HIGH-RESOLUTION CLIMATE VARIABLE GENERATION FOR THE WESTERN CAPE

SARAH JOAN JOUBERT

Thesis presented in partial fulfilment of the requirements for the degree of Master Natural Sciences at the University of Stellenbosch.

Supervisor: Mr. A van Niekerk

March 2007

DECLARATION

I, the undersigned, hereby declare that the work contained in this is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: -------------------------------

Date: -------------------------------

ABSTRACT

Due to the relative scarcity of weather stations, the climate conditions of large areas are not adequately represented by a weather station. This is especially true for regions with complex topographies or low population densities. Various interpolation techniques and software packages are available with which the climate of such areas can be calculated from surrounding weather stations' data. This study investigates the possibility of using the software package ANUSPLIN to create accurate climate maps for the Western Cape, South Africa.

ANUSPLIN makes use of thin plate smoothing splines and a digital elevation model to convert point data into grid format to represent an area's climatic conditions. This software has been used successfully throughout the world, therefore a large body of literature is available on the topic, highlighting the limitations and successes of this interpolation method.

Various factors have an effect on a region's climate, the most influential being location (distance from the poles or equator), topography (height above sea level), distance from large water bodies, and other topographical factors such as slope and aspect. Until now latitude, longitude and the elevation of a weather station have most often been used as input variables to create climate grids, but the new version of ANUSPLIN (4.3) makes provision for additional variables. This study investigates the possibility of incorporating the effect of the surrounding oceans and topography (slope and aspect) in the interpolation process in order to create climate grids with a resolution of 90m x 90m. This is done for monthly mean daily maximum and minimum temperature and the mean monthly rainfall for the study area for each month of the year.

Not many projects where additional variables have been incorporated in the interpolation process using ANUSPLIN are to be found in the literature, thus further investigation into the correct transformation and the units of these variables had to be done before they could be successfully incorporated. It was found that distance to oceans influences a region's maximum and minimum temperatures, and to a lesser extent rainfall, while aspect and slope has an influence on a region's rainfall.

In order to assess the accuracy of the interpolation process, two methods were employed, namely statistical values produced during the spline function calculations by ANUSPLIN, and the removal of a selected number of stations in order to compare the interpolated values with the actual measured values. The analysis showed that more accurate maps were obtained when additional variables were incorporated into the interpolation process.

Once the best transformations and units were identified for the additional variables, climate maps were produced in order to compare them with existing climate grids available for the study area. In general the temperatures were higher than those of the existing grids. For the rainfall grids ANUSPLIN's produced higher rainfall values throughout the study region compared to the existing grids, except for the Southwestern Cape where the rainfall values were lower on north-facing slopes and high-lying areas.

Keywords: Digital elevation models (DEM), spatial interpolation, ANUSPLIN, topographic dependence, thin plate smoothing splines, climate

OPSOMMING

Vanweë die relatiewe skaarsheid van weerstasies gebeur dit dikwels dat die klimaat van groot gebiede nie akkuraat voorgestel word nie. Dit is veral die geval in gebiede met komplekse topografie of lae bevolkingsdigthede. Verskeie interpolasiemetodes en sagteware-pakkette kan gebruik word om sulke gebiede se klimaat vanaf omliggende weerstasies se data te bereken. Hierdie studie ondersoek die moontlikheid om deur middel van die sagteware-pakket ANUSPLIN akkurate klimaatkaarte te skep vir die Wes-Kaap in Suid-Afrika.

ANUSPLIN is ’n statistiese pakket wat gebruik maak van vlakplaat-strykende latfunksies (*thin plate smoothing splines*) en ’n digitale elevasiemodel (DEM) om puntdata te omskep in roosterformaat wat ’n gebied se klimaattoestande verteenwoordig. Die sagteware-pakket is reeds suksesvol regoor die wêreld toegepas en heelwat literatuur oor die suksesse en beperkings van hierdie interpolasiemetode is beskikbaar.

Verskeie faktore, onder andere ligging (afstand van die pole of ewenaar), topografie (hoogte bo sëespiel), afstand vanaf groot wateroppervlaktes, en ander topografiese invloede soos helling en aansig beïnvloed ’n gebied se klimaat. Tot dusver is lengte- en breedtegrade en die hoogte bo seëspiel van ’n weerstasie as veranderlikes die meeste gebruik om kaarte te skep, maar die nuwe weergawe van ANUSPLIN (4.3) maak voorsiening vir addisionele veranderlikes. Hierdie studie ondersoek die moontlikheid om die effek van oseane en topografie (helling en aansig) te inkorporeer in die interpolasieproses om klimaatroosters met ’n resolusie van 90m x 90m te skep. Dit is gedoen vir maandelikse gemiddelde reënval en maksimum en minimum temperature vir die studiegebied vir elke maand van die jaar.

Min projekte is in die literatuur te vinde waar addisionele veranderlikes in ANUSPLIN gebruik word, dus was verskeie verwerkings nodig om die optimale transformasie en eenhede vir die veranderlikes te vind voor dit ingesluit kon word. Daar is bevestig dat afstand vanaf die naaste oseaan ‘n invloed op ‘n streek se maksimum en minimum temperature, en tot ‘n mindere mate reënval het, terwyl aansig en helling ‘n invloed op ‘n streek se reënval het.

Om die akkuraatheid van die interpolasieproses te bepaal, is twee metodes gebruik, naamlik statistiese waardes wat deur ANUSPLIN se latfunksie-berekening voorsien word, sowel as om ‘n sekere getal stasies se data te onttrek van die datastel en dan die geinterpoleerde waardes te vergelyk met die waargenome waardes vir die weerstasies. Die resultate bevestig dat akkurater

kaarte verkry is wanneer addisionele veranderlikes by die interpolasieprosess betrek word.

Nadat die beste transformasies en eenhede vir die addisionele veranderlikes geïdentifiseer is, is die klimaatkaarte met bestaande klimaatroosters van die studiegebied vergelyk. Oor die algemeen is ANUSPLIN se berekende temperature hoër as dié van bestaande roosters, terwyl die reënvalkaarte se waardes hoër is behalwe in die Suidwes-Kaap waar dit laer is vir noordelike hellings en hoogliggende gebiede.

Sleutelwoorde: Digitale elevasiemodel (DEM), ruimtelike interpolasie, ANUSPLIN, topografiese afhanklikheid, vlakplaat-strykende latfunksies, klimaat

ACKNOWLEDGEMENTS

I would like to thank the following people for assisting me with this research project:

- The Department of Geography and Environmental Studies, University of Stellenbosch for purchasing the software used in my study and the support of the staff.
- Dr MF Hurchinson, Australian National University, for assisting me with all my enquiries and giving useful advice on how to operate the software program.
- Wolfgang Lück for the help with solving so many technical difficulties and Dr M Kidd of University of Stellenbosch's Statistical Department for assistance with statistical procedures. Dr P de Necker, University of Stellenbosch for language editing and assitance.
- Dr MER and Mrs K Burns and my parents for their support and advice.
- Mr. Adriaan van Niekerk, my supervisor, for his support, advice and steering me in the right direction.
- My fellow geography students, for their support and friendship during my post-graduate studies.

CONTENTS

PAGE

DECLARATION II
ABSTRACT III
OPSOMMING V
ACKNOWLEDGEMENTS VII
FIGURES XI
TABLES XII
1. THE IMPORTANCE OF CLIMATE DATA 1
1.1 CLIMATE 1
1.1.1 Climate versus weather 1
1.1.2 Climate control factors 2
1.1.3 Climate data – where does it come from? 3
1.1.3.1 Climate modelling 3
1.1.3.2 Capturing of raw data 4
1.2 PROJECT DESCRIPTION 5
1.2.1 Research problem 5
1.2.2 Research aim 6
1.2.3 Objectives 6
2. FROM WEATHER STATIONS TO CLIMATE GRIDS 9
2.1 INTERPOLATION TECHNIQUES 9
2.1.1 Inverse distance weighting 10
2.1.2 Splines 10
2.1.3 Kriging 11
2.1.4 Regression 11
2.1.5 A South African study 11
2.2 INTERPOLATION PROGRAMS 12
2.2.1 PRISM 13
2.2.2 ANUSPLIN 13
2.2.3 Neural networks 14

2.2.4 MTCLIM and Daymet 14

2.2.5 CIAT 15

2.3 SOUTHERN AFRICA CLIMATE DATABASES 15

2.3.1 Climate data produced by climate modelling for Southern Africa 16

2.3.2 Completed Southern African interpolation projects 17

2.4 CONCLUSION 19

3. METHODOLOGY 20

3.1 THE STUDY AREA 20

3.1.1 Selection of the study area 20

3.1.2 Description of the study area 22

3.1.2.1 Topography 22

3.1.2.2 Climate 23

3.2 SELECTION OF AN INTERPOLATION SOFTWARE PACKAGE 24

3.2.1 ANUSPLIN as the chosen interpolator 25

3.2.2 ANUSPLIN: technical description 26

3.3 DESCRIPTION OF THE DATA 27

3.3.1 Digital elevation model 27

3.3.2 Weather stations 27

3.4 ADDITIONAL VARIABLE SELECTION 29

3.4.1 Topography (slope and aspect) 29

3.4.2 Distance to large water bodies 32

3.5 CORRELATION ANALYSES 33

3.5.1 Correlation with topography 33

3.5.2 Correlation with distance to oceans 37

3.5.3 Summary of the correlation analyses 39

3.6 SUITABLE TRANSFORMATIONS AND UNITS ADDITIONAL VARIABLES 41

3.6.1 Interpolation analysis: Monthly mean daily maximum temperature 45

3.6.2 Interpolation analysis: Monthly mean daily minimum temperature 49

3.6.3 Interpolation analysis: Mean monthly rainfall 51

3.6.3.1 Topography 53

3.6.3.2 Distance to ocean 56

3.6.3.3 Combining topography and distance to ocean variables 58

3.6.4 Summary of the parameterization investigations 61

3.7 COMPARISON OF ANUSPLIN-CREATED GRIDS WITH EXISTING GRIDS....... 63

3.7.1 Monthly mean daily maximum temperature....... 64

3.7.2 Monthly mean daily minimum temperature....... 67

3.7.3 Mean monthly rainfall....... 70

3.7.4 Conclusion of climate grid comparison....... 73

4. CONCLUSION AND RECOMMENDATIONS....... 74

4.1 INTRODUCTION....... 74

4.2 SUCCESSFUL INCORPORATION OF ADDITIONAL VARIABLES....... 75

4.3 RECOMMENDATIONS FOR FURTHER RESEARCH....... 78

3.7.5 Increasing the accuracy of the interpolation process....... 78

3.7.6 Taking climate interpolation further....... 80

4.4 CONCLUSION....... 80

REFERENCES....... 82

PERSONAL COMMUNICATION....... 89

APPENDICES....... 90

APPENDIX A: Brief summary of the programs which make up the ANUSPLIN package....... 90

APPENDIX B: Data flow diagram for creating climate grids....... 91

APPENDIX C1: Wind-roses representing dominant winds for January....... 92

APPENDIX C2: Wind-roses representing dominant winds for April....... 93

APPENDIX C3: Wind-roses representing dominant winds for July....... 94

APPENDIX C4: Wind-roses representing dominant winds for October....... 95

APPENDIX D: Monthly mean daily maximum and minimum temperature and mean monthly rainfall grids created by ANUSPLIN for each month of the year....... 96

