the developmental sample.
- One needs a reasonably large developmental sample if the resulting regression equation is to remain stable. While the number of regression coefficients that can be estimated with reasonable accuracy increases as the sample size increases, in forecasting practice it is often found that there is little to be gained from including more than about a dozen predictor variables in a final regression equation.

Studies by Wilby *et al.* (1998), Huth (2004) and Hessami *et al.* (2008) have shown that, when downscaling precipitation, a combination of circulation (e.g. geopotential heights or wind components), temperature and moisture (e.g. relative humidity) predictors is better than any single predictor. Hessami *et al*. (2008) used an automated statistical downscaling regression-based approach to reconstruct the observed daily precipitation and temperature series in eastern Canada. In their attempts to model the precipitation, Hessami *et al*. (2008) found that the most common combination of predictor variables for daily precipitation modelling are relative/specific humidity at 500 hPa, near-surface wind speed, 850 hPa zonal velocity and 500 hPa geopotential height. Maini *et al*. (2004) developed an operational model for forecasting location specific quantitative precipitation (24 hour accumulated) and probability of occurrence over India thirty days in advance. They observed that mean relative humidity (1000-500 hPa level) and meridional wind component (850 hPa) are frequently selected for the development of quantitative precipitation forecast equations, while the vorticity at 850 hPa is also important. According to Hessami *et al.* (2008), specific and relative humidity are not interchangeable, but they are strongly correlated. As their synchronous variation relies on the saturation phase of water vapour in the atmosphere, both are highly correlated to the occurrence of precipitation. Hence, using either relative or specific humidity should provide similar results for downscaling of precipitation since high relative humidities are always accompanied by commensurately high specific humidities. According to Hessami *et al.* (2008), the combination of humidity variables at various levels are often more significant to the precipitation process (occurrence and intensity) than a single value of humidity taken at a solitary level.

Table 4.2 provided a summary of all the GCM predictors available to this study. In each instance a parallel set of NCEP predictors, in addition to their 1-day lagged values, will be used in the model calibration and subsequent generation of ensemble members (for the current climatic period). Since SDSM can only screen 12 predictors at a time, the 52 potential predictors were evaluated in subsets according to the nature of the predictors. In this way a subset were created for the pressure/height variables (e.g. mean sea-level pressure, 850 and 500 hPa geopotential heights), low-level wind components (e.g.

surface and 850 hPa winds), low-level vorticity and divergence predictors and upper-air predictors (e.g. 500 hPa wind components, vorticity and divergence). In this way similar predictors were compared to each other so that the best ones could be chosen. Vertically integrated moisture flux were not available to this study. Correlation matrices were then calculated for each subset along with the partial correlations and P-values. The correlation matrix is equivalently the variance-covariance matrix of the standardised variables (Wilks, 1995).

During the screening process, all potential predictors were initially awarded a value of 1. For each subset the predictor variables that exhibited relatively high partial correlations with the observed rainfall and P-values in the vicinity of 0.05 or lower were retained for a second round of screening and awarded a value of 2. According to Wilby and Dawson (2007), higher correlation values imply a higher degree of association while smaller P-values indicate that the association is less likely to have occurred by chance. The potential predictors that made it through the first round were subsequently evaluated together in a final round of screening. Again the predictors adhering to the criteria mentioned above were retained as the predictor set for the selected QC. If in the final set of 12, the predictor performed better than its peers, it was then awarded a value of 3. This process of predictor selection was repeated for all 5 of the selected QCs on observational DJF data spanning the period 1961 – 1985. An independent dataset is used in order to minimise the danger of finding a spurious relationship due to overfitting. The allocated values were then added for all the potential predictors across all QCs and the top predictors were thus identified in terms of the highest accumulated values (for both the predictor and its 1-day lag). A physical restraint was placed by SDSM on the number of predictors as only 8 predictors could be included in the final set. This final set of top-scoring predictors thus formed a generic set that could be used for the Upper Olifants River catchment.

5.5 Model Calibration

During model calibration SDSM takes a user–specified predictand (in this case DJF daily rainfall for a specified QC) along with the chosen set of NCEP predictor variables, and computes the parameters of multiple linear regression equations through an ordinary least squares optimisation algorithm (Wilby & Dawson, 2007). This is carried out on independent observational DJF data spanning the period 1961 – 1985 which will not be used during the model validation. In the case of precipitation a conditional process is employed since daily rainfall amounts depend on the occurrence of wet-days, which in turn depend on regional-scale predictors such as humidity, atmospheric pressure and vorticity. This implies sequences of wet or dry days are first modelled, then the amounts of rain if it is a wet-day.

Wet-day precipitation amounts are assumed to be exponentially distributed and are modelled using the regression procedure of Kilsby *et al*. (1998, cited in Wilby *et al*., 2002). According to Wilby *et al*. (2002) the expected mean wet-day amount is empirically forced by the algorithm to equal the observed mean wet-day amount of the calibration period, while serial correlation between successive wet-day amounts may be incorporated implicitly by lagged predictor variables.

Since seasonal models for daily rainfall are developed in this study, all three months in the summer season (December, January and February) will have the same model parameters. In addition, SDSM will calculate residual statistics and display these on a scatter diagram (plot the residuals against the modelled predictor). Finally, the SDSM calibration procedure reports the percentage of explained variance and standard error for the regression model. The magnitude of deterministic forcing is indicated by the percentage of variance explained by the regression model while the significance of the indeterminate or noise fraction by the standard error of the calibrated model (Wilby *et al*., 2002). The Chow test is used to test for structural change in the parameters of a model. It tests for model stationarity by checking whether the coefficients estimated over one group of the data are similar to the coefficients estimated over another. High Chow

statistic values thus indicate that the fitted model may become unstable in future. The model is said to be calibrated when the regression coefficients, explained variance and standard error are within acceptable limits for the regression model (Lines *et al*., 2005).

5.6 Ensemble Generation

Given the set of NCEP predictor variables, an internal random number weather generator takes the calibrated output (model weights) and stochastically synthesises a number of ensembles that are statistically related to the original training dataset (Lines *et al*., 2005). The extent to which time series of different ensemble members differ is determined by the relative significance of the deterministic and stochastic components of the regression models (Wilby *et al*., 2002). Precipitation series, for example, display more 'noise' arising from local factors (Wilby *et al*., 2002). SDSM uses the standard errors to stochastically replicate the distribution of model residuals. A pseudo random number generator, as described by Rubinstein (1981, cited in Wilby *et al*., 2002), replicates values from a normal distribution with standard deviation equal to the calibration standard error. This stochastic value is subsequently added to each day's deterministic component in order to inflate the variance of the downscaled series and to enhance the agreement with daily observations. In the case of conditional processes, the model incorporates an additional stochastic process to determine the probability of precipitation occurring as prescribed by regional forcing. A random number generator is used to determine the outcome - if regional forcing indicates a probability of precipitation occurrence, $p = 0.65$, and the random number generator returns $r \leq 0.65$ the day is wet; alternatively if $r > 0.65$ the day is dry.

Individual ensemble members are considered equally plausible (Wilby *et al*., 2002). The ensemble means can be used to represent the synthesized downscaled values. This procedure also permits the verification of the calibrated models (using independent data) as well as the synthesis of artificial time series for present climate conditions (Wilby & Dawson, 2007). In this study 20 ensemble members were created for the

verification period spanning 1986 – 1999. After model verification, the procedure was repeated for the whole base period of 1960 – 1990 in order to facilitate a comparison between the current and future climates.

5.7 Model Validation

In the field of statistical weather/climate prediction it is customary for the regression equations to be tested on a sample of independent data that has been held back during the development of the forecast equations. In this way, the downscaling model can be verified against observed predictand data and if judged to be acceptable, the equations can be used operationally after recalibration for the whole base period. This procedure is actually a special case of a technique known as cross-validation. Cross-validation is a resampling technique whose mechanics are analogous to the bootstrap and permutation tests (Wilks, 1995). In cross-validation the available observational data are repeatedly divided into calibration and verification data subsets. The forecast model is developed using the calibration subset as input, and subsequently run for the verification period. The resultant model output (downscaled series) are subsequently evaluated against the verification subset. Ideally, sufficient observed data should be available to enable at least 20 years of data to be used to calibrate a spatial downscaling model, with a further 10 years of data (ideally more) being available for model verification (CICS, 2007).

In this study, cross-validation is carried out by splitting the observational DJF data into a calibration set spanning the period 1961 – 1985 and a verification set spanning the period 1986 – 1999. Since observational data will not be available to verify the downscaled rainfall for the future climatic periods, it seems reasonable to rather use diagnostic criteria to evaluate the performance of the statistical downscaling model. Such criteria include the following climatic indices:

- Maximum daily and total monthly rainfall;
- Variance;

- Percentage of wet days;
- Mean and maximum dry/wet spell length;
- Maximum 3-day precipitation total;
- Number of days with more than 25 mm of rain; and
- 90^{th} percentile of rain day amount.

These climatic indices are very useful in a number of economic sectors as they may help prepare that sector for adaptation to climate change. In the case of agriculture, total monthly rainfall and variance aid in determining crop or cultivar suitability. Maximum daily rainfall, the number of days with more than 25 mm of rain and the maximum 3-day precipitation totals are not only useful in evaluating groundwater replenishment but also help indicate potentially damaging heavy rainfall events. Of particular importance for agriculture is the wet and dry spell durations. Crops will be at risk from dry spells occurring during the growing season. The risk can sometimes be determined by assessing the probability that a long dry spell occurs when the plant is particularly sensitive, such as just after germination or at flowering (Stern & Coe, 1984).

In their respective studies Wilby and Wigley (2000), Wilby *et al*. (2002), Maini *et al.* (2004), Lines *et al*. (2005) and Hessami *et al.* (2008) made use of some subset of these climatic indices for precipitation forecasts. In addition to these, Hessami *et al.* (2008) also computed the mean and the standard deviation of observed and simulated values for total monthly precipitation. Additional verification measures will include the quantile-quantile (Q-Q) plot of the generated ensembles against the observed rainfall, the coefficient of determination (R^2), standard error (SE) and extreme value analyses.

The Q-Q plot is a scatterplot that is used to compare the dimensional values of the observed data with that of the modelled data (Wilks, 1995; Wilby & Dawson, 2007). The procedure works by constructing cumulative distribution functions (CDFs) of these data sets and plotting the percentiles of the observed and the predicted DJF rainfall against one another on a scatter chart with observed data on the vertical axis and modelled data on the horizontal axis (Wilks, 1995; Wilby & Dawson, 2007). A Q-Q plot for a

regression equation that modelled the predictand well would have the ensembles straddling the 45º diagonal line.

R^2 can be computed from the following equation:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

where SSR is the regression sum of squares, SST is the total sum of squared deviations of the predicted values around their mean, and SSE is the sum of squared differences between the residuals/errors and their means (Wilks, 1995). Qualitatively, R^2 can be interpreted as that portion of the variation of the predictand (proportional to SST) that is "described" or "accounted for" by the regression (SSR) (Wilks, 1995; Mendenhall & Sincich, 2003). For a perfect regression, SSR = SST and SSE = 0, so $R^2 = 1$. For a completely useless regression, SSR = 0 and SSE = SST, so that $R^2 = 0$. In such a case the least-squares regression line is almost indistinguishable from the sample mean of the predictand, so SSR is very small, and little of the variation in the forecast predictand can be ascribed to the regression (Wilks, 1995).

SDSM also allows the user to fit distributions to observed and downscaled data for the selected season in order to interpret the return period of extreme events (Wilby & Dawson, 2007). Following Wotling *et al*. (2000), the decision was made to similarly fit a Gumbel distribution (using the annual maximum series after the method of Shaw (1994, cited in Wilby & Dawson, 2007)) to the rainfall series. The results are plotted up to a return period of 100 years.

5.8 Generating Downscaling Scenarios of the Future Climate

The ultimate regression equations that are to be used for downscaling future climate projections were fit using all the observational data within the base period 1961 – 1990. The "Scenario Generator" operation in SDSM was used to produce 20 ensembles of synthetic daily weather series given the final set of daily atmospheric predictor variables

supplied by a GCM. The GCM predictor variables must be normalised with respect to the base period and available for all variables used in model calibration (Wilby & Dawson, 2007). The procedure is identical in all respects to the ensemble generation described in Section 5.6, except that it was necessary to specify different conventions for model dates. This was required since HadCM3 uses fixed year lengths of 360 days. For each QC and respective GCM experiment, the operation was first done for the base period in order to facilitate an evaluation of how closely the model can simulate the present climate. Later the procedure was repeated for 3 tri-decadal periods centred on the 2020s (spanning 2011 – 2040), 2050s (spanning 2041 – 2070) and the 2080s (spanning 2070 – 2099) in order to facilitate an evaluation of climate projections under future GHG forcing. Similar 30-year time slices were considered by Wilby *et al*. (2002) and Lines *et al*. (2005).

5.9 Analysis of Climate Forecasts

As mentioned earlier, the first statistical analysis was performed for the base period where several climatic indices were calculated for the observed and ensemble mean of the downscaled predictand. This was necessary in order to validate the downscaling model against observed data and to inform assessment of the significance of climate changes projected by the statistical downscaling. If for example the climate change projection results should later indicate changes smaller than the downscaling model's standard error, the model sensitivity to future climate forcing is said to be less than the model accuracy. In such cases the projected changes may result from model parameters rather than regional forcing (Wilby *et al*., 2002). This process was repeated for each GCM experiment and QC involved and the results were subsequently compared.

In a similar fashion, climatic indices were calculated for each tri-decade after which the downscaling results were compared to the downscaling results for the base period. Comparisons were facilitated by delta statistics which were calculated by taking the

absolute differences between the ensemble mean for each tri-decadal period and the ensemble mean for the base period. In this way the expected change in the various climatic indices mentioned in Section 5.7 could be assessed.

CHAPTER 6 RESULTS AND DISCUSSION

6.1 Rainfall Predictors

The screening process was performed on each of the five QCs using an independent observational dataset for the period 1961 – 1985. The totals for each of the 26 predictors and their lags are displayed in Appendix A. The columns on the right also indicate the rainfall predictors identified by other studies. Since only 8 predictors could be included in the final set (a SDSM restriction), the decision was made to select predictors based on both the unlagged and lagged totals. Table 6.1 summarizes the performance of the top predictors across the 5 QCs. This final set of top-scoring predictors thus form a generic set for the Upper Olifants River catchment. The selection process was complicated by the low explanatory power of the individual predictor variables. Similar low values of explained variance were also reported by other researchers who attempted to downscale daily rainfall e.g. Wilby *et al*. (2004) and Lines *et al*. (2005).