FIGURES

PAGE

Figure 1.1: Orographic lifting and the formation of rain shadow deserts 2
Figure 1.2: Research design 7
Figure 3.1: Main rivers and river catchments in the Western Cape 21
Figure 3.2: Topography of the Western Cape region 22
Figure 3.3: Rainfall seasonality of South Africa 23
Figure 3.4: Available long-term weather stations for the study region and surrounding areas 28
Figure 3.5a: Azimuth: angular direction of sun in degrees 30
Figure 3.5b: Slope or angle of the illumination source above the horizon in degrees 30
Figure 3.6: Hillshade with an azimuth of 180 degrees and an altitude of 45 degrees 31
Figure 3.7: Wind-roses indicating the wind directions during January for the study area 40
Figure 3.8: Distribution of weather stations selected for testing 44
Figure 3.9: Difference in maximum temperatures between ANUSPLIN and ATLAS grids during January 65
Figure 3.10: Difference in maximum temperatures between ANUSPLIN and ATLAS grids during June 66
Figure 3.11: Difference in minimum temperatures between ANUSPLIN and ATLAS grids during January 68
Figure 3.12: Difference in minimum temperatures between ANUSPLIN and ATLAS grids during June 69
Figure 3.13: Difference in rainfall between ANUSPLIN and ATLAS grids during January 71
Figure 3.14: Difference in rainfall between ANUSPLIN and ATLAS grids during June 72

TABLES

PAGE

Table 3.1: Standard Pearson's coefficient (r) and the coefficient of determination (r2) for topography (illumination value with a 315o azimuth) and rainfall and temperature................34

Table 3.2: Standard Pearson's coefficient (r) and the coefficient of determination (r2) for topography (illumination value with an 180o azimuth) and rainfall and temperature..............36

Table 3.3: Standard Pearson's coefficient (r) and the coefficient of determination (r2) for distance to nearest ocean and rainfall and temperature..38

Table 3.4: Aspect of weather stations 40

Table 3.5: Shortest distance to ocean from weather stations 41

Table 3.6: Number of weather stations and test stations in the three density zones 44

Table 3.7: Quadvariate model description for distance to ocean variable analysis for maximum temperature 45

Table 3.8: Spline function statistics for the maximum temperature quadvariate models 47

Table 3.9: Error values for the maximum temperature quadvariate models....................................48

Table 3.10: A and B: Spline function statistics (A) and error calculations (B) for maximum temperature for the trivariate model and the best-performing quadvariate model. 48

Table 3.11: Spline function statistics for the minimum temperature quadvariate models................50

Table 3.12: Error values for the minimum temperature quadvariate models 50

Table 3.13: A and B: Spline function statistics (A) and error calculations (B) for minimum temperature for the trivariate model and the best-performing quadvariate model. 51

Table 3.14: Trivariate model description for the standard variables for rainfall 52

Table 3.15: Spline function statistics for the rainfall trivariate models 52

Table 3.16: Error values for the rainfall trivariate models..53

Table 3.17: Quadvariate model description for the topography (slope and aspect) variable analysis for rainfall...54

Table 3.18: Spline function statistics for the rainfall quadvariate (topography) models 54

Table 3.19: Error values for the rainfall quadvariate (topography) models......................................55

Table 3.20: Quadvariate model description for the distance to ocean parameter analysis for rainfall 56

Table 3.21: Spline function statistics for the rainfall quadvariate (distance to ocean) models..........57

Table 3.22: Error values for the rainfall quadvariate (distance to ocean) models 57

Table 3.22 (Continued): Error values for the rainfall quadvariate (distance to ocean) models......... 58

Table 3.23: Spline function statistics for the trivariate (Triv) model, quadvariate models (distance: Quad-1 and topography: Quad-2) and the quinvariate model (Quin) for rainfall 59

Table 3.24: Error values for the trivariate (Triv) model, quadvariate models (distance: Quad-1 and topography: Quad-2) and the quinvariate model (Quin) for rainfall................................ 60

Table 3.25: Parameterization for the interpolation process for the three climate variables.............. 62

Table 3.26: Mean values of grids produced when the ATLAS grid was subtracted from the ANUSPLIN grid for maximum temperature 64

Table 3.27: Mean values of grids produced when the ATLAS grid was subtracted from the ANUSPLIN grid for minimum temperature 67

Table 3.28: Mean values of grids produced when the ATLAS grid was subtracted from the ANUSPLIN grid for monthly total rainfall... 70

CHAPTER 1: THE IMPORTANCE OF CLIMATE DATA

Climate data is a prerequisite for many fields of study and form part of various projects and decision making processes. A region's economic activities, especially in areas where farming is a primary economic enterprise, are best understood when taking its climate into account. Comprehending a region's climate and the prediction of future climatic changes are of utmost importance when managing regional resources. Complete and accurate climate datasets form part of simulation and decision-making support tools. Such datasets are also a requirement for the efficient modelling of most environmental processes. Climate data are also used in hydrological modelling and other scenario simulations which play important roles in the management of natural disasters, such as droughts and excessive precipitation, which can have a detrimental effect on a region's prosperity. Climate grids are used to classify ecological zones for the purpose of allocating land uses and for identifying priority areas for conservation of biodiversity. Clearly climate data plays a role in various daily human activities and in scientific research.

In the first chapter of this report the key concepts related to climate/weather will be underlined, the factors responsible for influencing a region's climate will be discussed and the various sources used to obtain climate data will be set out. The second chapter investigates the creation of climate grids (the format in which climate data is most often represented) and identifies climate grids available for the study region.

1.1 CLIMATE

1.1.1 Climate versus weather

Lutgens & Tarbuck (1998) define *climate* as the aggregate of weather at a given location for a given time period while *weather* conditions are the state of the atmosphere at any given time and can include temperature, precipitation, relative humidity, solar radiation, cloudiness, and wind speed and direction.

The Intergovernmental Panel on Climate Change's (2001: no page) glossary's definition for climate is:

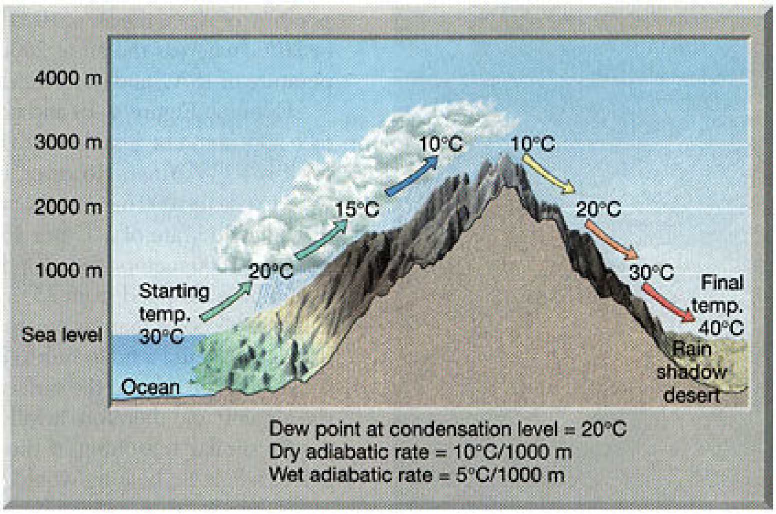
> "Climate in a narrow sense is usually defined as the "average weather", or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or

millions of years. The classical period is 30 years, as defined by the World Meteorological Organization (WMO). These quantities are most often surface variables such as temperature, precipitation, and wind. Climate in a wider sense is the state, including a statistical description, of the climate system."

During thus study all references will be made to climate, since weather is a short term phenomena, changing over short periods of time and from year to year. Climate is the average of weather and therefore provides a better overall picture of a region's normal atmospheric conditions.

1.1.2 Climate control factors

From a study of the literature (Hutchinson 1998a; Lutgens & Tarbuck 1998; Van Zyl 2003) it is clear that a region's climate is controlled by a couple of factors. The most important one is geographical position, because latitude determines the amount of solar energy received by a specific region. Another important factor is elevation since temperature and atmospheric pressure decrease with altitude (the former known as the temperature lapse rate). Apart from elevation, topography also plays a role in the amount of precipitation experienced in a region. For example, as air ascends a mountain slope, adiabatic cooling generates clouds and precipitation (Figure 1.1).



Source: Lutgens & Tarbuck 1998:99

Figure 1.1: Orographic lifting and the formation of rain shadow deserts

Land heats more rapidly and to higher temperatures than water and vice versa, thus water has a moderating influence on the climate of land masses. The proximity to large water bodies is thus also a factor, while wind and pressure systems also have an effect during the various seasons.

1.1.3 Climate data – where does it come from?

A researcher or decision-maker possesses different tools for collecting the climate data necessary for monitoring, simulation and modelling processes, either by means of capturing raw data or simulating weather conditions.

1.1.3.1 Climate modelling

In order to estimate a region's climate, it is possible to simulate the many processes that are responsible for the climatic conditions experienced by using a climate model. Such models attempt to describe the climate system in terms of basic physical, chemical and biological principles by means of a series of numerical equations expressing these laws (McGuffie & Henderson-Sellers 2005).

McGuffie & Henderson-Sellers (2005) give a brief overview on the history of climate models and the types of climate models available. It is important to realise that climate models are a simplification of the real world and that these models can be slow and expensive to use, even with the speed and processing capacity of modern computers. This is especially true when high resolution climate data are required.

Climate models require a wide range of input data (i.e. land-use, sea-surface information, topography, soil, land cover) and the pre-processing of this data is time-consuming and requires large amounts of disk storage space. Once an output dataset has been generated, more data processing is often required before it can be used in other research projects. When climate simulation is too costly and time-consuming, and a researcher requires accurate high resolution climate data, a simpler way of obtaining climate information for areas not represented by weather stations is interpolation where the data from surrounding stations are used. Interpolation is often used to obtain a higher resolution from the output generated by climate models (Bootsman *et al*. 2004; Cramer *et al*. 2000 and Price *et al.* 2004). Climate models are valuable for areas where no weather stations are located since some climate models (e.g. regional atmospheric climate model (RAMS)) do not essentially require observed climate data as input in order to simulate climate (Cautenet 2006, pers com). However, climate modelling software is provided with upper- and lower boundaries (sea-surface temperatures and atmospheric fields). Weather station data may also be used to provide the model with an initial state from which it simulates weather or climatic conditions (McGuffie & Henderson-Sellers 2005).

1.1.3.2 Capturing of raw data

The most modern approach to data capturing is the use of remote sensing devices such as satellites. Jeffrey, Carter, Moodie & Beswick (2001) summarize the shortcomings and difficulties of using remote sensing as a data capturing tool as follows: the development of calibration models to transform the signal and then to extract the required data is a difficult process requiring extensive research and highly sophisticated software and hardware. Remote sensing data are expensive, while the analysis and processing of these datasets require costly software and skilled scientists. The accuracy of satellite measurements are also less accurate compared to data generated at weather stations. Polar orbiting satellites (high spatial resolution) only provide data over limited domains at a specific time, whilst geostationary satellites (comparatively lower spatial resolution) monitor the same domain continuously (Engelbrecht 2007, pers com). Available literature makes it clear that until these limitations have been addressed, ground-based observational data remain the preferred source of climate data.

Ground-based observations take place in the form of collecting climate data from a weather station grid system which is set up and located throughout a region. Other less widely used data-capturing methods exist such as radiosondes which are balloon-carried weather measuring instruments released from weather stations situated on the ground at regular intervals (Van Zyl 2003).