During the screening process it was established that the 1-day lags of the surface divergence and surface specific humidity fared much better than their unlagged counterparts (see Appendix A), hence the relatively low totals for these predictors in the second last column of Table 6.1 and significantly higher values in the last column of Table 6.1. The physical significance of these 1-day lags may lie in the convective nature of the rainfall over the study area. After a dry-spell, it is often the case that a continuous influx of moist air is required before significant convective development can be sustained. The 1-day lag in surface specific humidity may therefore act as an indicator of preceding moisture at the low-levels, while the divergence may act as an indicator of an evolving convective environment. Wilby and Wigley (2000) pointed out some of the problems encountered with the use of screen temperature as a predictor for daily rainfall. On the one hand high summer temperatures may be a consequence rather than a cause of dry conditions, resulting from clear skies and low rainfall. On the other hand

high summer temperatures may enhance the instability in the lower troposphere and trigger convective development. For this reason the mean screen temperature was omitted in the final predictor set.

Table 6.1: Top predictors and their screening results for the grid box centred on 25ºS 30ºE

Predictor	Symbol in Appendix A	B32D	B20A	B11K	B12D	B41A	Total	Total (incl. lag) [a]
Mean temperature at 2m [b]	temp	2	3	1	2	1	9	14
Surface airflow strength	p__f	3	3	1	2	1	10	15
Surface vorticity	p__z	1	2	1	1	3	8	15
Surface divergence	p_zh	1	1	1	1	1	5	15
Surface specific humidity	shum	1	1	1	1	1	5	15
Wind direction at 850 hPa	p8th	3	3	1	3	2	12	17
Relative humidity at 850 hPa	r850	1	3	3	3	2	12	17
Relative humidity at 500 hPa	r500	3	3	3	3	3	15	20
Meridional wind velocity at 500 hPa	p5_v	3	2	3	3	2	13	18

[a] includes 1-day lag values [b] omitted from final predictor set

The partial correlations between the selected predictors and the observed daily summer rainfall are presented in Table 6.2, while the corresponding P-values are contained in Table 6.3. Note that these values do not include those for the 1-day lags which will be used in an autocorrelation setup in SDSM, but the accumulated value for the selected predictor and its 1-day lag. During the calibration period it is already evident that some predictors show considerable variation in their correlations even across such a relatively small study area. Again, this may be due to the convective nature of the rainfall over the study area. Although all the partial correlations are fairly low, we see that the surface airflow strength exhibits a higher correlation value in the north and west (B32D – Groblersdal and B20A – Delmas) than in the east (B41A – Belfast). Some of the spatial differences seem artificial. An example is the difference in the partial correlations for the 500 hPa meridional wind velocity between Witbank (B11K) and Middelburg (B12D)

which are only about 30 km apart. Although the differences in the terrain morphology between these two QCs have been highlighted in Section 3.1, there does not seem to be any physical reason why that should influence the strength of the association between the daily rainfall and the upper-air wind components within such short a distance.

Table 6.2: Partial correlations with observed daily rainfall for the period 1961 – 1985 (values for 1-day lags are not included)

Predictor	B32D	B20A	B11K	B12D	B41A
Surface airflow strength	0.115	0.076	0.047	0.036	0.006
Surface vorticity	0.058	0.025	0.113	0.013	0.103
Surface divergence	0.067	-0.050	-0.079	0.025	0.003
Surface specific humidity	0.099	0.028	0.015	0.008	-0.019
Wind direction at 850 hPa	0.059	0.077	-0.011	0.043	0.027
Relative humidity at 850 hPa	0.049	0.024	0.027	0.106	-0.036
Relative humidity at 500 hPa	0.072	0.049	0.075	0.117	0.126
Meridional wind velocity at 500 hPa	-0.080	-0.091	-0.075	-0.018	0.062

Table 6.3: P-values corresponding to the partial correlations in Table 6.2 (values for 1-day lags are not included)

Predictor	B32D	B20A	B11K	B12D	B41A
Surface airflow strength	0.009	0.074	0.249	0.377	0.558
Surface vorticity	0.197	0.456	0.005	0.534	0.021
Surface divergence	0.141	0.235	0.057	0.462	0.563
Surface specific humidity	0.026	0.430	0.522	0.552	0.506
Wind direction at 850 hPa	0.193	0.072	0.542	0.309	0.452
Relative humidity at 850 hPa	0.270	0.458	0.429	0.016	0.380
Relative humidity at 500 hPa	0.110	0.244	0.070	0.007	0.004
Meridional wind velocity at 500 hPa	0.078	0.031	0.070	0.508	0.174

Since a higher P-value indicates a higher likelihood that the corresponding partial correlation may be due to chance, it is interesting to note that the P-values for the 500 hPa relative humidity and 500 hPa meridional wind are relatively small. This result may stem from the higher expected correlations between widespread rainfall producing systems such as westerly and tropical-temperate troughs and the observed rainfall

across the QCs. It is perhaps also worth commenting here that Groblersdal (B32D) had no P-values larger than 0.3 while some of the other QCs such as Middelburg (B12D) had 6 P-values larger than 0.3.

In Section 3.3 it was highlighted that the bulk of the summer rainfall over the study area is in the form of thundershowers, with some orographic rain along the higher lying areas to the east. It is general knowledge amongst weather forecasters that the three conditions required for thunderstorm development are a continuous inflow of low-level moisture, a deep unstable layer that stretches at least 10 000 feet above the surface and a trigger action which will initiate the convection and overcome any initial negative buoyancy. With the warm Indian Ocean constituting a rich source of maritime air to the east of the subcontinent, the advection of low-level moisture from the north and east should be a governing factor in rainfall occurrence (see Figure 3.5). This influence is thought to be represented by predictors such as the surface specific humidity, relative humidity at 850 hPa and the wind direction at 850 hPa. It is also conceivable that the surface airflow strength influences not only the magnitude of any possible low-level confluence (or diffluence), but also the rate at which moist (or dry) air may be advected into the area. Surface convergence will supply the necessary trigger action required for the deep convection that occurs in thunderstorms, while vorticity is an indication of the horizontal rotation of fluid particles around the local vertical (Holton, 1992) and finds its relevance to weather forecasting through the divergence-vorticity relation.

The 500 hPa relative humidity may compensate for any mid to upper-air dryness that may enhance cloud evaporation (thereby reducing the chances of rainfall) and reduce the latent instability of the air mass. Apart from occasional air mass thunderstorms, the majority of the rainfall events over the study area can be traced back to either westerly waves or a combination of westerly and easterly waves (generally manifested in the form of tropical-temperate troughs). In either case we find that southerly-component flows are dry (upstream of the upper-air trough) while northerly component flows are relatively moist. In the case of westerly waves, it is well known that the area downstream (east) of the trough-line is characterised by cyclonic vorticity advection,

upper-air divergence, vertical uplift and thus enhanced thunderstorm development (Holton, 1992). The 500 hPa meridional wind speed therefore has a direct bearing on the nature of rainfall producing systems in this part of the world.

The calibration results, as described by the explained variance (R^2), standard error (SE) and Chow test statistic, are displayed in Table 6.4. The model results are not very encouraging, yet explained variances are in line with those reported by Lines *et al*. (2005). For the 25-year calibration period (spanning 1961 – 1985) the downscaling model could only account for approximately 7% of the variation in the daily DJF rainfall at Groblersdal (B32D). The explained variance dropped to a meagre 3% at Belfast (B41A) while standard errors averaged 0.45 across the 5 QCs. Results obtained by substituting the generic set of predictors, that is thought to be applicable across all 5 QCs, with those obtained for each QC, only showed slight improvements in the respective statistics for some QCs (e.g. an increase in R^2 in the order of 0.02) while almost any other combination of predictors, including those obtained from the literature, indeed proved to be inferior. This at least suggests that the screening procedure was not flawed, while a degree of collinearity amongst the candidate predictors may also account for the robust results.

Table 6.4: SDSM calibration results for the period 1961 – 1985

Statistic	B32D	B20A	B11K	B12D	B41A
R^2	0.066	0.036	0.056	0.054	0.029
SE	0.476	0.449	0.475	0.449	0.418
Chow	2.182	6.298	1.494	1.822	2.110

6.2 Cross-Validation Results

Validation of the downscaling model was performed with the use of observational data over the period 1986 to 1999. Table 6.5 presents the verification results, as described by the R^2, SE and Chow test statistics. As can be expected, the model performed worse over the verification period. For the 15-year verification period (spanning 1986 – 1999)

the downscaling model could only account for approximately 6% of the variation in the daily DJF rainfall at Groblersdal (B32D). The explained variance dropped to a meagre 2% at Delmas (B20A) while standard errors still averaged 0.45 across the 5 QCs. The Chow test value of 6.079 also indicates that the model may not be stable for Delmas (B20A). What is encouraging is that the explained variances are in line with those reported for the calibration period in Table 6.4, indicating at least that the model was not an overfit.

Table 6.5: SDSM verification results for the period 1986 – 1999

Statistic	B32D	B20A	B11K	B12D	B41A
R^2	0.057	0.024	0.049	0.043	0.027
SE	0.463	0.447	0.479	0.451	0.411
Chow	1.103	6.079	2.332	0.934	2.313

6.2.1 Quantile-quantile plots

After using the stochastic weather generator to create equally plausible ensembles of the downscaled rainfall over the verification period, the ensembles were subjected to a frequency analyses. Figure 6.1 presents the Q-Q plots for each QC and for all 20 ensemble members after adjusting the variance inflation to a value proposed by Wilby and Dawson (2007). SDSM performs variance inflation by adding or reducing the amount of "white noise" applied to model estimates (Wilby and Dawson, 2007). In the cases of Groblersdal (B32D) and Middelburg (B12D) it is evident that SDSM underforecasts low rainfall amounts while high daily rainfall amounts are overforecast. The Q-Q plots for Delmas (B20A), Witbank (B11K) and Belfast (B41A) show clear signs of model bias, with almost all daily rainfall amounts being overestimated. Apparently none of the Q-Q plots depict a regression equation that modelled the predictand well which would have been evident by ensembles straddling the 45º diagonal line. When it comes to climate change applications, these Q-Q plots are sometimes of more relevance than the daily verification statistics. The biases which were pointed out for Delmas, Witbank and Belfast imply that the downscaling model will tend to predict increased precipitation for any future period.

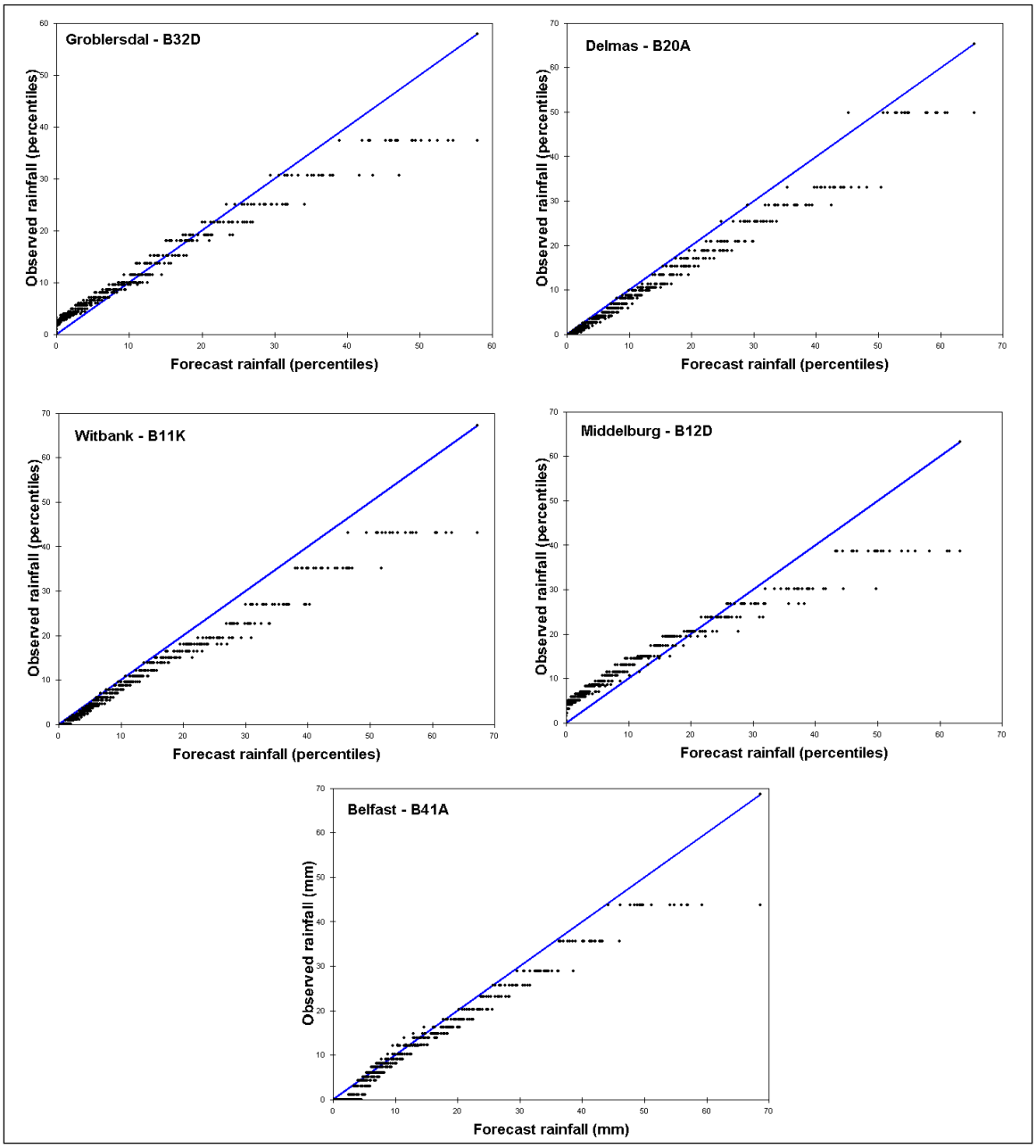

Figure 6.1: Q-Q plots of modelled versus observed rainfall percentiles for the verification period 1986 – 1999 (each point represents one of the 20 ensemble members)

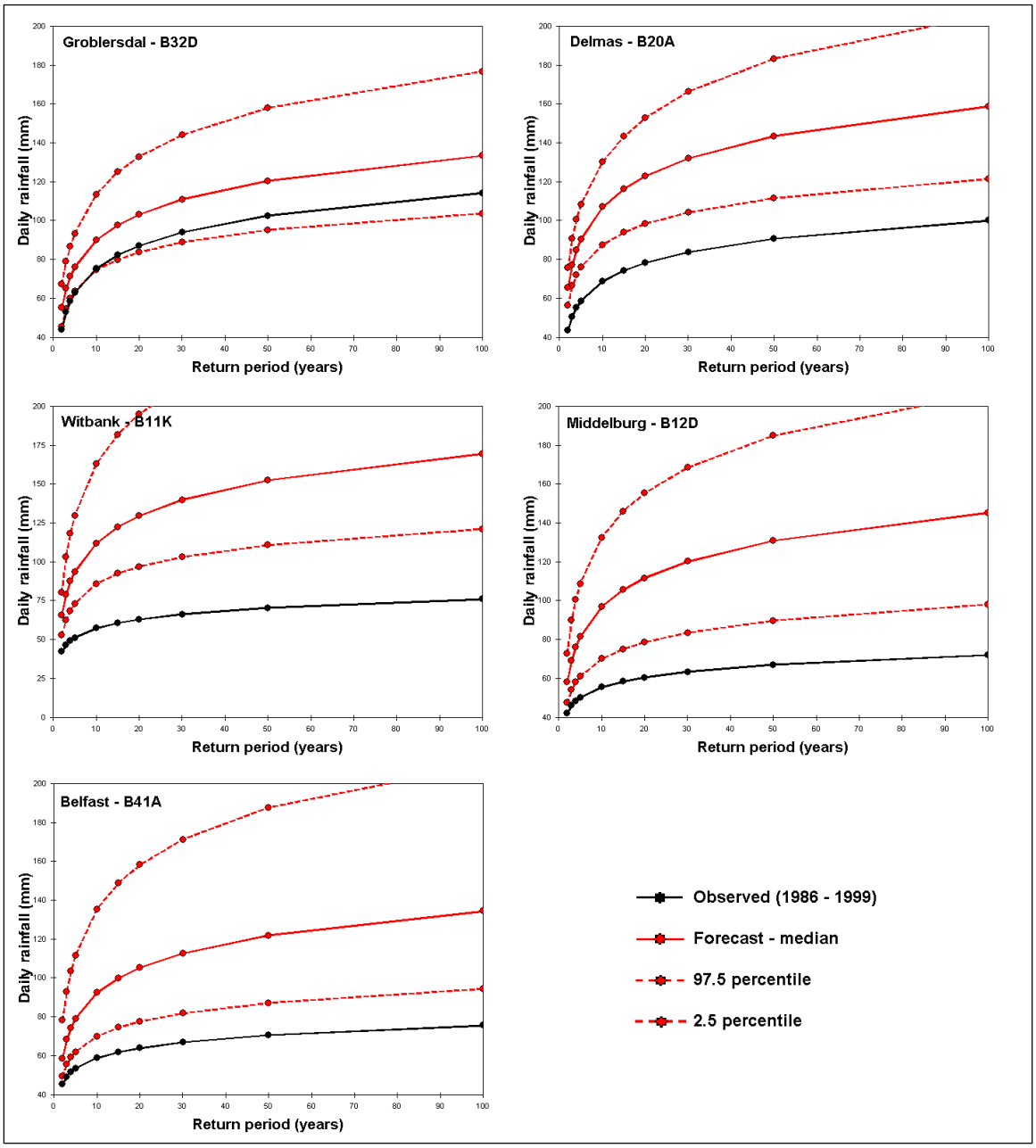

Figure 6.2: Extreme value analyses for the summer rainfall during the period 1986 – 1999 using a Gumbel fit (the black line represents the observed rainfall while the red lines depict the 97.5th, 50th and 2.5th percentiles of the forecast ensembles)

6.2.2 Extreme value analyses

The time series of observed and forecast daily rainfall were fit to a Gumbel distribution and analysed in terms of the return period of rainfall occurrences. The results of the extreme value analyses are shown in Figure 6.2. The respective return periods are

considerably underestimated over all 5 QCs. As an example we can see that the observed return period for a daily rainfall of 100 mm is approximately 50 years in Groblersdal (B32D) while the modelled return period for the same value is approximately 20 years (as measured by the median). SDSM only managed to generate ensembles that straddle the observed return period in Groblersdal (B32D) while the forecast return periods over the remainder of the QCs are unacceptably low.

In light of the low explained variances (Table 6.5) and observed biases (Figure 6.1) before, this must mean that the SDSM was unable to successfully downscale the daily rainfall over 4 of the QCs given the predictors that were available to this study. For this reason, the results and discussion pertaining to the projected climate change for these QCs will not be included and the remainder of the chapter will only focus on the results for Groblersdal (B32D).

6.2.3 Analysis of climatic indices

Figures 6.3a-c compares the downscaled and the observed values of selected climatic indices for the calibration period. These indices form part of the widely used STARDEX indices (STARDEX, 2008). In the case of Groblersdal it can be seen that the model performed fairly well in most instances. Perhaps the biggest problems were associated with an over-inflated variance, especially in the February daily rainfall (a difference of 187 mm in Figure 6.3a), over-predicting the number of significant rainfall events (here defined as days with 25 mm or more in Figure 6.3b) in February with approximately 6 days and under-predicting the percentage of wet days in December (as indicated in Figure 6.3b). For the purpose of this study wet days are defined as days on which any rainfall amount (more than 0 mm) was reported.

The mean and maximum dry- and wet-spell lengths (Figure 6.3c) compared reasonably well. The downscaling model over-forecasted the maximum dry-spell length for December with 5 days while it under-forecasted it with 5 days during February. The maximum wet-spell duration was also under-forecast with about 3 days for February.

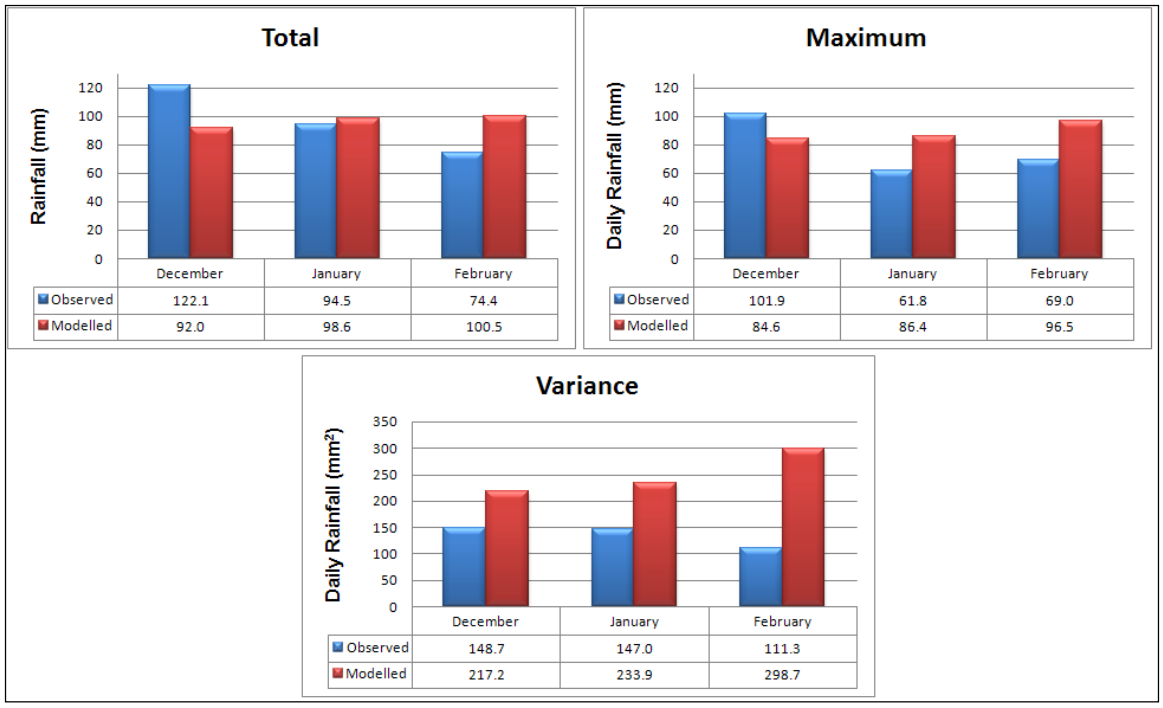

Figure 6.3a: Comparison between observed and forecast climatic indices for the verification period (1986 – 1999) for Groblersdal

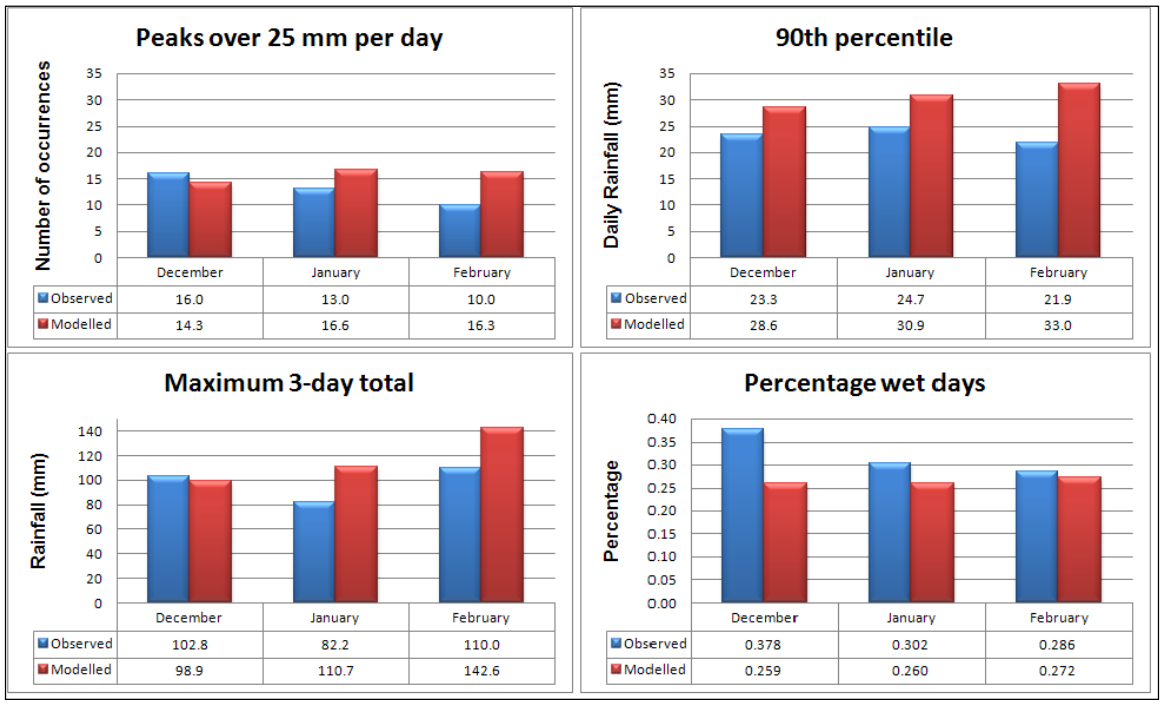

Figure 6.3b: Comparison between observed and forecast climatic indices for the verification period (1986 – 1999) for Groblersdal

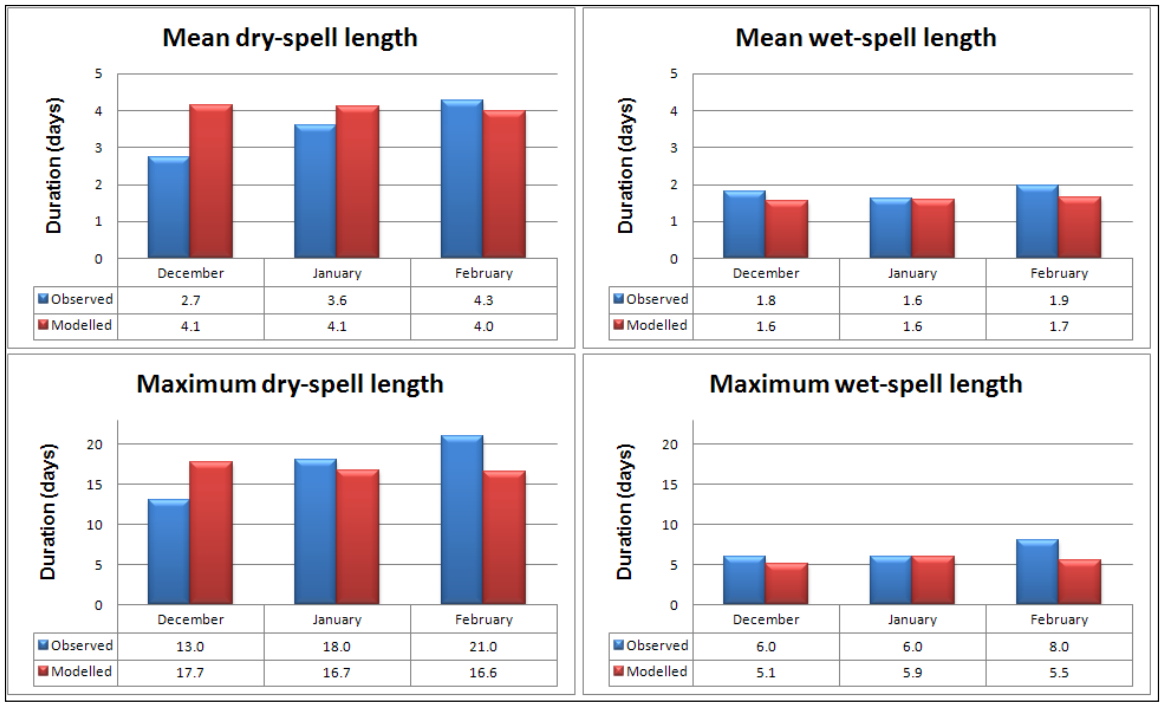

Figure 6.3c: Comparison between observed and forecast climatic indices for the verification period (1986 – 1999) for Groblersdal

6.3 Description of the Downscaled Current Climate

In order to facilitate a comparison between the projected future climates under the different SRES scenarios, a description of the downscaled current climate is necessary. This description is based on the same STARDEX climatic indices that were discussed in the model verification (Figures 6.3a-c). For the current climatic period (also referred to as the base period) the downscaled total rainfall for the summer months vary within a narrow band centred around 80 mm, while the variance itself ranges from 218.8 mm in February to 260.7 mm in December (Figure 6.4a). The corresponding observed values for the calibration period (shown in Figure 6.3a) vary from 9.2 to 10.4 mm. The modelled maximum daily rainfall values vary from 99.1 mm in January to 109.5 mm in December with standard deviations of 32 and 40 mm respectively.