Although the data from ground-based observations are more accessible, they also have problems. Similar to remote sensing, errors are created while transferring or storing data. Weather stations may also dysfunction and require servicing (Linacre 1992). The most significant problem with ground-based observations is that weather stations are often relatively sparsely situated, especially in mountainous regions or areas with low population densities. *This results in vast regions not being sufficiently represented by weather stations*.

The following chapter contains a comprehensive literature review done on interpolation techniques and software programs suitable for creating climate grids calculated from a network of ground-based weather stations. First a brief overview of the aim and objectives for this research project is given.

1.2 PROJECT DESCRIPTION

In the first part of this chapter the importance of climate and climate data has been highlighted. The next section provides an overview of the project, its aim and specific objectives.

1.2.1 Research problem

Climate data is available in a variety of formats depending on the use of the datasets, but climate grids – pixel by pixel representations – are most often used as part of simulation and modelling projects. A major problem with such projects is that the resolution of the climate data places a limitation on the accuracy and detail of the overall project. A preliminary literature review and a study of the climate grids available for the Western Cape region identified a need for high resolution (less than 1km x 1km) grids.

The Western Cape is South Africa’s major wine producing region. A method by means of which climatic patterns can be predicted at vineyards not covered by weather stations will be invaluable in understanding the terroir/vine/wine interaction. The improved knowledge of the terroir/vine/wine interaction will improve site selection for specific cultivars and will assist decision makers in their site-specific adaptation of management strategies.

The literature review indicated that climate interpolation programs are most often used for the creation of climate grids when long-term weather station data are available. It was found that the interpolated temperature grids were generally more accurate than the rainfall grids that were produced. This problem was generally experienced by interpolation projects performed throughout the world (Jeffrey *et al.* 2001; Price *et al.* 2004).

In order to achieve higher accuracy during the interpolation process, modern interpolation packages include the possibility of incorporating additional variables and changing the current variables by means of mathematical transformations. Limited experimentation with correct transformations and units for such additional variables has been done using the available interpolation packages, thus further research into transformation of and units for such variables is necessary in order to achieve better accuracy in the interpolation process.

1.2.2 Research aim

The aim of this study is to create high-resolution climate grids for the Western Cape region of South Africa by means of a suitable interpolation technique. Each cell in the climate grid will be associated with the value of a climatic variable, thus creating a dataset which can be used in a geographical information systems (GIS) environment. The climate variables will be the monthly mean daily maximum and minimum temperature and the mean monthly rainfall for each month of the year. Climate data collected over a period of 10 or more years will be used for the study.

1.2.3 Objectives

In order to achieve the aim set out in the previous section the followings steps were followed.

1. Undertake a literature study to examine interpolation techniques and software programs suitable for the three above-mentioned climate components.
2. Identify a suitable interpolation software package which will be appropriate for the study region.
3. Identify suitable input variables for the climate interpolation process.
4. Investigate by means of test runs possible transformations and units (i.e. parameterization) for the variables employed during the interpolation process.
5. Test the accuracy and suitability of the software package and the selected parameterization of the variables by performing statistical analysis on the program output and by extracting a selected set of weather station data and comparing the interpolated results to those of the actual recorded values.
6. Create high-resolution maps by the selected interpolation software package for the three climate components for each month of the year.
7. Compare the results with other available climate data sources.

Figure 1.2 summarizes the research design for this project, describing the steps that were taken and the research methodology followed.

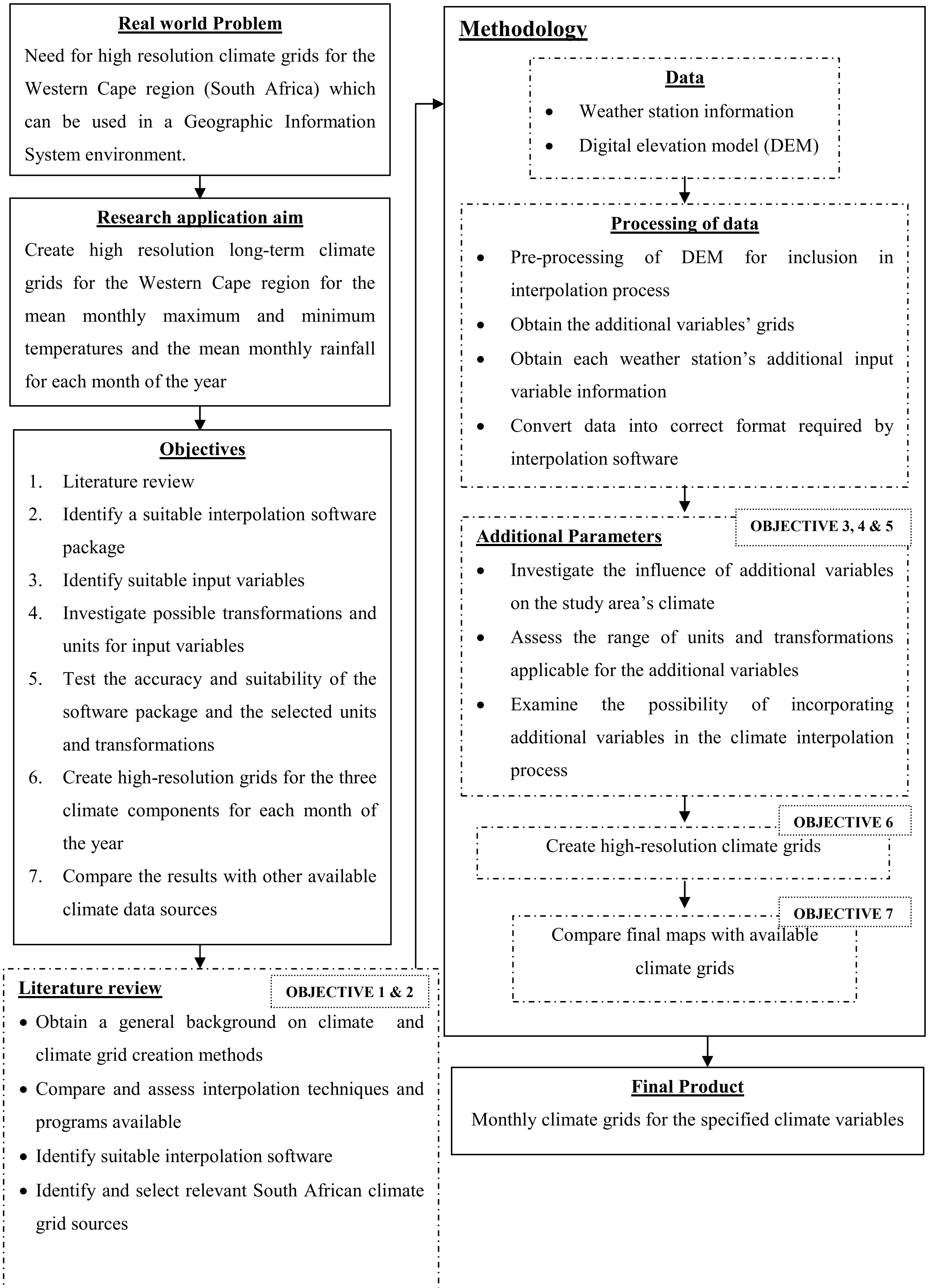


Figure 1.2: Research design

The first objective set out in this study aims at providing a good background on climate interpolation and software programs in order to assess the process of selecting the most suitable software package for the study area. The next section in this report addresses this objective by means of summarizing the available literature providing examples of projects where climate interpolation has been completed and present examples of relevant climate interpolation projects completed in South Africa.

CHAPTER 2: FROM WEATHER STATIONS TO CLIMATE GRIDS

When the climatic conditions of a region are not obtainable from a long-term weather station, spatial interpolation is a tried and tested method by which this problem can be resolved. A variety of interpolation methods have been used as revealed by a comprehensive literature review of the available methods and software packages (Joubert 2004). The salient findings of the literature study are presented next, followed by a discussion of the available climate datasets for Southern Africa.

2.1 INTERPOLATION TECHNIQUES

A useful definition of interpolation is: "The procedure of estimating the value of properties at un-sampled sites within an area covered by sampled points, using the values of properties from those points" (Hartkamp *et al.* 1999: vi). A basic technique of interpolation is to draw boundaries according to the distribution of sampled data points, i.e. one data point located in the centre of a polygon. This technique predicts the attributes of an un-sampled point to be the same as that of its nearest sampled point and is best for data where external factors (such as temperature lapse rate) do not play a role, therefore it calls for more sophisticated techniques.

A review of the literature indicates that most climate interpolation projects have been done with temperature variables (maximum and minimum) and to a lesser extent on rainfall. The studies have been done at various scales and resolutions. The most popular methodology followed in such projects well demonstrated by Kurtzman & Kadmon (1999) - is to identify a range of suitable interpolation techniques and then to compare these techniques with one another using practical examples (i.e. for well defined study areas). From the results obtained, the merits and demerits of each technique were assessed.

A good example of such a project is that of Collins & Boldstad (1996) in which they extended their study to include a range of temporal scales. Eight interpolation techniques were compared across two regions (eastern and western North America) using two temperature variants (maximum and minimum temperature) over three temporal scales (10-year mean, seasonal mean and daily). This study compared weather and climate data interpolation.

The available literature highlights the fact that rainfall interpolation is more complex than temperature interpolation, since a region's rainfall is influenced by more external factors. Given this anomaly, Naoum & Tsanis (2004) have developed a GIS-based decision support system to investigate and select the appropriate interpolation techniques in order to study rainfall's spatial variability. Rainfall interpolation has also received attention from Borga & Vizzaccaro (1997) and Dirks *et al.* (1998) where they have investigated a wide range of interpolation techniques.

In order to understand the advantages and disadvantages of the available interpolation techniques, the above-mentioned examples were used to identify the interpolation techniques most appropriate to the envisaged research. These selected sources, together with geographical information systems (GIS) package help files, are summarized below.

2.1.1 Inverse distance weighting

Inverse distance weighting (IDW) is based on the assumption that the nearby values contribute more to the interpolated values than distant observations – the influence of a known data point is inversely related to the distance from the unknown location that is being estimated (Anderson 2004). The IDW method is deterministic since it uses specific mathematical formulae in order to calculate the weighting function, and in the case of IDW the weighting function is strictly a function of distance. Deterministic methods can be further divided into two groups, namely local and global. Global techniques calculate interpolations using the entire dataset, while local techniques calculate interpolated values from the measured points within neighbourhoods, which are smaller spatial areas within the larger study area (ArcGIS 1999). IDW is classified as a local and "exact" interpolator. An exact interpolator honours the data points on which the interpolation is based, while an inexact interpolator predicts a value that is different from the measured values for the known data points. The main disadvantage of IDW is that the results rely on size of search window and the choice of weighting parameter. Furthermore IDW does not provide any statistical output that can be used for error assessment, and it is sensitive to outliers. IDW works best for evenly distributed data points and is intuitive and efficient to use (Joubert 2004)

2.1.2 Splines

The spline method is analogous with fitting a rubber-sheeted surface through known points using a mathematical function (Anderson 2004). Most sources define it as deterministic, but Hartkamp *et al.* (1999) go further and define it as deterministic with a local stochastic interpolator. Unlike deterministic methods, stochastical methods use statistical criteria to determine weighting factors. Like the IDW interpolation technique, splining is sensitive to outliers, but unlike IDW, the splining function may have different minimum and maximum values than those provided by the dataset.

Splining has the ability to generate sufficiently accurate surfaces from only a few sampled points and it also retains small local features (Joubert 2004).