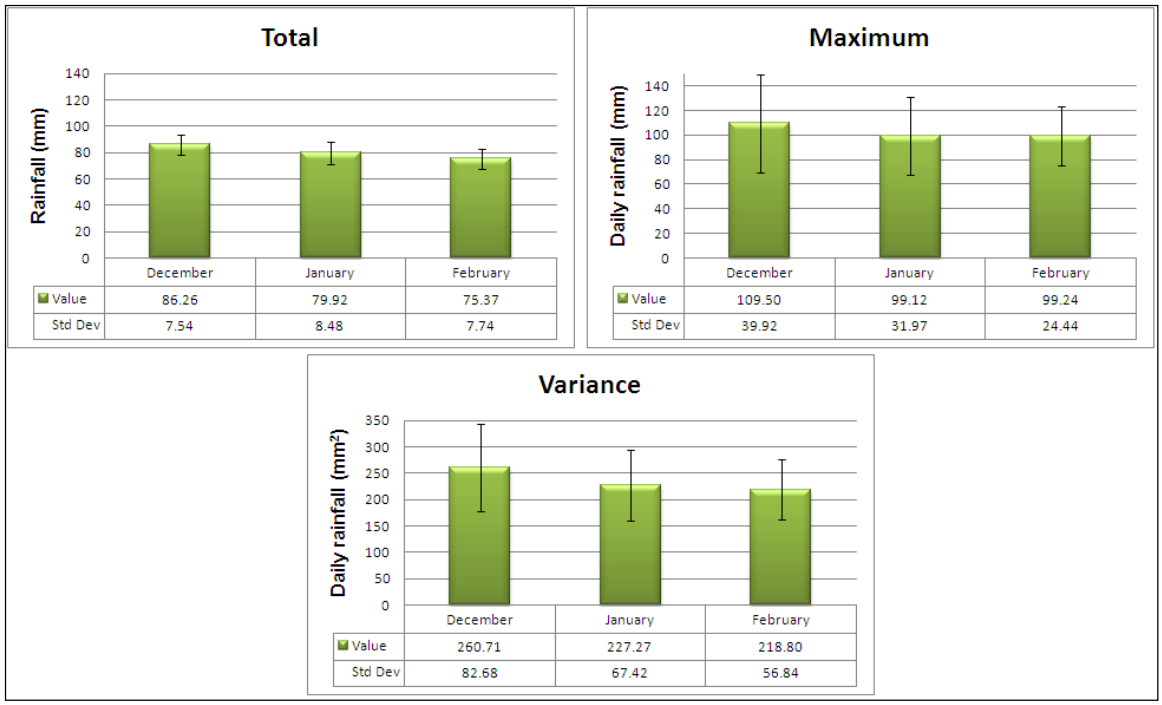

Figure 6.4a: Climatic index values and standard deviations for Groblersdal for the modelled base period (1961 – 1990)

As depicted in Figure 6.4b, the number of days with 25 mm of rain or more drops from an average value of 30.8 in December to 25.1 in February, with a comparable decreases in the values of the 90th percentile (from 30.6 mm in December to 28.1 mm in February), maximum 3-day rainfall total (122 mm in December to 112 mm in February) and percentage wet days (0.24% in December to 0.22% in February). All of these index values for DJF are higher than those observed in the real world (as presented in Figure 6.3b). It is again interesting to note that compared to the values obtained for the calibration period, the downscaling model tend to over-forecast all the index values provided in Figure 6.4b. Figure 6.4c indicates that the mean dry-spell duration for the three summer months is in the order of 4.5 days, while the wet-spell duration is only about 1.5 days. The maximum dry-spell duration ranges between 20.6 and 23.1 days. The modelled maximum wet-spell duration is approximately 5 days while the observed value for calibration period was between 6 and 8 days (Figure 6.3c). Except in the case of the variation and peaks over 25 mm, the climatic index values fall within the range of the observed climatic index values as depicted in Figures 6.3a-c.

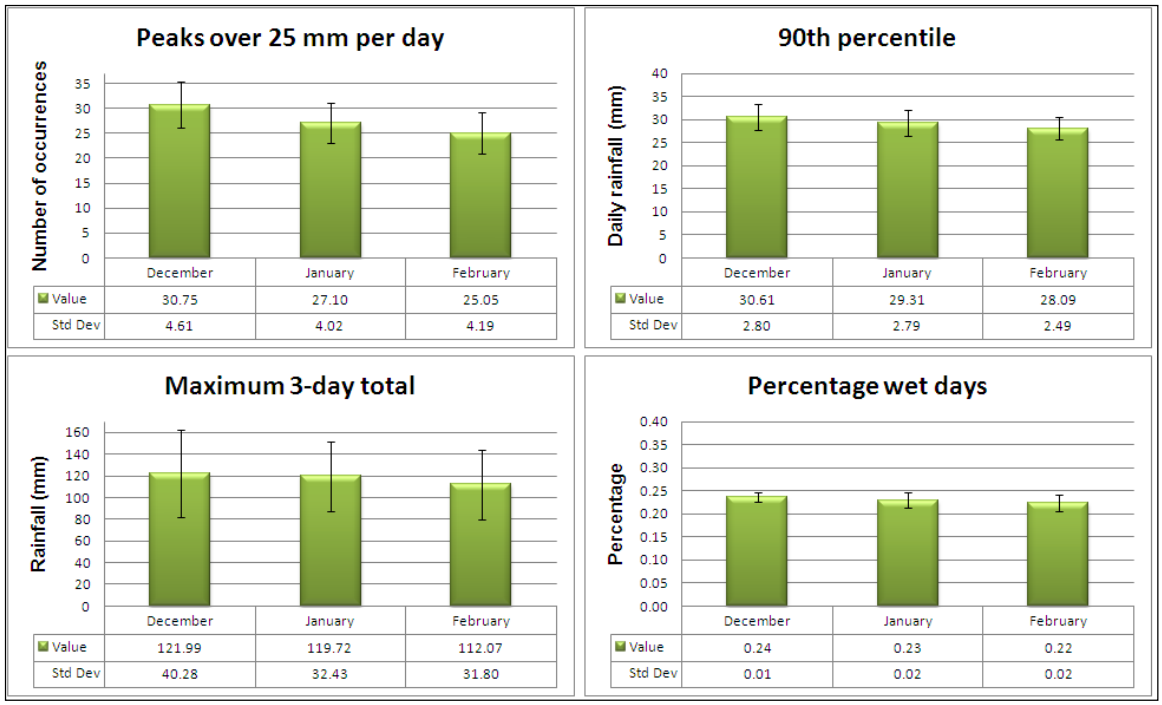

Figure 6.4b: Climatic index values and standard deviations for Groblersdal for the modelled base period (1961 – 1990)

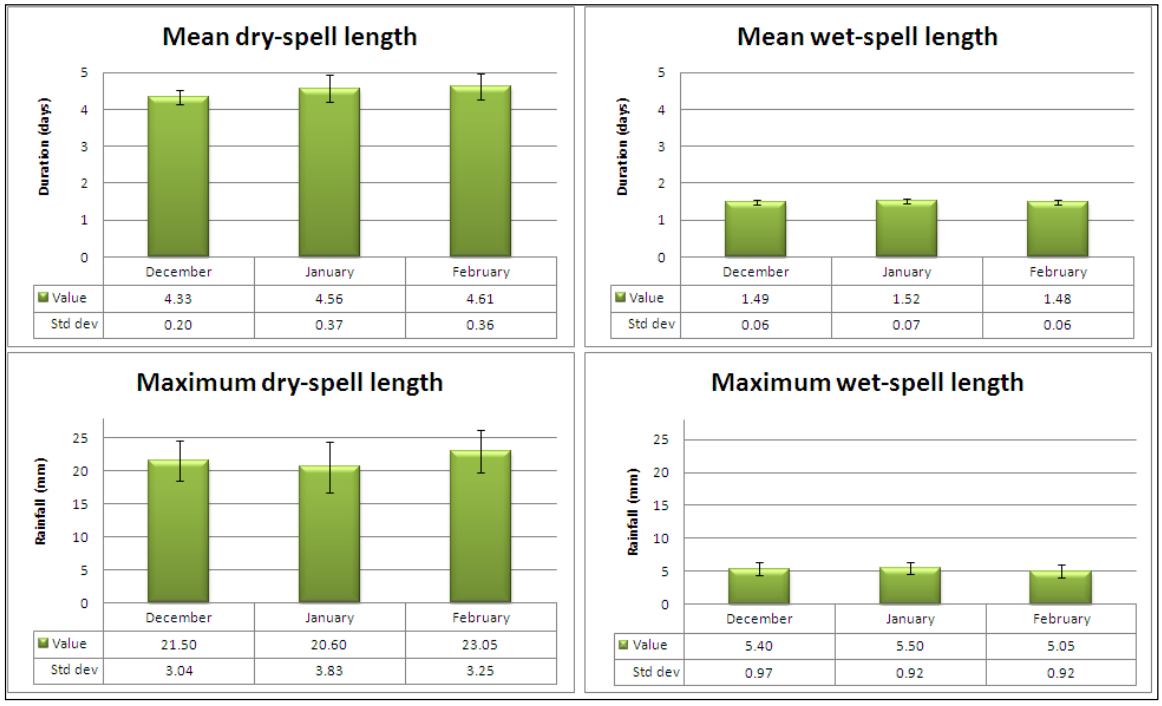

Figure 6.4c: Climatic index values and standard deviations for Groblersdal for the modelled base period (1961 – 1990)

6.4 Description of the Downscaled Future Climates

6.4.1 Downscaling of the A2 scenario

The description of the future climatic periods, as projected by the A2 scenario, focuses on the so-called delta statistics of selected climatic indices (Figures 6.5a-c) and the expected return period for extreme values (Figure 6.6). In each case the diagnostic indices are derived for the summer season using the mean of all 20 generated ensemble members. Figures 6.5a-c contain bar charts depicting the absolute difference between the ensemble mean of each of the tri-decadal future periods and that of the base period (1961 – 1990).

Figure 6.5a indicates the projected changes with regards to the mean, maximum and total rainfall along with the expected change in variance. According to the A2 scenario it seems that the mean daily rainfall will increase in January and February (0.8 and 1.1 mm higher for the 2070-2099 period) while inconsistent changes are forecast across the three future periods for December. Similarly, no clear linear trend emerges for the other climatic indices (viz. maximum, total and variance). This is somewhat surprising as one would initially expect changes indicated for the first future period (2011 – 2040) to continue and perhaps grow in the successive periods (2040 – 2070 and 2070 – 2099). At first, one plausible explanation for the inconsistent changes may lie in oscillatory variations in the summer rainfall over the north-eastern interior of South Africa as first identified by Tyson and Dyer (1975). According to Tyson (1986) the rainfall over this region exhibits an unambiguous 18-year cycle with an apparent oscillation between a 9-year dry period and a 9-year wet period. Although a full cycle is contained within a 30-year period, an extrapolation of Tyson's observed cycles into the future should result in generally dry conditions during the early 2020s and 2050s, while wet conditions should prevail during the latter half of the 2020s and 2080s. When examining Figure 6.5a such an extrapolation seems to be consistent with at least the projected changes for the maximum and total rainfall for the three tri-decadal periods centred on the 2020s, 2050s and 2080s.

It must however be mentioned that the observed rainfall in the 1990s and early 2000s did not fit into the Tyson cycles described above. It therefore seems unreasonable to expect this weak amplitude cycle to dominate the projected climate changes. For example, a region may become wetter in future due to an increase in the occurrence of well-developed tropical-temperate troughs and dry out again as these systems weaken and their average positions shift further west. Changes in the atmospheric circulation that may result in such a shift in the average position of the tropical-temperate troughs have been noted by Engelbrecht *et al*. (2009). This implies that the inconsistent projected changes may not be problematic after all.

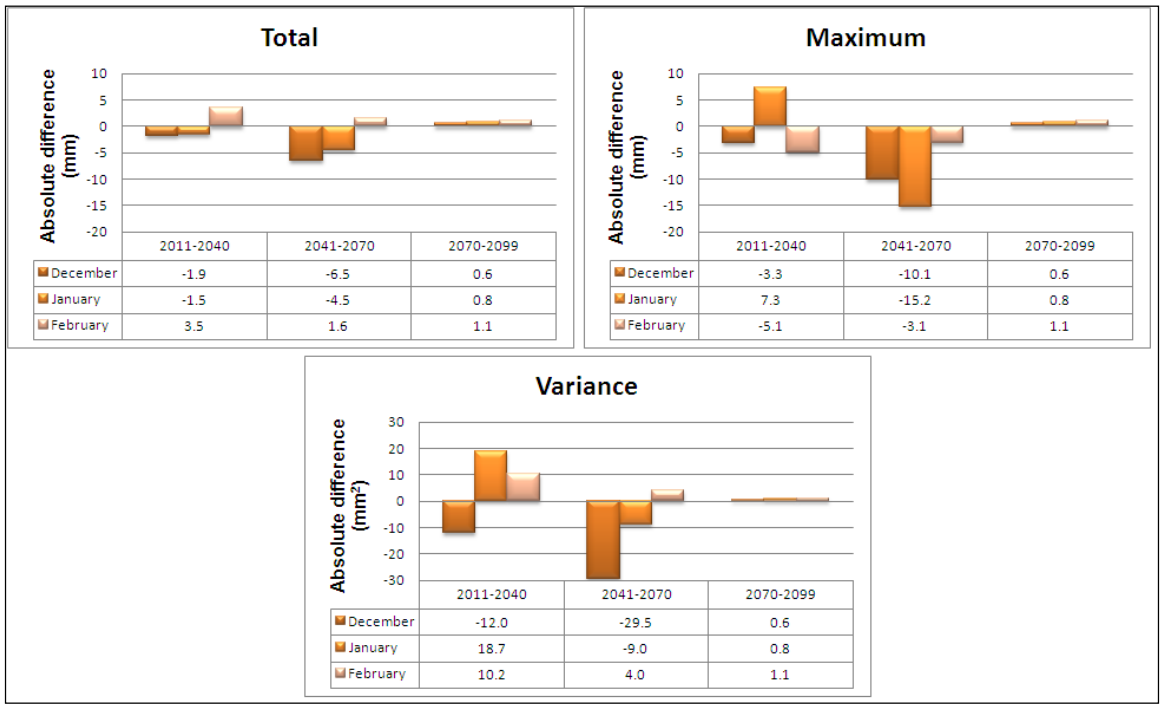

Figure 6.5a: Delta statistics (downscaled future period minus modelled base period) for selected climatic indices according to the A2 scenario (indicated changes are absolute differences)

The projected changes with regard to the number of occurrences greater than or equal to 25 mm per day and the rainfall amount corresponding to the 90th percentile are depicted in Figure 6.5b. The maximum rainfall total accumulated over 3 consecutive days and the percentage of days on which rainfall occurred are also presented. According to the A2 scenario it seems that consistent increases can be expected for the

month of February in terms of the number of occurrences exceeding 25 mm per day and the amount of rainfall corresponding to the 90th percentile. The 3-day rainfall totals are expected to change little and the percentage of wet days will decrease slightly. No major changes are expected for January in terms of these four climatic indices, while December exhibits fluctuating changes consistent with the rainfall changes described earlier. In light of the convective nature of the rainfall, a more likely reason for these results may lie in a stronger signal for the late summer as opposed to the early summer. For the early summer heat thunderstorms, with their irregular occurrence, are perhaps more dominant, while cloud band formation becomes more frequent in February as the CAB shifts further south (see Section 3.3.2).

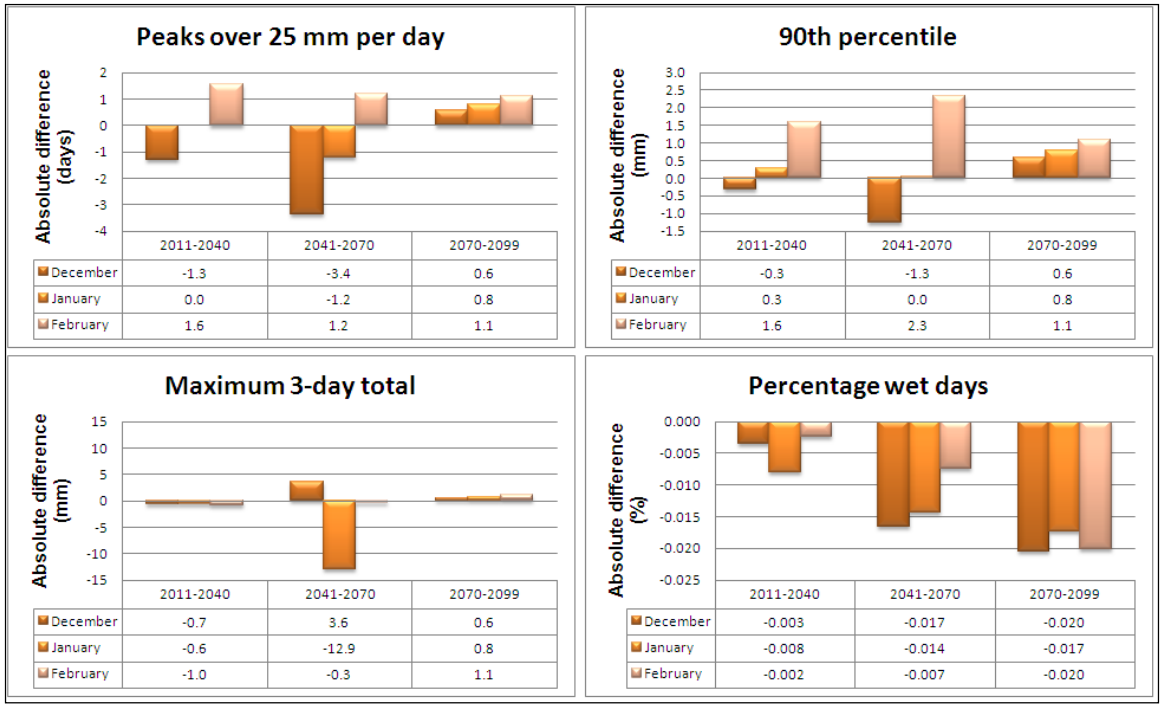

Figure 6.5b: Delta statistics (downscaled future period minus modelled base period) for selected climatic indices according to the A2 scenario (indicated changes are absolute differences)

According to the A2 scenario it seems that the average summer dry-spell lengths are expected to increase slightly (with less than one day) across all three tri-decadal periods (Figure 6.5c). The maximum dry-spell lengths are expected to increase slightly with 1 to 2 days by the 2080s during December and January, while a decrease is forecast for

February (albeit only with an average of 1.5 days). Mean wet-spell lengths are expected to decrease slightly during the summer months, while the maximum wet-spell duration shows decreases for January and increases for February, while the signal for December is somewhat inconsistent across the three tri-decadal periods.

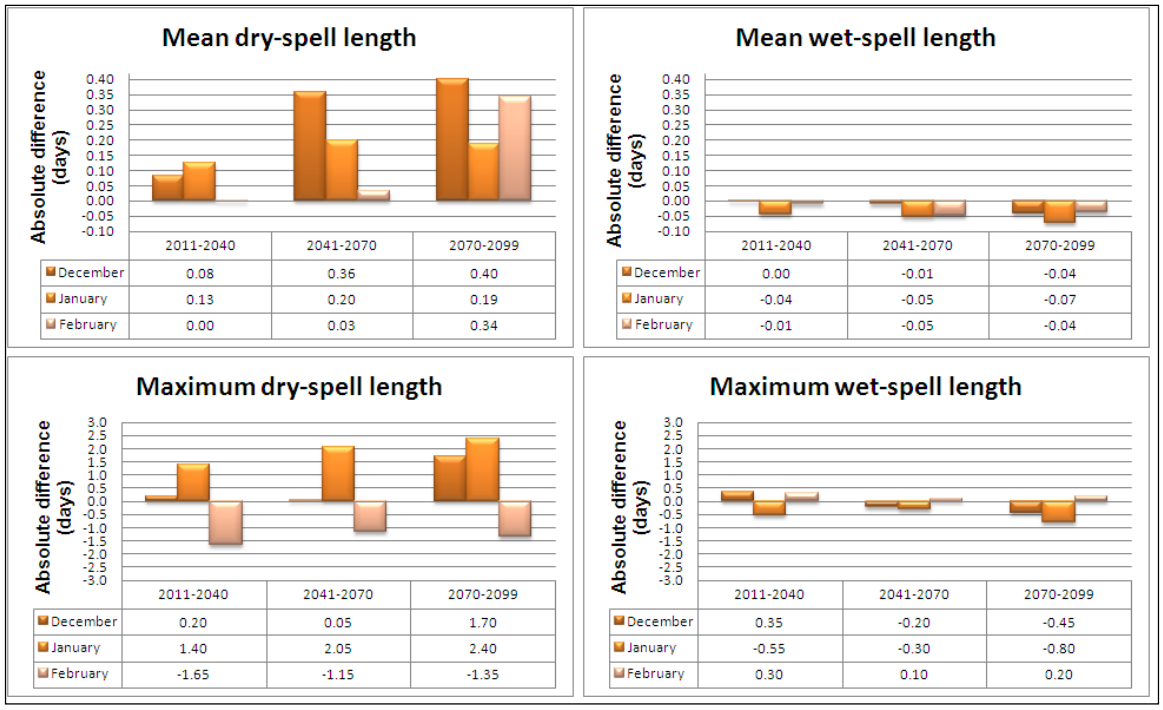

Figure 6.5c: Delta statistics (downscaled future period minus modelled base period) for selected climatic indices according to the A2 scenario (indicated changes are absolute differences)

The expected return period for extreme values under the A2 scenario, using a Gumbel fit, are depicted in Figure 6.6. It can be seen that the return period for a daily rainfall amount of 100 mm at Groblersdal increases to 30 years during the drier period centred on the 2050s and remains approximately 20 years throughout the remainder of the 21st century as judged against the median.

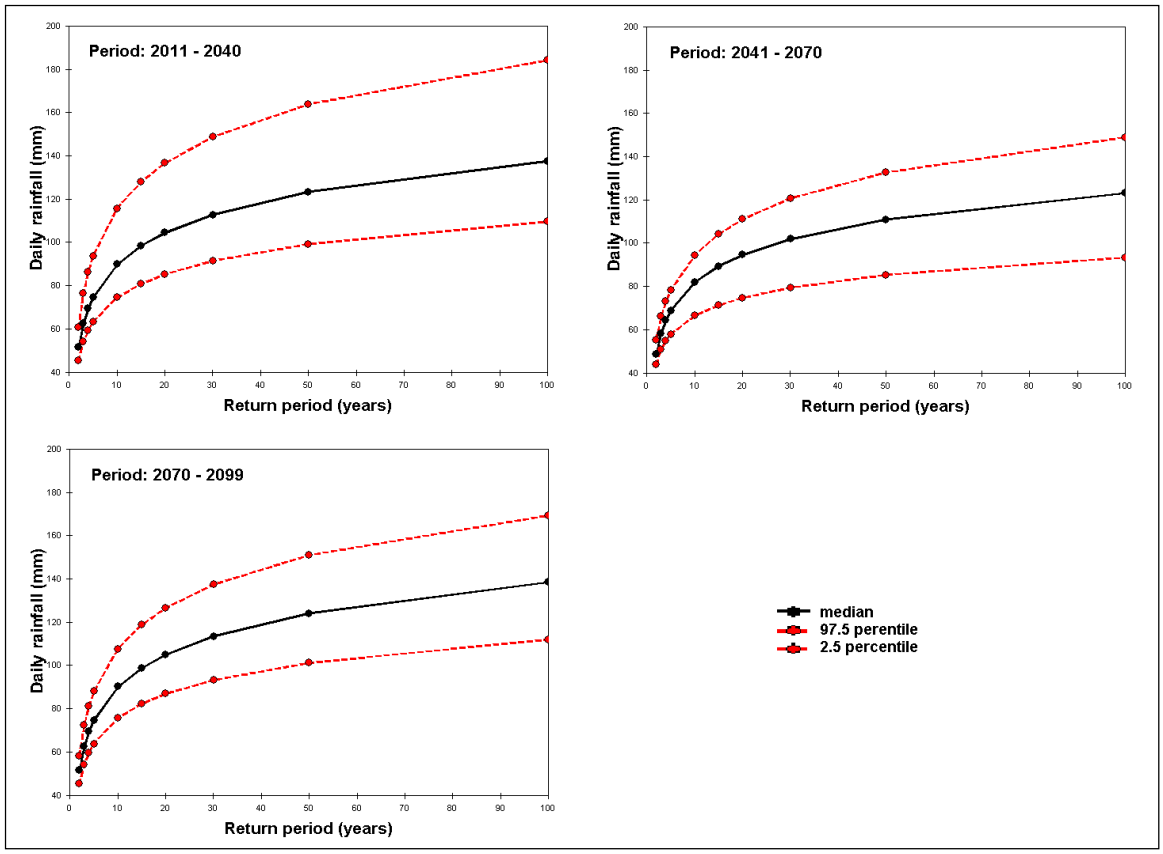

Figure 6.6: Extreme value analyses of daily rainfall for Groblersdal for the various downscaled tri-decadal periods according to the A2 scenario

6.4.2 Downscaling of the B2 scenario

Figure 6.7a indicates the projected changes with regard to the mean, maximum and total rainfall along with the expected change in variance. With regards to all four climatic indices the projected changes correspond closely to those predicted under the A2 scenario. The most noteworthy deviations from the A2 scenario come to light in the total rainfall and the variance for the period centred on the 2080s, where the total rainfall is now expected to decrease slightly during the first two months and the variance is set to increase rather significantly during February. Again, no clear linear trend emerges for the maximum rainfall and variance during the 21st century.

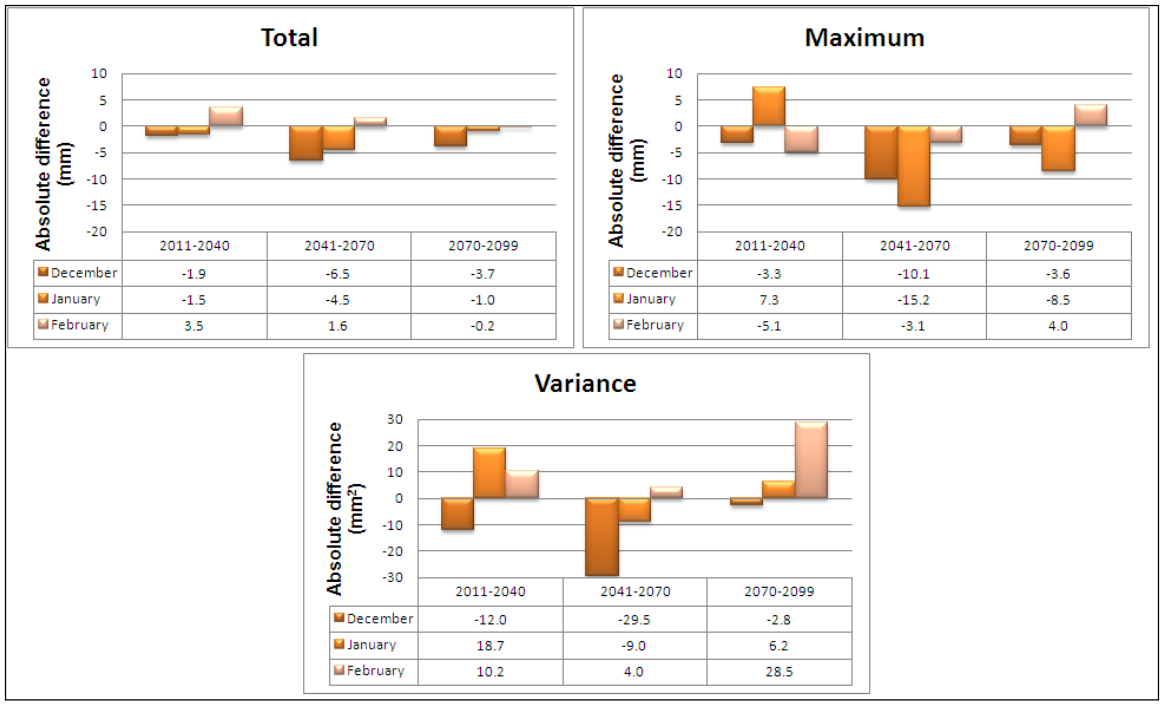

Figure 6.7a: Delta statistics (downscaled future period minus modelled base period) for selected climatic indices according to the B2 scenario (indicated changes are absolute differences)

The projected changes with regard to the number of occurrences greater than or equal to 25 mm per day, the rainfall amount corresponding to the 90th percentile, the maximum rainfall total accumulated over 3 consecutive days and the percentage of days on which rainfall occurred are depicted in Figure 6.7b. Again the differences in the expected GHG forcing under the B2 scenario does not seem to influence any of these climatic indices differently from the A2 scenario until the late 21st century (i.e. 2070 – 2099). According to the IPCC (2007) about half of the early 21st century warming is committed in the sense that it would occur even if atmospheric concentrations were held fixed at year 2000 levels. For the 2080s, the biggest deviations from the projections under the A2 scenario include a decrease in the number of rainfall days with 25 mm or more in December, a decrease in the maximum 3-day rainfall total for January and a contrasting increase in the same value for February.