2.1.3 Kriging

Similar to IDW, kriging uses weighting which assigns more influence to the nearest data points in the interpolating process, but it is not deterministic since it extends the proximity weighting process of IDW (Joubert 2004). Kriging weights for the surrounding measured points are more sophisticated than those of IDW. IDW uses a simple algorithm based on distance, but kriging weights come from a semivariogram developed by looking at the spatial structure of the data (ArcGIS 1999). This means that kriging is a stochastic interpolator since the first step for this method is to determine the statistical relationship between sampled points. The advantages of kriging are the incorporation of variable interdependence (Anderson 2004) and the stochastical methods provide the ability to compute and assess error (NOAA Coastal Services Center 2004). Compared IDW, kriging requires more computing and modelling time, and more input from the user.

2.1.4 Regression

Regression interpolation estimates the relationship between variables so that a given variable can be predicted from one or more other variables. It is classified as an inexact deterministic interpolator which is global, but which can be refined locally (Hartkamp *et al.* 1999). The simplest form of regression is linear regression where it is assumed that the relationship between two variables, the one independent and the other dependant, can be explained by a straight line fit. A more popular form of regression is multiple regression where there can be two or more independent variables. In this case the values of the parameters are estimated by least squares method (minimizing the sum of squares of the residuals). Another regression interpolation method is polynomial regression which is a stochastic interpolation technique (Hartkamp *et al.* 1999). Although regression methods are relatively easy to understand, they have the disadvantage of being affected by extreme values and uneven distribution of observational data points (Joubert 2004).

2.1.5 A South African study

Lynch (1999) has used Arc/Info, a GIS software package to compare five precipitation data interpolation techniques for the winter rainfall area of South Africa. He compared IDW, Schäfer's daily rainfall estimation method (expresses the daily rainfall at each measured site as a ratio of the median monthly precipitation), multiple regression analysis and spline and kriging interpolation methods. Lynch (1999) concluded that the Schäfer methods performed the best when the dataset does not contain the maximum amount (i.e. a specific area within a region may have the highest rainfall, but this is not recorded (no weather station) and is thus not included in the dataset). He warns however that the “GIS user cannot simply throw data into the first available interpolation method and hope for a swift and satisfactory solution” (Lynch 1999: no page). In order to achieve accurate results it is necessary to choose the method most applicable to the data and the region.

Spatial interpolation tools have been integrated within GIS software packages, but these are often implemented in such a way that users have little or no choice in selecting input parameters (Dubois 1998). The interpolation functions provided in GIS packages only allow for using one input variable, sufficient when interpolating a phenomenon whose distribution is only dependant on a single factor, but not adequate when interpolating climate data since a region’s climate is influenced by multiple factors. For this reason, special climate interpolation software packages have been developed and are available for commercial use.

2.2 INTERPOLATION PROGRAMS

Traditionally, methods for mapping climate from point data were done geographically which involved the manual preparation of climate maps. Due to the advent of computers as a common workspace tools, interpolation methods became largely statistical (Climate Source 2000). A more modern approach to interpolation is a combination of human-expert knowledge and statistical methods (Daly *et al.* 2002).

Statistical methods: “[u]se a numerical function, calculated or prescribed, to weight irregularly spaced point data to estimate a regularly spaced prediction grid” (Daly *et al.* 2002: 100). The most popular statistical methods incorporated into statistical packages have been briefly examined in the previous section. Although statistical methods have been accused of lacking a spatial climate knowledge base that experts provide, they have the advantages of being faster, producing more consistent results and having repeatability.

Human-expert methods is a “… modern approach [which] makes use of human experience, expertise and knowledge acquisition capabilities to infer climate patterns from meteorological regimes, physiographic features, biotic characteristics and other information sources” (Daly *et al.* 2002: 100). This method is less rigid and incorporates additional information in the interpolation process, but it is relatively new and still under development so that most of the studies and comparisons of the results of these packages are performed by the creators and programmers of the packages. Human-expert packages require constant upgrading and maintenance, making it difficult to distribute the package commercially.

The following section briefly describes some software packages identified by Joubert (2004) as being the most popular and widely used in the industry and by researchers for a range of climate related projects and data-generating exercises.

2.2.1 PRISM

The parameter-elevation regression on independent slopes model (PRISM) is an example of a knowledge-based model (KBM) used for climate interpolation and is being developed by the Oregon State University. Daly *et al.* (2002) provides a comprehensive description of the model. The statistical component of this model is based on regression and the following climate controlling factors are incorporated in the interpolation process:

- Elevation influence on climate
- Terrain-induced climate transitions
- Coastal effects
- Two-layer atmosphere
- Orographic effect of terrain

PRISM has been used to create climate layers for commercial distribution (Natural Resources Conservation Service 2004), and also to produce digital climate layers that are easily viewed in a variety of software programs, such as GRASS, Arc/Info, and WinGIF. PRISM is compatible with UNIX, DOS/Windows and Macintosh operating systems and recently a “PRISM Tools” extension has been developed for ArcView, a popular GIS program.

2.2.2 ANUSPLIN

The Australian National University, under the direction of Dr MF Hutchinson, was responsible for the creation of ANUSPLIN a statistical package which makes use of thin plate smoothing splines to create continuous surfaces from point data. The package is a popular interpolator which is often used for climate research and for the production of commercial climate maps (Jeffrey *et al.* 2001; Johnson *et al.* 1999). In a comprehensive study by Price *et al.* (2000), ANUSPLIN was compared to another statistical model, ‘gradient plus inverse-distance-squared’ (GIDS), in order to investigate the performance of these programs when interpolating the spatial distribution of temperature and precipitation over complex topography. According to the results, ANUSPLIN is superior to GIDS and more reliable in rough terrains, and it can also be successfully applied to rainfall data.

Until recently ANUSPLIN had only been able to incorporate the effects of latitude, longitude and elevation in the interpolation process, but ANUSPLIN version 4.3, released in 2004, enables researchers to incorporate additional influences, such as aspect and slope.

2.2.3 Neural networks

The development of algorithms and programs for climate data interpolation is an ongoing and dynamic process closely related to development in computer systems and computer science. Often, models and algorithms created for other applications are also applicable to climate interpolation, for example neural networks which are a form of artificial intelligence. Silva (2003: 2) defines artificial neural networks as "adaptable system[s] which can derive relationships between different sets of data from successful cases of training." Silva (2003) identified various types of networks applicable to climate variables, explained how these systems work, and described a selection of training cases in order to provide a better understanding of this new form of interpolation.

In studies by Antonić *et al.* (2001) and Bryan & Adams (2002), neural networks were successfully used for spatial-temporal interpolation of climatic variables over regions with complex terrain. No studies where neural network output was compared to other interpolation methods were found in the available literature.

2.2.4 MTCLIM and Daymet

The mountain climate simulator (MTCLIM) is based on the spatial convolution of a truncated Gaussian weighting filter performed on a set of sparsely located weather stations (Thornton, Running & White 1997). The site's (a specific location) maximum and minimum temperatures are calculated by using temperature lapse rates from another location's (the "base" site) information, while precipitation estimates are based on the daily record of precipitation from the base site and a user-specified ratio of annual total precipitation between the site and the base. The base and the site can be at different elevations, and can have different slope and aspect characteristics.

According to the information website of the Numerical Terradynamic Simulation Group at the University of Montana (2005) School of Forestry, MT-CLIM is limited by the fact that (a) it uses a single base station for observations and (b) the need for the user to specify the temperature and precipitation relationships with elevation. The group developed an expanded version of the MT-CLIM logic, called Daymet, which uses observations from a large number of base stations.

The Daymet U.S. Data Centre (2005) website provides climate data compiled using Daymet for an 18-year daily dataset (1980 to 1997) of temperature, precipitation, humidity and radiation as a continuous surface at a 1km-resolution for the United States as well as a list of relevant Daymet literature published in scientific journals. According to the literature this software program is not a popular interpolation package.

2.2.5 CIAT

This method makes use of a simple interpolation algorithm based on the inverse square of the distance between the five nearest stations and the interpolated point (Chapman, Muńoz & Koch 2005). Due to the fact that this method only uses five points during the process, it relies less on the underlying DEM, than do PRISM and ANUSPLIN. The CIAT method has the advantage of speed and simplicity, but the results can be severely compromised by bad data points. According to Jones (2006, pers com), there is no standard software package for CIAT since the FORTRAN code is rewritten for each project. The Centro Internacional de Agricultura Tropical (CIAT) provides various climate grids for regions across the world produced from the CIAT method, and these are available from the FloraMap website (Centro Internacional de Agricultura Tropical 2001).

The literature available on climate interpolation software packages indicates that climate interpolation projects (at various resolutions), using a range of climate interpolation programs, have been completed. This provides an extensive climate database covering most of the globe. The next section specifically describes the available climate data available for Southern Africa.

2.3 SOUTHERN AFRICA CLIMATE DATABASES

The importance of climate data and its role in various fields of study has been established in the previous chapter. The next section deals with climate databases available for Southern Africa, specifically focussing on datasets that can be used in a GIS environment. Climate data in a GIS-suitable format enables researchers or decision-makers to adapt and analyse the data and to incorporate the data into larger projects.

2.3.1 Climate data produced by climate modelling for Southern Africa

Most often climate modelling performed for Africa is done by non-African institutions. For example, the United Kingdom's Department for Environment, Food and Rural Affairs subcontracted researchers from the Hadley Centre at the University of East Anglia (Climatic Research Unit) to perform simulations of the present-day and future climate over Southern Africa using HadAM3H a general circulation model (GCM). The climate grids produced were considered to be "high" resolution at ≈ 150km (Hudson & Jones 2002).

In South Africa, the University of Cape Town's Climate Systems Analysis Group is a leader in the field of climate modelling and together with a few other research institutions the group is responsible for most of the publications and data produced on the subject of climate modelling, mainly using GCMs (Crane & Hewitson 2003; Hewitson 2003; Hewitson & Crane 1996; Wilby *et al.*1998). The group's website provides a list of publications produced by the institution over the past 15 years (Climate Systems Analysis Group 2006). Other climate modelling projects have been done by the University of Pretoria's Geography, Geoinformatics and Meteorology department. Their Laboratory for Research in Atmospheric modelling is responsible for running projects such as numerical weather prediction, regional wind simulation and seasonal forecasting (Engelbrecht, Rautenbach, MecGregor & Katzfey 2002; Engelbrecht 2005).

Unfortunately the coarse resolution of GCMs restricts their usefulness for the study of a smaller area's atmospheric processes. Important sub-grid scale features such as clouds and thunderstorms and the study of the influence of topography are often not noticed (Wilby, Dawson & Barrow 2002). In addition to the downscaling of GCM's in order to obtain higher resolution data, regional climate models have been developed and are known as mesoscale climate models. The resolution of these models enables the simulation of smaller scale atmospheric processes. It is even possible to use regional climate models and GCM's collectively to study an area's climate. An example of such a model is the regional atmospheric modelling system (RAMS), which requires high computational power and memory. The South African Agricultural Research Council (ARC) employed RAMS to monitor sea-breeze mechanisms and their effects on the wine-producing areas in the Western Cape (Bonnardot, Planchon & Cautenet (2005).

Olwoch *et al.* (2003) used available climate datasets, at a resolution of 60km x 60km, for predictive species modelling in order to compare the efficiency of the various datasets. The interpolated datasets were provided by the Australian National University's Centre for Resource and

Environmental Studies (CRES), the Climate Research Unit (CRU) from the University of East Anglia in Norwich, while the modelled data was created using the Commonwealth Scientific and Industrial Research Organization's (CSIRO) Division of Atmospheric Research limited-area model (DARLAM). Using data obtained from climate modelling enable researchers to incorporate future climate scenarios.