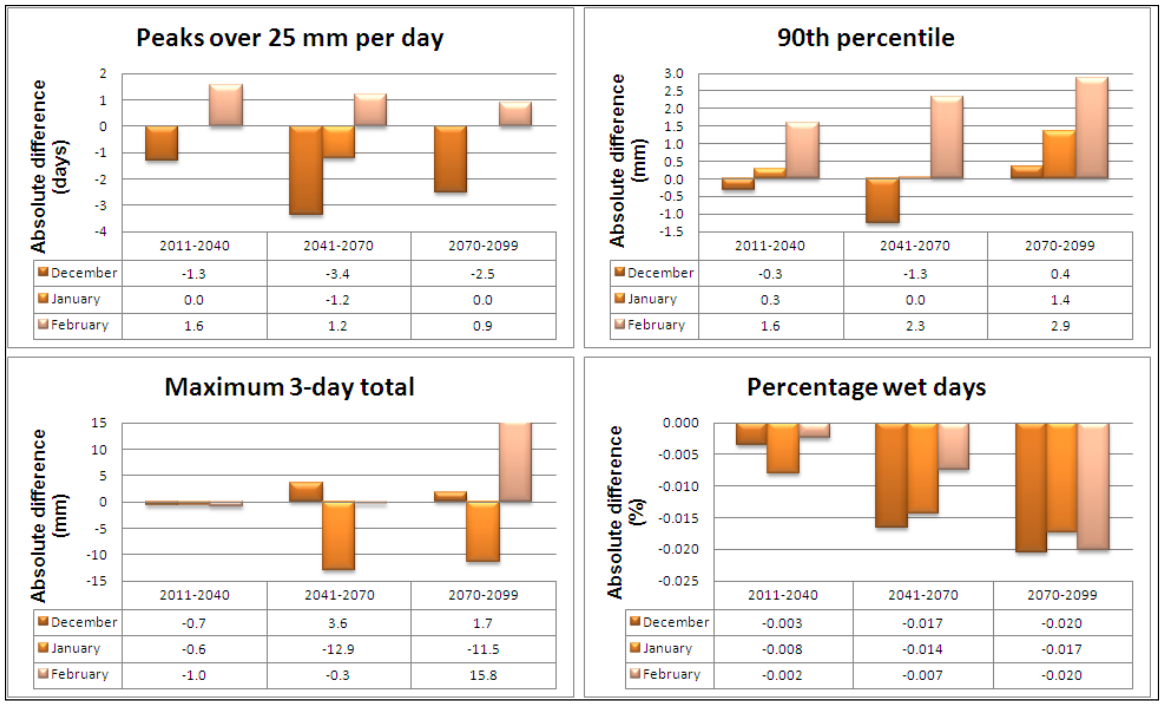

Figure 6.7b: Delta statistics (downscaled future period minus modelled base period) for selected climatic indices according to the B2 scenario (indicated changes are absolute differences)

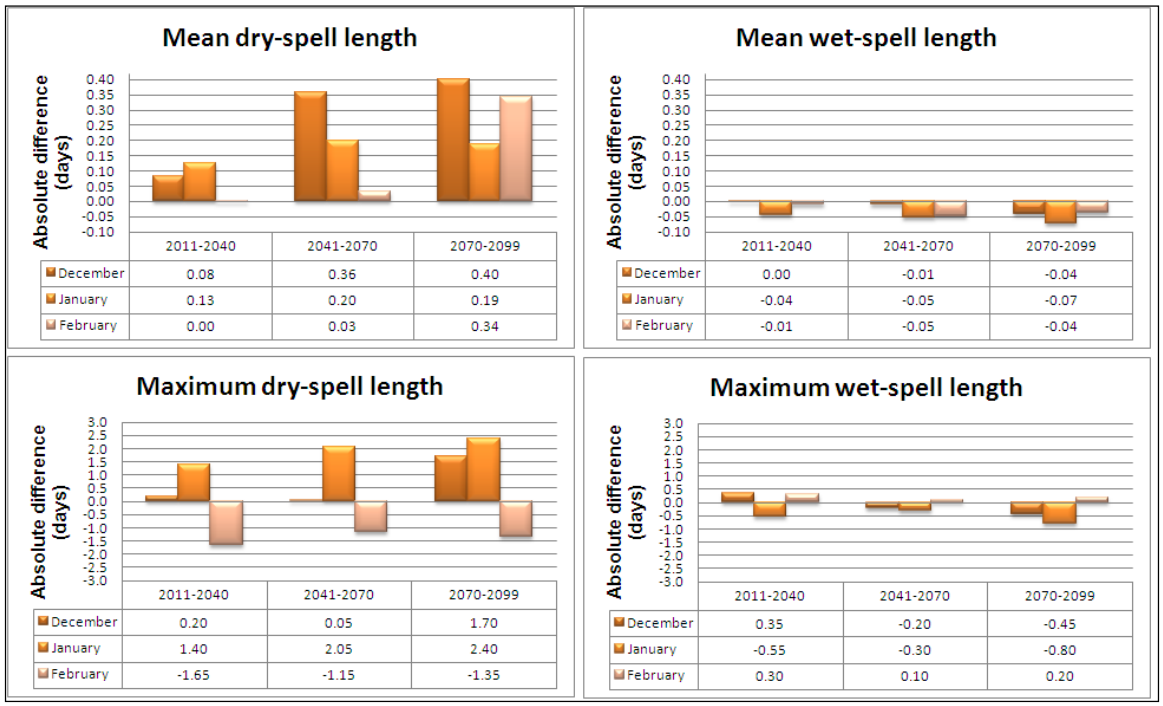

Figure 6.7c: Delta statistics (downscaled future period minus modelled base period) for selected climatic indices according to the B2 scenario (indicated changes are absolute differences)

The changes predicted under the B2 scenario for the average and longest dry- and wet-spell lengths, as illustrated in Figure 6.7c, correspond very closely to those predicted under the A2 scenario in Figure 6.5c. This implies that both the mean and maximum dry-spell lengths are expected to increase slightly during December and January, while the maximum dry-spell duration is expected to decrease during February. Similarly, the mean wet-spell lengths are expected to decrease slightly during the summer months, while the maximum wet-spell duration shows consistent decreases for January and consistent increases for February. This shows that for Groblersdal downscaled wet- and dry-spell length is not sensitive to different GHG forcings as prescribed by the A2 and B2 scenarios in the HadCM3 model.

The expected return period for extreme values under the B2 scenario, using a Gumbel fit, are depicted in Figure 6.8. It can be seen that the return period for a daily rainfall amount of 100 mm at Groblersdal increases to 25 years during the drier period centred on the 2050s and remains approximately 20 years throughout the remainder of the 21st century as judged against the median.

Although projected changes have been obtained from the downscaling model for the future climates, it should be noted that in most cases these projected changes are comparable or smaller than the errors that were made by the model when tested against the verification set (see Figures 6.3a-c). For example, both the A2 and B2 scenarios indicate that the total rainfall should decrease with 6.5 mm per month in December towards the 2050s, but during verification it was shown that for the current climatic period the model underestimated the total rainfall in December with 30.1 mm! This means that the downscaling model is simply not sensitive enough for these projected changes to be taken at face value and the results must only be used with caution.

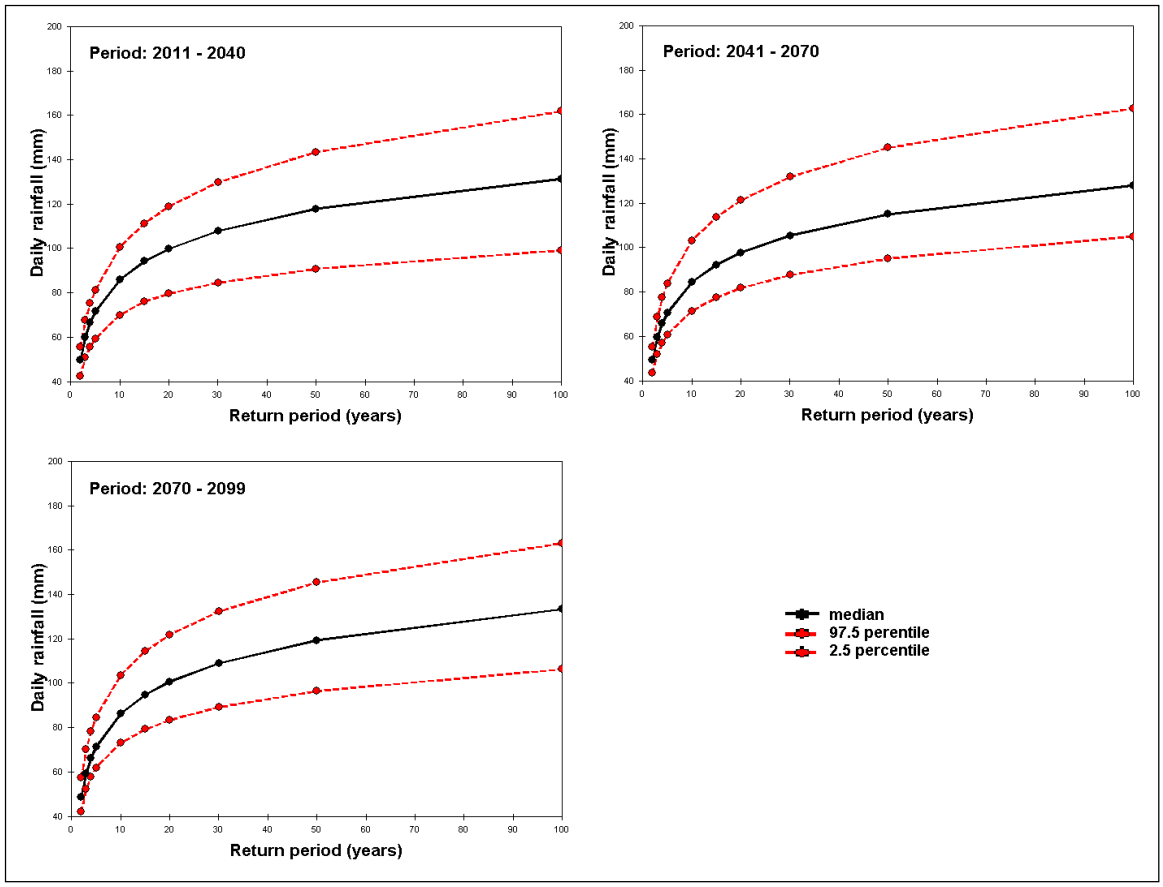

Figure 6.8: Extreme value analyses of daily rainfall for Groblersdal for the various downscaled tri-decadal periods according to the B2 scenario

6.5 Exploring the Effects of Expanding the Predictor Domain

Discouraged by the poor model performance, an attempt was made to explore the feasibility of employing predictors from a larger spatial domain surrounding the currently used grid-box. According to Wilby *et al*. (2002), the grid-box nearest to the target site does not always yield the strongest predictor-predictand relationship. Spatial correlations were calculated between the observed rainfall for Groblersdal (B32D) and the NCEP predictors identified in Table 6.2 for all eight of the GCM grid-boxes bordering the grid-box overlying the QC. The results of these spatial correlations are presented in Table 6.6 where the grid-box overlying the study area is situated at 25ºS 30ºE.

Table 6.6: Spatial correlations between Groblersdal observed rainfall and selected NCEP predictors at indicated GCM grid boxes for the period 1961-1985

Predictor	Coordinates	26,25ºE	30,00ºE	33,75ºE
Surface airflow strength	**22,50ºS**	0.022	0.117	-0.008
	25,00ºS	0.028	0.129	0.042
	27,50ºS	0.038	0.024	0.027
Surface vorticity	**22,50ºS**	-0.070	0.030	-0.013
	25,00ºS	0.062	0.084	-0.086
	27,50ºS	0.104	0.076	-0.102
Surface divergence	**22,50ºS**	0.042	0.072	0.062
	25,00ºS	0.046	0.110	0.101
	27,50ºS	0.043	0.137	0.117
Surface specific humidity	**22,50ºS**	0.081	0.048	0.072
	25,00ºS	0.053	0.026	0.059
	27,50ºS	0.004	-0.023	0.059
850 hPa wind direction	**22,50ºS**	0.116	0.068	-0.001
	25,00ºS	0.059	0.117	-0.017
	27,50ºS	-0.004	0.004	0.022
850 hPa relative humidity	**22,50ºS**	0.087	0.091	0.099
	25,00ºS	0.119	0.121	0.123
	27,50ºS	0.138	0.119	0.096
500 hPa relative humidity	**22,50ºS**	0.069	0.147	0.058
	25,00ºS	0.047	-0.145	0.143
	27,50ºS	-0.002	0.113	0.166
500 hPa meridional wind velocity	**22,50ºS**	-0.143	-0.150	-0.100
	25,00ºS	-0.114	0.170	-0.113
	27,50ºS	-0.062	-0.099	-0.076

For some of the predictors such as surface airflow strength, 850 hPa wind direction and 500 hPa meridional wind velocity the grid-box overlying the QC yielded the strongest correlations when compared to other adjacent grid-boxes. Surface divergence had a higher correlation to the south, while surface vorticity and 850 hPa relative humidity revealed marginally stronger correlations to the south-west. In the case of 500 hPa relative humidity the highest correlation values lay to the south-east, while only the surface specific humidity favoured a position to the north-west. Referring back to our

discussion in Section 3.3 it is to be expected that near-surface humidities exhibit stronger correlations over grid-boxes to the east as these are closer to the source of maritime air. However, there does not seem to be any logical reason why the grid-box to the south-east yielded a stronger correlation between Groblersdal's DJF daily rainfall and the observed 500 hPa relative humidity as the majority of weather-producing systems approach this area from the west.