It is apparent from climate research literature that climate modelling and climate prediction projects receive much attention from research institutions throughout the world. The resolution at which climate data is produced is usually coarse since finer scale projects require more computing power and special computer networks. Regional climate models (RCMs) are typically integrated over large domains (often covering entire continents) since the purpose of the simulations is to simulate the synoptic and meso-scale climate of the region over a certain time period. Present day computational power allows such simulations to be performed at a horizontal resolution of about 50km (Engelbrecht 2007, pers com). Climate models also require good data subsets (i.e. soil, land cover, sea surface temperatures) and therefore climate interpolation techniques still remain the best way to obtain climatic information for a region when the climate variables required are only temperature and rainfall. Due to the relatively straightforward techniques used for climate interpolation, it is possible to perform such processes on modern personal computers to produce climate grids with high resolutions suitable for input in GIS packages.

2.3.2 Completed Southern African interpolation projects

The literature review identified a couple of climate interpolation projects done by diverse institutions yet again often not located in Africa. The most complete and relevant example of such a project was requested by the International Laboratory for Research into Animal Diseases (ILRAD) and the International Institute for Tropical Agriculture in order to address pressing problems associated with improving food production, managing pests and diseases, and preserving biodiversity in Africa. The Australian National University's Centre for Resource and Environmental Studies was contracted to develop a topographic and climate database for the African continent. For the second part of the task, ANUSPLIN was used to construct the climate dataset (for the monthly mean values of rainfall, daily minimum temperature and daily maximum temperature). The outcome was a gridded database with a resolution of 5km x 5km including standard errors of about 0.5^{o}C for monthly mean temperatures and errors of approximately 10-30% for monthly mean precipitation (Hutchinson *et al.* 1996).

Climate data creation by means of interpolation has been completed for the entire world including Southern Africa. Hijmans *et al.* (2005) used ANUSPLIN to create a climate database at a resolution of 1km from records available for 1950 to 2000. It is referred to as the WorldCLIM dataset. The results were compared to those produced by Daymet and PRISM, and recommendations were made regarding regional interpolation projects which require finer resolution.

In the field of climate interpolation for Southern Africa, Roland Schulze and Steven Lynch in the Department of Agricultural Engineering at the University of Natal have done significant research over the past decades. The most recent temperature and rainfall databases created for Southern Africa were created by them in association with the South Africa's Water Research Commission (WRC).

Schulze was responsible for producing daily maximum and minimum temperature maps for Southern Africa to a one arc minute spatial resolution (≈1.6km) with a 50-year base period (Schulze & Maharaj 2004). This publication provides a detailed description of similar studies performed in Southern Africa and assesses how factors influencing a region's temperature have been incorporated in these projects. A major problem encountered was that most stations had missing values or did not extend over the entire time period. Schulze & Maharaj (2004) provide an explanation of how this problem was addressed. During this interpolation process the regional and intra-annual lapse rates for temperature where determined and used.

Another project aimed at developing a raster database of annual, monthly and daily rainfall for Southern Africa in order to provide data for various hydrological simulation models using GIS and other techniques was initiated by Dr. Steven Lynch. The study highlights the disturbing phenomenon of a decline in the numbers of active rain gauges in South Africa, which results in many weather stations not having complete datasets. In order to compensate for missing rainfall information, infilling of rainfall datasets by means of various techniques was investigated. This called for an examination of available interpolation techniques for estimating rainfall at ungauged locations. The eventual method chosen was geographically weighted regression (GWR) approach, a method which requires South Africa to be divided into rectangular regions in order to manually select a regression equation for each region. The residuals (i.e. the differences between the observed and estimated values) were interpolated onto a one-minute raster grid using IDW. "Inverse distance weighting technique was identified as the technique of choice when the density of point data is sufficient and the variation in parameter to be converted to raster is not to complex" (Lynch 2003: 27). This raster was then added to the one-minute rainfall raster in order to calculate the final product. Lynch's (2003) report provides a detailed exposition of the methodology followed, together with explanations of other methods available, and a list of literature sources referring to similar projects completed for South Africa.

2.4 CONCLUSION

There is a large body of climate related scientific and research literature available in scientific journals, popular magazines and the Internet. The Internet is a valuable tool for locating climate research groups and their home pages, which in turn provide detailed descriptions of the work done by these institutions and also offer links to other institutions and relevant scientific sources. This is a useful starting point when looking for climate data and investigating interpolation techniques. The literature consulted provided useful examples of methodologies and techniques for undertaking an interpolation analysis project.

The literature applicable to South Africa is mostly concerned with climate modelling and climate change prediction analysis. The only study in which a specialised climate interpolation software package was used and analysed was on a continental/sub-continental scale, where the resolution of the final product is too coarse for use in projects on a regional scale (Hutchinson *et al.* 1996). This is also true for interpolation projects performed for South Africa, where the databases created by the WRC have resolutions of 1 arc minute (≈1.6km x 1.6km). This is not sufficient for analysis and simulation projects of areas where the phenomena under investigation have a spatial extent of hundreds of square metres.

In the Western Cape region where the topography is extremely complex, a climate database with a spatial resolution of about 100 metres is more applicable and of more use to researchers and decision-makers. No such project has yet been attempted for climate variables (maximum and minimum temperature and rainfall) for the Western Cape Region. This research aims at filling this lacuna. The following section describes the methodology followed to accomplish this aim.

CHAPTER 3: METHODOLOGY

This chapter presents the methodology followed to complete the objectives set out during the project description. Firstly it is necessary to describe the reason for selecting the study area. Following this, a brief overview of the selected interpolation package is provided. Once the interpolation package is described the data requirements are addressed. The next step is to investigate how the selected factors, in addition to co-ordinates and elevation, influence the climate of the study area. Once the additional variables have been selected it is crucial to explore suitable transformations of and units for the variables incorporated in the interpolation process. The final step is to create climate grids which in turn can be compared to existing climate grids.

3.1 THE STUDY AREA

An understanding of a region's topography, its environmental and social issues and in this case, the climatic conditions, is of utmost importance when selecting a study area for an exploratory investigation of interpolation techniques for creating high-resolution climate grids in weather station sparse areas. The reasons for selecting the study area are set out in the following section.

3.1.1 Selection of the study area

In order to investigate the aptness of creating climate grids using a selected climate interpolation software package, the Western Cape province, situated at the southwestern part of South Africa, was selected for a number of reasons:

- The Cape Floristic Region (CFR) is one of the five floral kingdoms recognised in the world. It is the smallest and the richest floral kingdom on the planet, having the highest known concentration of plant species: 1 300 per 10 000km^2. It is important to preserve this biome since roughly 70% of the plant species are endemic (Maneveldt 2005). The CFR is under threat due to the spread of alien plants, too frequent veld fires and the development of housing estates and farms. Consequently the protection and management of this biome is of utmost importance, therefore requiring accurate high-resolution climate datasets in order to look after and manage the biome and to simulate the effects of future climate change on the various species.
- The Western Cape region is South Africa's main wine-producing region and it is also an important fruit and grain production area. Predicted climate changes and their effects on the agricultural sector of the Western Cape will have major implications for the region's

economy. Climate information and the simulation of future scenarios receive plenty of attention from researchers, but in order to make decisions where areas can be influenced on a small scale; high resolution climate grids are required (Joubert 2000; Hudson & Jones 2002; Olwoch *et al.* 2003).

- The region is already facing major water shortages which will increase in severity in the near future. Consequently decision-makers and scientists are looking at alternative sources and calling attention to management schemes and research which will require detailed climate information for hydrological modelling and simulation projects (New 2002). Thus, when defining the extent of the region for which climate grids need to be produced, all the main rivers and catchments were identified and included in the study region.

Figure 3.1 shows the size of the study area. The intent was to include all the major Western Cape river catchments, thus incorporating areas outside the Western Cape. The interpolation process requires grids in a rectangular format therefore the extent of the study area includes parts of the Northern and Eastern Cape.

Figure 3.1: Main rivers and river catchments in the Western Cape

Unlike the rest of Southern Africa, part of the Western Cape receives the majority of its rainfall during the winter months. The region also has a number of topographical features which have considerable influence on its climate. The next section provides a brief overview of the region's climate and topographical features.

3.1.2 Description of the study area

The Western Cape province situated in the southwestern region of South Africa, covers an area of 129 370km^2 (Winter 2002) and is bordered seaward by the Indian Ocean in the south and the Atlantic Ocean in the west. The northern and eastern regions are bordered by other South African provinces (Figure 3.2).