Table 6.7: Calibration results for the spatially diverse set of predictors for the period 1961 – 1985

Predictor	Selected grid-box	Partial correlations	P-values
Surface airflow strength	25,00ºS 30,00ºE	0.091	0.043
Surface vorticity	27,50ºS 26,25ºE	0.018	0.510
Surface divergence	27,50ºS 30,00ºE	0.033	0.406
Surface specific humidity	22,50ºS 26,25ºE	0.047	0.284
Wind direction at 850 hPa	25,00ºS 30,00ºE	0.074	0.104
Relative humidity at 850 hPa	27,50ºS 26,25ºE	0.011	0.543
Relative humidity at 500 hPa	27,50ºS 33,75ºE	0.068	0.133
Meridional wind velocity at 500 hPa	25,00ºS 30,00ºE	-0.088	0.049

A new set of NCEP predictors were then constructed by obtaining the predictors from the GCM grid-boxes with the strongest spatial correlations (as indicated in Table 6.6). The partial correlations and corresponding P-values between the selected predictors and the observed daily summer rainfall at Groblersdal are presented in Table 6.7. The calibration results for this spatially diverse set of predictors only yielded an R^2 of 0.047, a SE of 0.481 and a Chow statistic value of. Compared to the calibration results of the original generic predictor set supplied in Table 6.4, this alternative spatially diverse set of predictors did not prove to be superior after all. The explained variance was found to be lower (R^2 of 0.047 versus 0.066 for the set consisting of predictors from 25ºS 30ºE alone) while the standard error was slightly bigger (SE of 0.481 versus 0.476 for the set consisting of predictors from 25ºS 30ºE alone). The Chow statistic was of comparable magnitude (2.868 versus 2.182 for the set consisting of predictors from 25ºS 30ºE alone) indicating at least that the model is fairly robust.

CHAPTER 7 CONCLUSIONS

7.1 Statistical Downscaling Results

As was mentioned in Chapter 3, the bulk of the summer rainfall over the study area is in the form of thundershowers, with some orographic rain along the higher lying areas to the east. The irregular distribution of thunderstorms in space and time makes convective rainfall difficult to predict or downscale. This inherent property of thunderstorms may well account for the low explained variance of the predictors (Wilby *et al*., 2002).

The set of generic predictors which were identified across all five QCs included airflow strength, vorticity, divergence and specific humidity at the surface, wind direction and relative humidity at 850 hPa as well as relative humidity and meridional wind velocity at 500 hPa. Physical relationships between the daily rainfall at the QC-scale and the synoptic-scale circulation were established for each of these predictors. Generally, all the predictors exhibited a reasonably low explanatory power. The considerable variation in the resultant correlations between the large-scale predictors and the observed daily precipitation at the selected QCs may very well have stemmed from the convective nature of the rainfall patterns, being irregularly distributed in space and time. This may also be the root of the seemingly artificial spatial differences in some of the correlations as demonstrated by the 500 hPa meridional wind velocity. In contrast, relatively low P-values for the 500 hPa relative humidity and 500 hPa meridional wind may stem from the higher correlations between widespread rainfall producing systems such as westerly and tropical-temperate troughs and the observed rainfall across the QCs.

The results from the downscaling model produced with SDSM were not very encouraging. Validation of the downscaling model over the period 1986 to 1999 failed to produce satisfactory results for four of the five QCs as judged against the R^2, standard error, Chow-statistic, Q-Q plots and extreme value analyses. Yet, explained variances

for Groblersdal (B32D) were in line with those reported by Lines *et al*. (2005). An attempt to include predictors from neighbouring GCM grid-boxes did not yield better results.

In the case of Groblersdal it was evident that SDSM underforecasted low rainfall amounts while high daily rainfall amounts were overforecast. This was also reflected in a commensurate underestimation of the return period for extreme daily rainfall amounts. When judged with the STARDEX indices the downscaling model for Groblersdal performed fairly well. The biggest problems were associated with an over-inflated variance, especially in the February daily rainfall, over-predicting the number of significant rainfall events in February and under-predicting the number of wet days in December. However, except in the case of the variation and peaks over 25 mm, the STARDEX index values did fall within the range of the observed climatic index values. It should also be kept in mind that the downscaling model was calibrated against predictors which were selected for the DJF season and subsequently asked to predict daily rainfall for December, January and February separately. This might have contributed to differing biases for the individual months.

The projected changes for the future climate (as dictated by both the A2 and B2 scenarios) were assessed by calculating several delta-statistics. Only a few of the indices revealed a clear linear trend (e.g. a decrease in the percentage wet days for all three months), while most indices exhibited inconsistent changes for DJF across the three future periods. Although most of the inconsistent changes could be fit to an extrapolation of the well documented weak oscillatory variations in the summer rainfall over the north-eastern interior of South Africa, such a fit is thought to be highly coincidental as most GCMs do not contain such cyclic climate forcings. It therefore seems unreasonable to expect this weak amplitude cycle to dominate the projected climate changes. One can rather imagine that a continued westward shift in the average positions of tropical-temperate troughs may cause an area to become wetter during one tri-decadal period as these systems move over it and dry out again during another

tri-decadal period as these systems are being displaced further to the west on the average.

According to the A2 scenario a relatively drier period is expected for Groblersdal during the 2050s. Consistent increases can be expected for the month of February in terms of the number of occurrences exceeding 25 mm per day and the amount of rainfall corresponding to the 90th percentile, while 3-day rainfall totals are not expected to change much. The percentage of wet days will decrease slightly. It seems that the average summer dry-spell lengths are expected to increase across all three tri-decadal periods, while mean wet-spell lengths are expected to decrease slightly during the three summer months. The return period for a daily rainfall amount of 100 mm at Groblersdal increases to 30 years during the 2050s but remains at approximately 20 years throughout the remainder of the 21st century (judged against the median).

With regard to all the rainfall indices discussed, the projected changes under the B2 scenario closely resemble those predicted under the A2 scenario. In general it seems that the differences in the expected GHG forcing under the B2 scenario does not seem to affect any of the rainfall indices differently from the A2 scenario until the late 21st century (i.e. 2070 – 2099). Thus, the downscaling suggests that the future rainfall characteristics are not particularly sensitive to the differences in GHG forcing as described by the A2 and B2 scenarios. Some of the biggest deviations of the B2 projections (from the A2 projections) for the 2080s include:

- a slight decrease in the total monthly rainfall for December and January;
- an increase in the variance during February;
- a decrease in the number of rainfall days with 25 mm or more in December;
- a decrease in the maximum 3-day rainfall total for January; and
- an increase in the maximum 3-day rainfall total for February;
- the return period for a daily rainfall amount of 100 mm at Groblersdal increases to 25 years during the drier period centred on the 2050s but remains at approximately 20 years throughout the remainder of the 21st century (judged against the median).

From an agricultural point of view, the projected changes mentioned above may point to a decrease in groundwater replenishing events during the months of December and January while an increase is projected for February.

It should be noted that the projected changes are often smaller than the model errors which implies that the downscaling model is simply not sensitive enough for these projected changes to be taken at face value. This means that detailed projections cannot be made with any useful confidence. However, the fact that the downscaling procedure provides similar results for the A2 and B2 scenarios suggests that it is at least to some extent robust and stable.

7.2 Future Research

In an attempt to overcome the problems arising from discontinuous daily rainfall – especially when downscaling to a single site – future research should attempt to downscale gridded rainfall from gridded predictor fields over a larger domain. Such an approach will however have to incorporate mathematically complex procedures such as principal component analyses in order to identify the key modes of variability contained in the large gridded variable fields. It may have the advantage that correlations will be sought between the spatial distribution of predictors and the spatial distribution of rainfall, rather than to seek for such relationships over an individual catchment and the near-local values of predictors. Consideration should also be given to incorporate more complicated non-linear statistical downscaling techniques in order to better account for the highly spatially and temporally heterogeneous and discontinuous nature of the daily rainfall over Southern Africa. For example, self-organising maps can be used to statistically downscale daily precipitation over South Africa as Hewitson and Crane (2006) reported reasonable results with such an approach.

Future research should also include projections from an ensemble of models which may include other GCMs such as the Conformal-Cubic Atmospheric Model (C-CAM) which is

run at the modelling group of the Department of Geography, Geo-informatics and Meteorology at the University of Pretoria, South Africa (Engelbrecht *et al*., 2009). This will facilitate the use of a combined dynamical-statistical downscaling approach and candidate predictors that are perhaps more relevant to simulating daily precipitation. These may include stability indices and vertically integrated moisture flux convergence, as opposed to the relative humidity at one or two levels, and thicknesses rather than temperature or geopotential heights. Predictors describing the atmospheric stability were not available to this study. If found to be successful, that approach can also be tested on other study areas within South Africa.

REFERENCES

Ahrens, C.D. 2003. Meteorology Today: An Introduction to Weather, Climate and the Environment. 7th Ed. Pacific Grove, Brooks/Cole. 545pp.

Badas, M.G., Deidda, R. and Piga, E. 2005. Orographic influences in rainfall downscaling. *Advances in Geosciences* 2, 285 – 292.

Bass, B. 1996. Interim report on weather generator project. Focus 4 of IGBP Biospheric Aspects of the Hydrological Cycle (BAHC). Ontario, Environmental Adaption Research Group, Atmospheric Environment Service.

Benestad, R.E. 2004. Empirical-statistical downscaling in climate modeling. *Eos Trans* 85 (42), 417 – 422.

Brunet, N., Verret, R. and Yacowar, N. 1988. An Objective Comparison of Model Output Statistics and "Perfect Prog" Systems in Producing Numerical Weather Element Forecasts. *Weather and Forecasting* (4), 273 – 283.

Burger, G. 1996. Expanded downscaling for generating local weather scenarios. *Climate Research* 7, 111 – 128.

Carbone, G.J. and Bramante, P.D. 1995. Translating monthly temperature from regional to local scale in southeastern United States. *Climate Research* 5, 229 – 242.

CCCma (Canadian Centre for Climate Modelling and Analysis). 2008. CCCma website. Available from http://www.cccma.ec.gc.ca/eng_index.shtml *[Accessed 19 Oct 2008]*.

CICS (Canadian Institute for Climate Studies). 2007. Canadian Climate Impacts and Scenarios FAQ: Downscaling Background. Available from http://www.cics.uvic.ca/scenarios/index.cgi *[Accessed 10 Jul 2007]*.

CICS. 2008. Data access integration portal. Available from http://gaia.ouranos.ca/DAI/predictors-e.html *[Accessed 29 Sep 2008]*.

COMET Program. 2008. Meteorology Education & Training: Statistical Guidance Approaches Module. Available from http://www.meted.ucar.edu/ *[Accessed 12 Oct 2008]*.

Daly, C., Neilson, R.P. and Phillips, D.L. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33, 140 – 158.

DEAT (Department of Environmental Affairs and Tourism). 2000. Environmental Potential Atlas for South Africa. University of Pretoria, Pretoria.

DEAT. 2007. South African Environment Outlook. A Report on the State of the Environment. Department of Environmental Affairs and Tourism, Pretoria. 370pp.

Dent, M.C., Lynch, S.D. and Schulze, R.E. 1989. Mapping Mean Annual and Other Rainfall Statistics over Southern Africa. WRC Report, 109/1/89. Water Research Commission, Pretoria. 230pp.

Díez, E., Primo, C., García-Moya, J.M., Gutiérrez, J.M. and Orfila, B. 2005. Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus* 57A(3), 409 – 433.

Engelbrecht, F.A., McGregor, J.L. and Engelbrecht, C.J. 2009. Dynamics of the Conformal-Cubic Atmospheric Model projected climate-change signal over southern Africa. *International Journal of Climatology* 29, 1013 – 1033.

Gachon, P., Radojevic, M., Harding, A. and Dimitri, P. 2008. Predictor Datasets Derived from the CGCM3.1 T47 and NCEP/NCAR Reanalysis. Canadian Centre for Climate Modelling and Analysis, Montreal. 18pp.

Goldstein, J., Dimitri, P., Gachon, P. and Milton, J. 2004. Development of Climate Scenarios from Statistical Downscaling Methods. Available from http://www.criacc.qc.ca/project/ACFAS_May2004.pdf *[Accessed 27 Oct 2008]*.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B. and Wood, R.A. 2000. Simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics* 16, 147 – 168.

Hay, L.E., McCabe, G.J., Wolock, D.M. and Ayers, M.A. 1991. Simulation of precipitation by weather type analyses. *Water Resource Research* 27, 493 – 501.

Hessami, M., Gachon, P., Ouarda, T.B.M.J. and St-Hilaire, A. 2008. Automated regression-based statistical downscaling tool. *Environmental Modelling and Software* 23(6), 813 – 834.

Hewitson, B.C. and Crane, R.G. 1996. Climate downscaling: techniques and application. *Climate Research* 7, 85 – 95.

Hewitson, B.C. and Crane, R.G. 2002. Self-organizing maps: applications to synoptic climatology. *Climate Research* 22, 13 – 26.

Hewitson, B.C. and Crane, R.G. 2006. Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa. *International Journal of Climatology* 26, 1315 – 1337.

Holton, J.R. 1992. An Introduction to Dynamic Meteorology. 3rd Ed. San Diego, Academic Press. 511pp.

Hostetler, S.W. 1994. Hydrologic and atmospheric models: the (continuing) problem of discordant scales. *Climatic Change* 27, 345 – 350.

Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Kattenberg, A. and Maskell, K. 1996. Climate Change 1995. The Science of Climate Change. Cambridge University Press, Cambridge. 572pp.

Huth, R. 2004. Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. *Journal of Climate* 17, 640 – 652.

Institute for Water Quality Studies. 2001. River Health Programme State of the Rivers Report: Crocodile, Sabie-Sand & Olifants River Systems. DWAF. 43pp.

IPCC. 2007. Summary for Policymakers. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, Cambridge University Press. 18pp.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, D., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77, 437 – 471.

Kalnay, E. 2003. Atmospheric Modeling, Data Assimilation and Predictability. *Cambridge,* Cambridge University Press. 341pp.

Kanamaru, H. and Kanamitsu, M. 2007. Dynamical Downscaling of Global Analysis and Simulation over the Northern Hemisphere. *Monthly Weather Review* 136 (7), 2796–2803.

Kanamitsu, M. 1989. Description of the NMC global data assimilation and forecast system. *Weather Forecasting* 4, 334 – 342.

Kanamitsu, M., Alpert, J.C., Campana, K.A., Caplan, P.M., Deaven, D.G., Iredell, M., Katz, B., Pan, H.-L., Sela, J. and White, G.H. 1991. Recent changes implemented into the global forecast system at NMC. *Weather Forecasting* 6, 425 – 435.

Kilsby, C.G., Cowpertwait, P.S.P., O'Connell, P.E. and Jones, P.D. 1998. Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *International Journal of Climatology* 18, 523 – 539.

Kotze, A.V. 1980. Waarskynlike In- en Uittreedatums van Ryp in Suid-Afrika. Navorsingsinstituut vir Grond en Besproeiing, Pretoria. 247pp.

Kruger, A.C. 2004. Climate of South Africa: Climate regions. WS45. South African Weather Service, Pretoria. 24pp.

Kruger, G.P. 1983. Terrain Morphology Map of Southern Africa. Soil and Irrigation Research Institute, Department of Agriculture, Pretoria.

Kunz, R.P. 2004. Daily Rainfall Data Extraction Utility: User Manual v 1.0. Institute for Commercial Forestry Research, Pietermaritzburg.