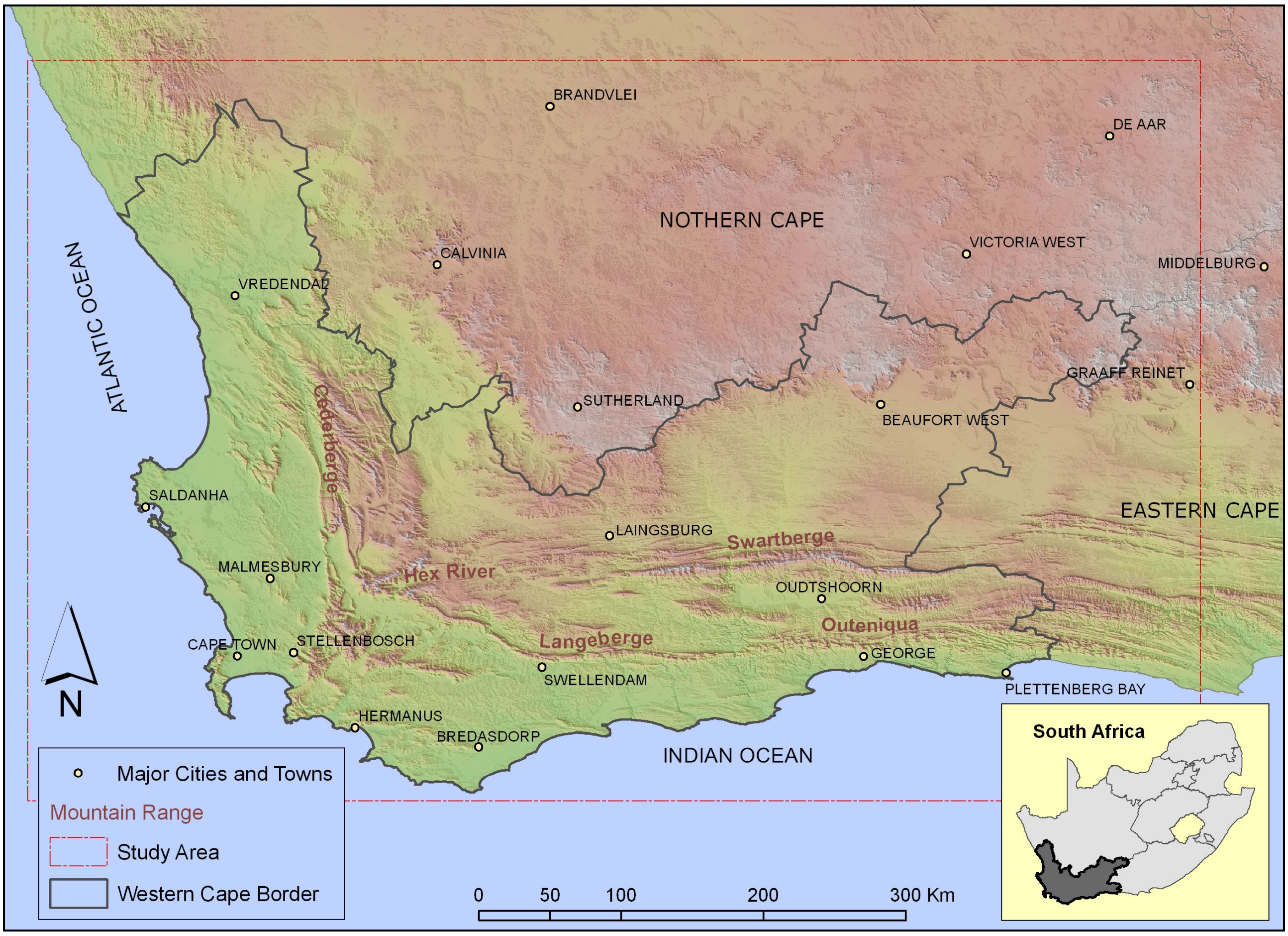


Figure 3.2: Topography of the Western Cape region

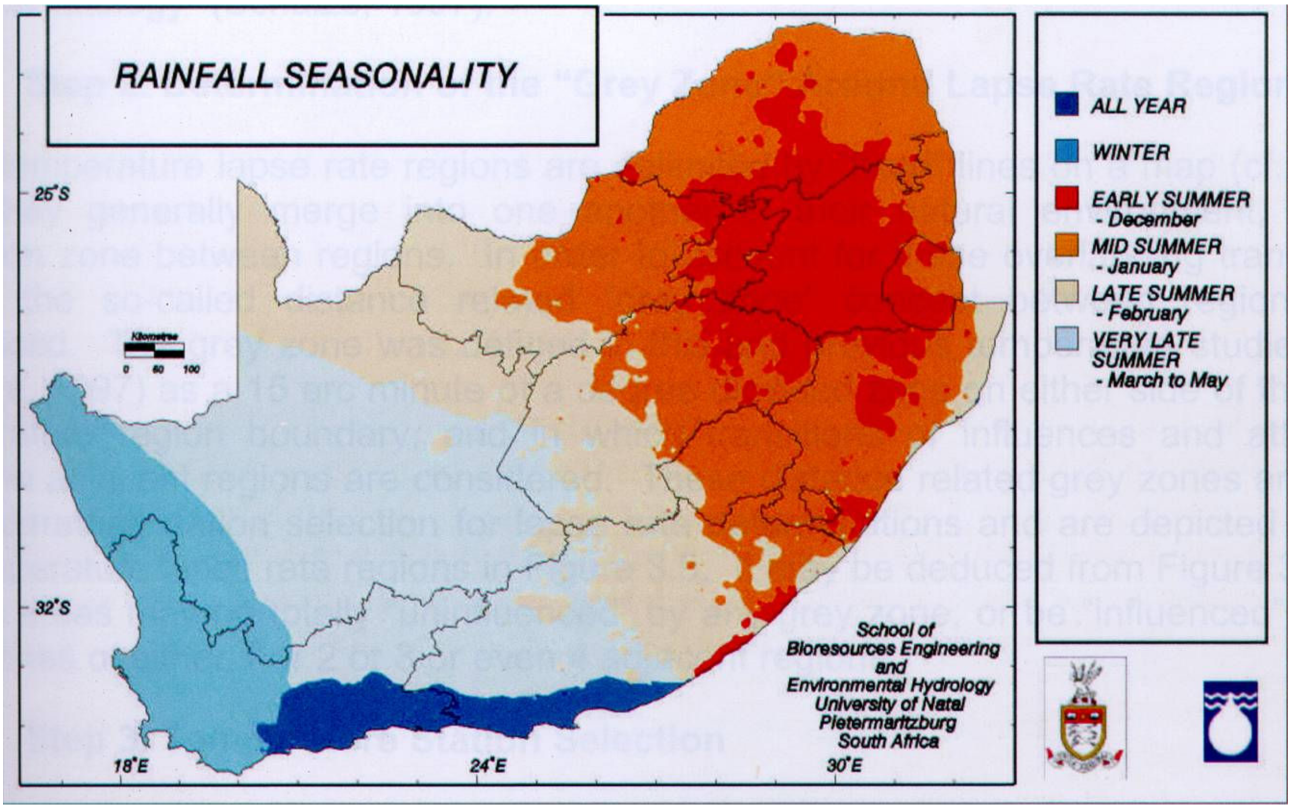
3.1.2.1 Topography

The study area has a complex topography ranging from coastal plains (along side the oceans) to the central plateau of the interior, separated by a series of complex mountain ranges and valleys. The

Cape Fold Belt essentially dominates the topography of the Western Cape consisting of mountain ranges formed in an L-shape, running north-south and east-west (Department of Environmental Affairs and Tourism 2005).

3.1.2.2 Climate

The climate regions covered by this study are diverse, ranging from the winter rainfall region of the southwestern and western Cape to the inland late summer rainfall region of the Karoo. Figure 3.3 shows the rainfall regions found in South Africa as identified by Schulze (1997). The brief descriptions of the major climate regions in the study area and their topography given in South African Weather Bureau (1996) and Van Zyl (2003) are summarized below.



Source: Schulze (1997: 40)

Figure 3.3: Rainfall seasonality of South Africa

"Mediterranean": winter rainfall

The main physical features of this region are the complex system of folded mountains, the coastal plains and the cold Benguela ocean current in the Atlantic Ocean which borders the region in the west. The region experiences warm to hot summers and mild to cool winters, and receives the majority of its rainfall during the winter months from May to August. Due to the northward displacement of high pressure systems during the winter months, the westerlies and polar air masses have a significant influence on the region's climate. The precipitation during the winter is prefrontal rain and postfrontal showers, while the rainfall variability is great from place to place due to the heavy orographic rains caused by the mountain ranges.

The South Coast: all-year rainfall

The South Coast region experiences rainfall throughout the year, but with a definite peak during the spring and autumn months. Temperature wise the region is warm and humid during the summer, while the winters are mild. A marked decrease in temperature is experienced with an increase of altitude. The reason for this is the influence of the Indian Ocean which borders the region in the south and the southern Cape mountain ranges in the northern parts which form a natural divide between the South Coast and Karoo climate regions.

Karoo: rain in the very late summer

The Karoo region is separated from the "Mediterranean" and South Coast areas by mountains which create rain shadows. Situated on the inland plateau of Southern Africa, it is a semi-arid region with low and unreliable rainfall. Temperatures vary dramatically from winter to summer, with average temperatures of over 30^{o}C during the summer and freezing nights during the winters.

Due to the complexity of the study area's topography and other factors influencing its climate, it is important to select an interpolation package suitable for the interpolation of the selected climate variables for this exceptional region.

3.2 SELECTION OF AN INTERPOLATION SOFTWARE PACKAGE

The literature survey affirmed PRISM and ANUSPLIN to be the most popular and widely used interpolation software packages. PRISM, a program which combines human-expert knowledge and statistical procedures to perform interpolation, is not commercially available since it is still undergoing modification and requires significant technical support (Daly 2005, pers com).

ANUSPLIN, developed by the Australian National University, is commercially available and can be obtained from Centre for Resource and Environmental Studies (2004). Sections 3.2.1 and 3.22 provide a more in-depth description of ANUSPLIN, identifying its strengths and limitations, explaining why it is regarded as a suitable interpolator for the study region and provides examples of projects in which it has been used.

3.2.1 ANUSPLIN as the chosen interpolator

The first objective of the study was to identify examples and projects across the world where ANUSPLIN was selected as a climate grid creator. The versatility of ANUSPLIN is demonstrated in the following paragraphs which summarize projects in which ANUSPLIN was used. The projects, which range in extent from regional to global scale, illustrate the confidence climate interpolators and climate data users have in the suitability of this interpolation package.

Hijmans *et al.* (2005) performed an extensive study on the performance of ANUSPLIN on a global scale, but they suggest that better results could be obtained when interpolation is carried out on a regional scale and merged afterwards. Continental-scale projects for Australia (Jakob, Taylor & Xuereb 2005; Jefrey e*t al.* 2001), Africa (Hutchinson *et al.* 1996) and North America (McKenney 2000) have been completed. ANUSPLIN has been used in regional projects in Canada (Price *et al.* 2000), Madagascar (Chapman 2003), China, Thailand, Vietnam, Laos, Cambodia and the Malay Peninsula (Zuo *et al.*1996) and Guyana (Funk & Richardson 2002).

Literature in which various interpolation packages are compared is available and provides insights into the performance and limitation of the software and also gives an indication of the most suitable package for a specific type of region. According to Hartkamp *et al.* (1999), Hijmans *et al.* (2005) and Price *et al.* (2000) ANUSPLIN performed well in these comparative tests.

Data produced by ANUPSLIN are often created for use as primary or secondary data components in larger projects where climate information forms part of the modelling and/or simulation processes. The possibility of using ANUSPLIN datasets was examined by Wilson, Mitasova & Wright (2000) where various combinations of GIS and simulation models were used for the advancement of knowledge about water resource assessment and management.

ANUSPLIN is also a popular interpolator for projects where environmental modelling relies heavily on accurate climate data (Chapman, Muńoz & Koch 2005; Murray 1998). In studies where predicted climate change is investigated, ANUSPLIN data has been used as baseline data (Bootsman, Anderson & Gameda 2004) or to interpolate climate grids with a high resolution from the coarse resolution data produced by general circulation models (Price *et al.* 2004).

3.2.2 ANUSPLIN: technical description

A brief description of the various programs (and their functions) that makes up the ANUSPLIN package is attached in Appendix A. In short, ANUSPLIN calculates and optimises thin plate smoothing splines fitted to datasets distributed across an unlimited number of climate station locations (Hutchinson 2004).

As mentioned before (paragraph 2.1.1) the spline method can be seen of as fitting a rubber-sheeted surface through the known points using a mathematical function. In fitting surfaces to data points, thin plate smoothing splines determine an optimal trade-off between accuracy of fit and surface smoothness by minimizing the generalized cross-validation (GCV) index. The GCV value is an estimate of the interpolation error obtained by removing each data point in turn and fitting a spline surface to the remaining data to see how well each omitted point can be predicted (Hutchinson *et al.* 1996).

ANUSPLIN treats climate data as variables dependent upon spatial dimensions (latitude, longitude and elevation), and potentially dependent on other factors such as slope and aspect, and spatially varying rainfall/temperature distributions (Price *et al.* 2004). A digital elevation model (DEM) is a prerequisite for developing data layers of topographically influenced climatic variables (Barringer & Lilburne 2000), but additional data layers such as slope and aspect can be introduced. The climate grids are calculated by applying the fitted smoothing spline surface to an underlying DEM (and additional variable grids) using the LAPGRD program (Hutchinson *et al.* 1996). Appendix B shows the data flow during the creation of climate grids.

ANUSPLIN inputs include the numbers of dependent, independent and covariate spline variables, the lower and upper limits for each independent variable, optimal transformation of each variable, the order of derivative to be minimized and the number of surfaces (Hutchinson 2004). The user manual supplied with the software package provides comprehensive technical description and refers to relevant literature sources on the mathematical functions on which the processes are based. The manual is available online from the Centre for Resource and Environmental Studies (2004) website.

3.3 DESCRIPTION OF THE DATA

In order to use ANUSPLIN to create climate grids, various data sources are required. This section provides brief descriptions of the data used in this project.

3.3.1 Digital elevation model

The DEM consists of a regular array of elevation information, normally squares, over a terrain. The sizes of these squares indicate the resolution of the DEM and may vary. The Shuttle Radar Topography Mission (STRM) DEM used for this project was generated using the C- and X-bands from the data obtained by the Shuttle Radar Topography Mission (SRTM) which successfully collected Interferometric Synthetic Aperture Radar (IFSAR) data over the Earth between 60^{o} north and 56^{o} south latitudes (United States Geological Survey 2006). The STRM DEM grids have a resolution of 3 arc minutes (≈ 90m x 90m), are in geographical co-ordinates, with elevation in metres and the World Geodetic System 1984 (WGS84) is used as horizontal and vertical datums (Rabus *et al.* 2003). This DEM dataset is readily available from the USGS website and was selected due to its high accuracy level. Finer resolution DEMs have been created for the study area using 20m contour lines from topographical maps, but this is an estimation of elevation since it is created from secondary data, while the STRM DEMs are calculated from a satellite image. The elevation of each pixel is therefore far more accurate since it is obtained from a primary source.

3.3.2 Weather stations

The monthly mean data for both the daily maximum and minimum temperatures and rainfall were supplied by the South African Weather Services (SAWS) (64 stations) and the Agriculture Research Council's (ARC) (61 stations) long-term weather stations situated in and around the study region. Stations situated outside the study area is also used since it enhances the accuracy of the spline function calculations. The underlying DEM determines the extent the climate grids produced by ANUSPLIN.

Figure 3.4 shows the location of the weather stations in and around the study area. Note that some regions, for example the southwestern region, have a greater density of stations compared to the number of stations found in the northern and northeastern parts of the study area. ANUSPLIN does not allow stations to be situated at the same location (or within close proximity), but since two datasets are used this does occur. In such cases the station with the smallest time range was removed from the dataset.

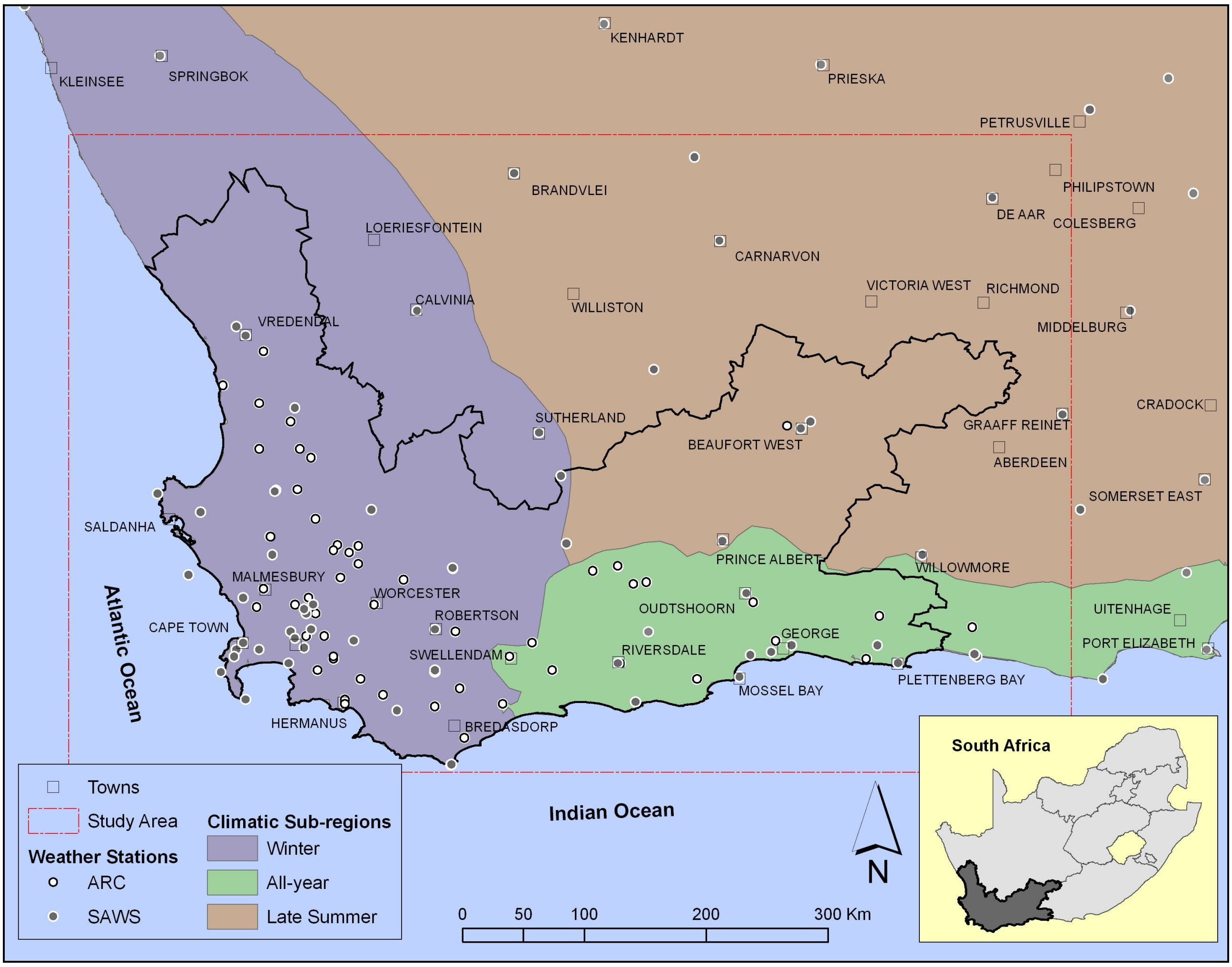


Figure 3.4: Available long-term weather stations for the study region and surrounding areas

The weather station data collection period had an average time range of thirty years, but may vary between 10 and 35 years. The data provided by SAWS and ARC specify each weather station's location (latitude and longitude in geographical co-ordinates) and elevation. If additional data is required for a weather station (i.e. distance to nearest ocean), this information must be calculated using suitable GIS tools.

3.4 ADDITIONAL VARIABLE SELECTION

ANUSPLIN requires the latitude, longitude and elevation for each station as independent input variables before the spline function can be calculated. ANUSPLIN version 4.3, which was made commercially available in April 2004, makes provision for the use of additional input variables such as topography (slope and aspect) and distance to nearest large water bodies. For the purpose of this study it was decided to first investigate the effect of these factors on the study area's climate in order to attain whether it is necessary to incorporate their influence when creating climate grids.

3.4.1 Topography (slope and aspect)

Owing to the recency of ANUSPLIN version 4.3 examples of commercial or research projects where additional variables (other than latitude, longitude and elevation) have been incorporated in the interpolation process are scarce. The only relevant work is by Hutchinson (1998b) (the creator of the ANUSPLIN package) in which an analysis of topographic dependence performed during the interpolation of rainfall data for a single day is reported. The effect topographical features such as slope and aspect have on rainfall was modelled by using two components, namely p and q, as predictor variables, namely

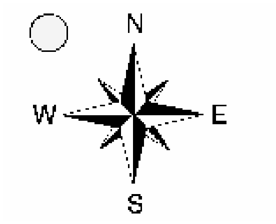
$$p = cos\,(\alpha)\,sin\,(\theta)$$

$$q = sin\,(\alpha)\,sin\,(\theta)$$

where α is the aspect angle and θ is the angle of the slope itself. Using both p and q in the rainfall grid interpolation process permits the incorporation of topographical effects and allows the direction of these effects to be determined without reference to the prevailing wind direction (Hutchinson 1998b).

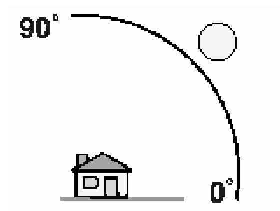
Since Hutchinson's (1998b) analysis was done for one specific day, other methods still have to be investigated for different synoptic patterns, especially frontal rainfall. His project confirmed the importance of incorporating additional topographic dependencies, but he admits that there is room for further studies in incorporating such factors using different variable units and transformations. Furthermore, this method requires two dataset grids for the two separate variables to represent topography. This is cumbersome as it decreases the computation speed of the interpolation process and increases the size of the final climate grids, especially, when high-resolution output grids are required.

In order to combine aspect and slope in one variable, the study area's DEM was used to create a hillshade grid. Hillshading, also known as relief-shading, reproduces the looks of an area when simulating the interaction of sunlight with the area's surface features. The hillshade function determines a theoretical illumination for each cell in a grid by setting the position of a hypothetical light source and calculating the cell's relative radiation value in relation to neighbouring cells. There are four factors which are taken into consideration when calculating a cell's relative radiance, namely the azimuth (i.e. the angular direction of the sun, measured clockwise, from 0^{o} to 360^{o} as in Figure 3.5a), the altitude (illustrated in Figure 3.5b, which is the angle of the illumination source above the horizon, from 0^{o} (on the horizon) to 90^{o} (overhead)), slope (measured from 0^{o} to 90^{o}), and aspect (0^{o} to 360^{o}).



Source: ArcGIS 1999: no page

Figure 3.5a: Azimuth: angular direction of sun in degrees



Source: ArcGIS 1999: no page

Figure 3.5b: Slope or angle of the illumination source above the horizon in degrees

The algorithm used to compute a hillshade value for each cell is

$$R_f = \cos(A_f - A_s)\sin(H_f)\cos(H_s) + \cos(H_f)\sin(H_s)$$

with R_f the relative radiance of a raster cell, A_f the aspect of the cell, A_s the sun's azimuth, H_f the cell's slope while H_s is the sun's altitude. R_f ranges in value from 0 too1 and is multiplied by a constant 255 to obtain the *illumination value* (Chang 2006).

The hillshade function calculates whether a cell falls in a shadow (ArcGIS 1999). Figure 3.6 is an example of hillshade calculated for the southwestern part of the Western Cape. The lighter the colour of the hillshade map, the more directly the slope is exposed to the 'external light source' and the higher the illumination value is.

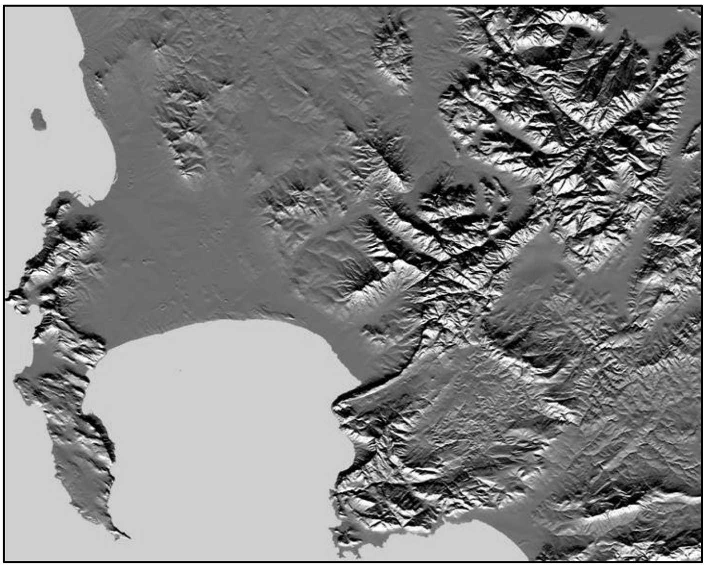


Figure 3.6: Hillshade with an azimuth of 180 degrees and an altitude of 45 degrees

A major part of the Western Cape's rainfall is associated with prefrontal rain and postfrontal showers, usually linked with polar frontal systems developing into depressions approximately 200km southwest of Cape Town (South African Weather Bureau 1996). These cold fronts or cyclones are associated with clockwise wind motion and are prominent features during the winter months.