Landman, W.A., Mason, S.J., Tyson, P.D., Tennant, W.J., 2001. Statistical downscaling of GCM simulations to Streamflow. *Journal of Hydrology* 252, 221 – 236.

Lines, G.S. and Barrow, E.M. 2002. Regional Climate Change Scenarios in Atlantic Canada Utilizing Downscaling Techniques: Preliminary Results. AMS Preprint, 13-17 January 2002, Olrando.

Lines, G.S., Pancura, M. and Lander, C. 2005. Building Climate Change Scenarios of Temperature and Precipitation in Atlantic Canada using the Statistical Downscaling Model (SDSM). Meteorological Service of Canada, Dartmouth. 41pp.

Lynch, S.D. 2004. Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa. WRC Report 1156/1/04. Water Research Commission, Pretoria. 78pp.

Lynch, S.D. and Schulze, R.E. 2006. Rainfall Database. *In:* Schulze, R.E. (Ed). South African Atlas of Climatology and Agrohydrology. WRC Report 1489/1/06. Water Research Commission, Pretoria. 18pp.

MacKellar, N.C., Hewitson, .C. and Tadross, M.A. 2006. Namaqualand's climate: Recent historical changes and future scenarios. *Journal of Arid Environments* 70(4), 604 – 614.

Maini, P., Kumar, A., Singh, S.V. and Rathore, L.S. 2004. Operational model for forecasting location specific quantitative precipitation and probability of precipitation over India. *Journal of Hydrology* 288, 170 – 188.

Marzban, C., Sandgathe, S. and Kalnay, E. 2005. MOS, Pergect Prog, and Reanalysis. *Monthly Weather Review* 134, 657 – 663.

McGregor, J.J. 1997. Regional climate modelling. *Meteorological and Atmospheric Physics* 63, 105 – 117.

McGregor, J.L. and Dix, M.R. 2001. The CSIRO conformal-cubic atmospheric GCM. IUTAM Symposium on Advances in Mathematical Modelling of Atmosphere and Ocean Dynamics. Hodnet, P.F. Ed. Kluwer, 197 – 202.

Mendenhall, W. and Sincich, T. 2003. A Second Course in Statistics: Regression Analysis. 6th Ed. Upper Saddle River, Pearson. 880pp.

Midgley, D.C., Pitman, W.V. and Middleton, B.J. 1994. Surface Water Resources of South Africa 1990, User's Manual. Water Resources 1990 Joint Venture. WRC Report 298/1/94. Water Research Commission, Pretoria. 191pp.

Murphy, J. 1998. An evaluation of statistical and dynamical techniques for downscaling local climate. *Journal of Climate* 12, 2256 – 2284.

Nakićenović, N., Davidson, O., Davis, G., Grübler, A., Kram, T., La Rovere, E.L., Metz, B., Morita, T., Pepper, W., Pitcher, H., Sankovski, A., Shukla, P., Swart, R., Watson, R. and Dadi, Z. 2000. Emission Scenarios: A Summary of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. 599pp.

NCEP (National Centers for Environmental Prediction). 2008. About NCEP. Available from http://www.ncep.noaa.gov *[Accessed 19 Oct 2008]*.

NOAA (National Oceanic and Atmospheric Administration). 2008. Climate Diagnostics Center map room. Available from http://www.cdc.noaa.gov *[Accessed 27 May 2008]*.

Oelschlägel, B. 1995. A method for downscaling global climate model calculations by a statistical weather generator. *Ecological Modelling* 82, 199 – 204.

Pope, V.D., Gallani, M.L., Rowntree, P.R. and Stratton, R.A. 2000. The impact of new physical parameterizations in the Hadley Centre climate model - HadAM3. *Climate Dynamics* 16, 123 – 146.

Rossel, F. and Garbrecht, J. 2001. Spatial variability and downscaling of precipitation. *Physics and Chemistry of the Earth (B)* 26, 11-12, 863 – 867.

Rubinstein, R.Y. 1981. Simulation and the Monte Carlo Method. Wiley, New York.

Semenov, M.A. and Brooks, R.J. 1999. Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain. *Climate Research* 11, 137 – 148.

Schulze, R.R. 1994. Climate of South Africa. Part 8: General Survey. Pretoria, Government Printer.

Schulze, R.E. 1997. South African Atlas of Agrohydrology and Climatology. TT82/96. Water Research Commission, Pretoria.

Schulze, R.E. (Ed). 2006. South African Atlas of Climatology and Agrohydrology. WRC Report 1489/1/06. Water Research Commission, Pretoria.

Shaw, E. 1994. Hydrology in Practice. 3rd Ed. London, Chapman & Hall. 592pp.

STARDEX (Statistical and Regional dynamical Downscaling of Extremes for European regions). 2008. Available from http://www.cru.uea.ac.uk/projects/stardex/ *[Accessed 11 Oct 2008]*.

Stern, R.D. and Coe, R. 1984. A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Association* 147, 1 – 34.

Taljaard, J.J. 1994. Controls of the Weather and Climate of South Africa. Technical Paper No. 27. South African Weather Bureau, Pretoria. 45pp.

Taljaard, J.J. 1996. Rainfall in South Africa. Technical Paper No. 32. South African Weather Bureau, Pretoria. 98pp.

Trenberth, K.E., Jones, P.D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J.A., Rusticucci, M., Soden, B. and Zhai, P. 2007. Observations: Surface and Atmospheric Climate Change. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. 102pp.

Tyson, P.D. and Dyer, T.G.J. 1975. Mean annual fluctuations of precipitation in the summer rainfall region of South Africa. *South African Journal of Science* 76, 340 – 341.

Tyson, P.D. 1986. Climatic Change and Variability in Southern Africa. Cape Town, Oxford University Press. 218pp.

Von Storch, H., Langerberg, H. and Feser, F. 2000. A spectral nudging technique for dynamical downscaling purposes. *Monthly Weather Review* 128, 3664 – 3673.

Walker, N.J. and Schulze, R.E. 2006. An assessment of sustainable maize production under different management and climate scenarios for smallholder agro-ecosystems in KwaZulu-Natal, South Africa. *Physics and Chemistry of the Earth* 31, 995 – 1002.

Walker, N.J. and Schulze, R.E. 2008. Climate change impacts on agro-ecosystem sustainability across three climate regions in the maize belt of South Africa. *Agriculture, Ecosystems and Environment* 124, 114 – 124.

Wallis, T.W.R. and Griffiths, J.F. 1995. An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). *Agricultural and Forest Meteorology* 73, 115 – 133.

Wigley, T.M.L., Jones, P.D., Briffa, K.R. and Smith, G. 1990. Obtaining sub-grid scale information from coarse resolution general circulation model output. *Journal of Geophysical Research* 95, 1943 – 1953.

Wilby, R.L., Barnsley, N. and O'Hare, G. 1995. Rainfall variability associated with Lamb weather types: the case for incorporating weather fronts. *International Journal of Climatology* 15, 1241 – 1252.

Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P. and Mearns, L.O. 2004. Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. IPCC Task Group on Scenarios for Climate Impact Assessment (TGCIA). 27pp.

Wilby, R.L. and Dawson, C.W. 2007. SDSM 4.2 User Manual. Available from https://co-public.lboro.ac.uk/cocwd/SDSM/ *[Accessed 1 Nov 2007]*.

Wilby, R.L., Dawson, C.W. and Barrow, E.M. 2002. SDSM - a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling and Software* 17(2), 145 – 157.

Wilby, R.L., Hassan, H. and Hanaki, K. 1998. Statistical downscaling of hydrometeorological variables using general circulation model output. *Journal of Hydrology* 205, 1 – 19.

Wilby, R.L., Whitehead, P.G., Wade, A.J., Butterfield, D., Davis, R.J. and Watts, G. 2006. Integrated modelling of climate change impacts on water resources and quality in a lowland catchment: River Kennet, UK. *Journal of Hydrology* 330, 204 – 220.

Wilby, R.L. and Wigley, T.M.L. 1997. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography* 21 (4), 530 – 548.

Wilby, R.L. and Wigley, T.M.L. 2000. Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology* 20, 641 – 661.

Wilks, D.S. 1995. Statistical Methods in the Atmospheric Sciences. San Diego, Academic Press. 467pp.

Wilks, D.S. 1999. Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research* 11, 125 – 136.

WDC (World Data Centre for Climate). 2008. The IPCC data distribution centre. Available from http://pluto.dkrz.de/IPCC_DDC/IS92a/ *[Accessed 17 Oct 2008]*.

Wotling, G., Bouvier, C, Danloux, J and Fritsch, J.M. 2000. Regionalization of extreme precipitation distribution using the principal components of the topographical environment. *Journal of Hydrology* 233, 86 – 101.

Yarnal, B. 1993. Synoptic Climatology in Environmental Analysis: A Primer. London, Belhaven Press. 195pp.

Yoshimura, K. and Kanamitsu, M. 2008. Dynamical global downscaling of global reanalysis. *Monthly Weather Review* 136 (8), 2983 – 2998.

APPENDIX A PREDICTOR SCREENING

Evaluation of potential predictors (refer to Section 6.1)

Predictor	B32D	B20A	B11K	B12D	B41A	Total for QCs	Total (Incl. Lags)	Lines *et al.* (2005)	Wilby & Wigley (2000)	Maini *et al.* (2004)	Hessami *et al.* (2008)
mslp	1	1	1	1	3	7	12		×		
l1mslp	1	1	1	1	1	5					
p850	1	1	1	1	1	5	10				
l1p850	1	1	1	1	1	5					
p500	1	1	1	1	1	5	13				×
l1p500	2	1	3	1	1	8					
temp	2	3	1	2	1	9	14				
l1temp	1	1	1	1	1	5					
p__f	3	3	1	2	1	10	15				×
l1p__f	1	1	1	1	1	5					
p__u	2	1	2	1	1	7	13	×	×		
l1p__u	1	1	1	2	1	6					
p__v	2	1	1	1	2	7	15	×			
l1p__v	1	2	2	2	1	8					
p8_f	1	1	1	1	1	5	12				
l1p8_f	1	1	1	1	3	7					
p8_u	2	2	1	1	1	7	14				
l1p8_u	1	1	1	3	1	7					
p8_v	1	1	1	1	1	5	12			×	×
l1p8_v	1	1	2	1	2	7					
p__z	1	2	1	1	3	8	15				
l1p__z	1	1	3	1	1	7					
p_th	3	1	1	2	1	8	15				
l1p_th	1	1	1	1	3	7					
p_zh	1	1	1	1	1	5	15	×			
l1p_zh	2	2	3	2	1	10					
p8_z	1	1	1	1	1	5	10			×	
l1p8_z	1	1	1	1	1	5					
p8th	3	3	1	3	2	12	17				
l1p8th	1	1	1	1	1	5					
p8zh	1	1	1	1	1	5	10				
l1p8zh	1	1	1	1	1	5					
p5_f	1	1	3	1	1	7	12				
l1p5_f	1	1	1	1	1	5					
p5_u	1	1	1	1	1	5	10				
l1p5_u	1	1	1	1	1	5					
p5_v	3	2	3	3	2	13	18		×		
l1p5_v	1	1	1	1	1	5					
p5_z	1	1	1	1	1	5	11	×			
l1p5_z	1	1	2	1	1	6					
p5th	1	1	1	1	1	5	10		×		
l1p5th	1	1	1	1	1	5					
p5zh	1	1	1	3	2	8	15	×			
l1p5zh	1	3	1	1	1	7					
rhum	1	1	1	1	1	5	10			×	
l1rhum	1	1	1	1	1	5					
shum	1	1	1	1	1	5	15	×	×		
l1shum	3	2	2	1	2	10					
r850	1	3	3	3	2	12	17	×		×	
l1r850	1	1	1	1	1	5					
r500	3	3	3	3	3	15	20			×	×
l1r500	1	1	1	1	1	5					

Explanation of abbreviations used for potential predictors (refer to Section 6.1)

Abbreviation	Level	Climatic Element	Lag
mslp	Mean sea level	Pressure	none
l1mslp	Mean sea level	Pressure	1-day
p850	850 hPa	Geopotential height	none
l1p850	850 hPa	Geopotential height	1-day
p500	500 hPa	Geopotential height	none
l1p500	500 hPa	Geopotential height	1-day
temp	Surface	Temperature	none
l1temp	Surface	Temperature	1-day
p__f	Surface	Geostrophic airflow	none
l1p__f	Surface	Geostrophic airflow	1-day
p__u	Surface	Zonal airflow component	none
l1p__u	Surface	Zonal airflow component	1-day
p__v	Surface	Meridional airflow component	none
l1p__v	Surface	Meridional airflow component	1-day
p8_f	850 hPa	Geostrophic airflow	none
l1p8_f	850 hPa	Geostrophic airflow	1-day
p8_u	850 hPa	Zonal airflow component	none
l1p8_u	850 hPa	Zonal airflow component	1-day
p8_v	850 hPa	Meridional airflow component	none
l1p8_v	850 hPa	Meridional airflow component	1-day
p__z	Surface	Vorticity	none
l1p__z	Surface	Vorticity	1-day
p_th	Surface	Wind direction	none
l1p_th	Surface	Wind direction	1-day
p_zh	Surface	Divergence	none
l1p_zh	Surface	Divergence	1-day
p8_z	850 hPa	Vorticity	none
l1p8_z	850 hPa	Vorticity	1-day
p8th	850 hPa	Wind direction	none
l1p8th	850 hPa	Wind direction	1-day
p8zh	850 hPa	Divergence	none
l1p8zh	850 hPa	Divergence	1-day
p5_f	500 hPa	Geostrophic airflow	none
l1p5_f	500 hPa	Geostrophic airflow	1-day
p5_u	500 hPa	Zonal airflow component	none
l1p5_u	500 hPa	Zonal airflow component	1-day
p5_v	500 hPa	Meridional airflow component	none
l1p5_v	500 hPa	Meridional airflow component	1-day
p5_z	500 hPa	Vorticity	none
l1p5_z	500 hPa	Vorticity	1-day
p5th	500 hPa	Wind direction	none
l1p5th	500 hPa	Wind direction	1-day
p5zh	500 hPa	Divergence	none
l1p5zh	500 hPa	Divergence	1-day
rhum	Surface	Relative humidity	none
l1rhum	Surface	Relative humidity	1-day
shum	Surface	Specific humidity	none
l1shum	Surface	Specific humidity	1-day
r850	850 hPa	Relative humidity	none
l1r850	850 hPa	Relative humidity	1-day
r500	500 hPa	Relative humidity	none
l1r500	500 hPa	Relative humidity	1-day