Appendixes C1, C2, C3 and C4 illustrate the major wind directions experienced throughout South Africa for January, April, July and October respectively based on long term wind monitoring station data. The dominant wind direction for the *southwestern part* of study region is southerly during the summer, while the winter is dominated by northwesterly to northerly winds, usually associated with the passing cold fronts. The *South Coast region's* prevailing winds have a tendency to blow with an off-shore component in winter and an on-shore component during summer. The *northern interior* region experiences easterly winds during the winter and westerly winds during summer, while for the *eastern interior* southerly to southwesterly winds dominate during summer and northerly to northeasterly winds prevail during winter (Kruger 2002). It is apparent that the dominant winds not only change from season to season, but also from one region to another in the study area. In general, the study area is dominated by northerly or southerly winds depending on the season.

This pattern dictated that two hillshade grids be constructed from the study area's DEM. The first hillshade grid has an azimuth of 315^{o} (from a northwestern direction) and the second using an azimuth from a southern direction (180^{o}), both with an angle of the illumination of 45^{o}. This horizontal component was used as the default and further studies will have to be done using different angles in order to obtain the optimum angle. Figure 3.6 is an extract of the hillshade grid calculated for part of the study region using an azimuth of 180^o.

The hillshade component indicates whether a specific cell is situated on the side of a mountain where orographic rainfall occurs or whether it is on the rain shadow side. Cells that are in the shadow of another cell are coded as 0, while the rest of the cells are coded between 1 and 255. High values indicate a high illumination value and appear white in Figure 3.6. For both hillshade grids created, the minimum value is 0, indicating that some cells fall within a shadow of another cell when using an illumination angle of 45^o.

3.4.2 Distance to large water bodies

The two large oceans are prominent in the topography of the study region. The proximity to such large water bodies has a marked effect on the region's climate since water bodies warm more slowly and loose heat less quickly than land masses. This has a moderating influence on areas close to the water, causing them to have less of a difference in maximum and minimum temperatures (Lutgens & Tarbuck 1998). For the purpose of this study only the oceans were considered to be large water bodies since the influence of dams are on a micro-scale. Too few stations are located near the dams to provide information on the influence of dams on the surrounding region's climate.

The weather station data provided by SAWS and ARC includes each station's latitude, longitude and elevation, but no additional data. Therefore it was necessary to calculate each station's distance from the closest ocean using the distance function of ArcView (version 3.3). This tool calculates each weather station's distance to nearest ocean (in the specified units). The next step will be to investigate the correlation between a weather station's illumination value and its distance to large water body with its climate data in order to ascertain the necessity of incorporating these factors into the interpolation process.

3.5 CORRELATION ANALYSES

Two additional variables, namely distance to nearest ocean and topography (the latter represented as hillshade – illumination value), have been identified as factors which could affect the study area's climate. To investigate whether these factors do influence the region's climate, and if so, how strongly, correlation analyses were performed on the location of the weather station and the climate recorded at a weather station using the STATISTCA (version 7.1) statistical package.

According to Schulze & Maharaj (2004) the study region comprises various sub-climatic and rainfall areas. Consequently correlation analyses were done for the three climatic sub-regions identified in the study area (Figure 3.3), namely "Medditerranean" (83 stations), South Coast (28 stations) and Karroo (14 stations). Correlation analyses were also performed for the study region as a whole.

Table 3.1, 3.2 and 3.3 give the results of correlation analysis. The Standard Pearson's coefficient (r) provides an index of the correlation and can be either positive (direct) or negative (indirect). The closer the r value is to + 1.0 or – 1.0 the higher the correlation between the two variables. The coefficient of determination (r^2) states the amount of common variation between the two variables and is expressed in percentage (r^2 x 100) in the tables below. When the p-level (statistical significance – represents a decreasing index of the reliability of a result) is greater than 0.05 the correlation between the two variables is insignificant and can be disregarded (STATISTICA 2005). The r and r^2 values with the appropriate p-level are indicated in the grey and are discussed in the following sections.

3.5.1 Correlation with topography

Table 3.1 gives the correlation coefficients from which the relationship between a weather station's topography (aspect and slope) and the climatic conditions experienced at that station can be gauged. Aspect and slope is represented by the illumination value calculated with an azimuth from a northwestern direction. For precipitation a positive r-value indicates that the higher the illumination value, the higher the rainfall experienced while a negative coefficient for temperature indicates that a station with a low illumination value has higher temperatures.

Table 3.1: Standard Pearson's coefficient (r) and the coefficient of determination (r^2) for topography (illumination value with a 315° azimuth) and rainfall and temperature

		Karoo Late Summer		**South Coast All-year**		**Mediterranean Winter**		**All the Stations**	
		r	**r^2 x 100**	**r**	**r^2 x 100**	**r**	**r^2 x 100**	**r**	**r^2 x 100**
Rainfall	January	-0.15	2.21%	-0.14	2.07%	-0.08	0.63%	-0.16	2.53%
	February	-0.06	0.32%	-0.15	2.24%	-0.23	5.48%	-0.17	2.98%
	March	-0.23	5.29%	-0.09	0.89%	-0.16	2.49%	-0.15	2.38%
	April	-0.05	0.25%	-0.01	0.00%	-0.10	1.07%	-0.15	2.28%
	May	0.01	0.00%	0.13	1.59%	-0.03	0.10%	-0.06	0.34%
	June	-0.16	2.66%	-0.09	0.90%	-0.04	0.19%	-0.09	0.76%
	July	-0.02	0.03%	-0.02	0.05%	-0.02	0.04%	-0.07	0.46%
	August	-0.06	0.32%	0.09	0.76%	-0.03	0.09%	-0.07	0.54%
	September	-0.15	2.16%	-0.15	2.32%	-0.03	0.12%	-0.16	2.46%
	October	0.22	4.76%	-0.14	2.04%	-0.13	1.67%	-0.20	4.12%
	November	-0.11	1.27%	-0.07	0.44%	-0.04	0.20%	-0.15	2.24%
	December	0.32	10.05%	-0.22	4.74%	-0.20	4.06%	-0.21	4.52%
	Annual	*-0.04*	*2.44%*	*-0.07*	*1.50%*	-0.09	*1.35%*	*-0.14*	*2.13%*
Maximum Temperature	January	0.03	0.07%	-0.16	2.66%	-0.09	0.82%	0.04	0.14%
	February	0.09	0.84%	-0.15	2.38%	-0.06	0.39%	0.04	0.16%
	March	0.10	0.96%	-0.12	1.38%	-0.08	0.71%	0.02	0.06%
	April	0.14	1.84%	-0.12	1.39%	-0.12	1.43%	-0.04	0.15%
	May	0.12	1.37%	-0.20	4.15%	-0.13	1.70%	-0.12	1.39%
	June	0.09	0.73%	-0.04	0.18%	-0.16	2.60%	-0.15	2.24%
	July	-0.03	0.10%	-0.04	0.16%	-0.14	1.88%	-0.13	1.71%
	August	0.02	0.04%	-0.10	0.98%	-0.14	1.99%	-0.07	0.51%
	September	0.00	0.00%	-0.13	1.71%	-0.10	0.96%	0.01	0.00%
	October	0.07	0.52%	-0.14	1.95%	-0.08	0.57%	0.04	0.17%
	November	-0.04	0.12%	-0.16	2.68%	-0.08	0.65%	0.03	0.11%
	December	0.01	0.01%	-0.14	1.98%	-0.07	0.52%	0.05	0.26%
	Annual	*0.05*	*0.55%*	*-0.13*	*1.80%*	-0.10	*1.81%*	*-0.02*	*0.57%*
Minimum Temperature	January	0.29	8.21%	0.02	0.03%	-0.22	4.66%	-0.06	0.36%
	February	0.36	12.79%	-0.02	0.04%	-0.20	4.03%	-0.06	0.33%
	March	0.34	11.59%	0.04	0.12%	-0.17	2.98%	-0.05	0.30%
	April	0.37	13.45%	0.06	0.38%	-0.11	1.26%	-0.06	0.38%
	May	0.34	11.58%	0.06	0.42%	-0.03	0.08%	-0.05	0.25%
	June	0.26	6.85%	0.09	0.81%	0.02	0.04%	-0.04	0.19%
	July	0.29	8.46%	0.09	0.85%	0.03	0.12%	-0.03	0.11%
	August	0.35	12.28%	0.08	0.68%	0.01	0.00%	-0.04	0.13%
	September	0.35	12.26%	0.08	0.62%	-0.04	0.19%	-0.03	0.08%
	October	0.38	14.37%	0.03	0.09%	-0.10	1.10%	-0.04	0.18%
	November	0.31	9.35%	0.01	0.01%	-0.19	3.71%	-0.08	0.66%
	December	0.35	12.36%	0.02	0.06%	-0.19	3.69%	-0.04	0.18%
	Annual	*0.33*	*11.13%*	*0.05*	*0.34%*	-0.10	*1.2%*	*-0.05*	*0.26%*

Note: Values where $p \leq 0.05$ are printed in grey blocks.

The correlation analysis detailed in Table 3.1 indicates that except for October and December, the observed correlation coefficients between the two variables are not an accurate indication of the relationship between the variables since the *p*-values are all higher than 0.05. Slopes facing north receive less rain compared to south-facing slopes for October and December, indicating that rainfall during these months is associated with on-shore winds. A low coefficient of determination indicates that the variance in climate can not be related to topography, therefore there is no significant relationship between topography and climate when using a 315^{o} azimuth.

Table 3.2 details the correlation coefficients for maximum and minimum temperature and rainfall when the illumination values were calculated with an azimuth from a southern direction. The interior displays an almost negligible relationship between the climate experienced in this region and its topography due to the fact that there are no major mountain ranges present (Figure 3.2).

A moderate positive correlation between rainfall and topography exists for the South Coast region from December to March. The "Mediterranean" region also experiences a positive correlation (but low) from April to October. This indicates that south-facing slopes experience higher rainfall compared to north-facing slopes. Combining all the data to calculate the correlation for the entire study region indicates a low correlation between south facing-slopes and rainfall. The coefficient of determination indicates a low (10% - 20%) amount of common variation between the topography and rainfall.

When frontal air masses move into the Western Cape they typically cause huge drops in maximum temperatures (SAWS 1996). This is true for both northern and southern facing slopes, thus although cold air masses is the dominant feature influencing the region's temperature the effect is not slope or aspect dependant.

According to Lutgens & Tarbuck (1998) a region's temperature can also be influenced by the amount of direct sunlight exposure. South Africa, situated in the southern hemisphere, receives sunlight from a northern direction, therefore north-facing slopes should experience higher maximum temperatures compared to south-facing slopes. Consequently the slopes with low illumination values should have higher temperatures, indicated by a negative correlation coefficient. The correlation coefficients set out in Table 3.2 indicate a definite, but small inverse relationship between the illumination value and the temperature of a station for the entire study area. For all the regions as well as the entire study area the *p*-values for minimum temperature are most often higher than 0.05 and where it is not the coefficient of determination is significantly low (>10%). This indicates a negligible relationship between minimum temperature and a station's surrounding topography.

Table 3.2: Standard Pearson's coefficient (r) and the coefficient of determination (r^2) for topography (illumination value with an 180° azimuth) and rainfall and temperature

		Karoo Late Summer		South Coast All-year		Mediterranean Winter		All the Stations	
		r	r^2 x 100	r	r^2 x 100	r	r^2 x 100	r	r^2 x 100
Rainfall	January	-0.17	2.85%	0.64	41.19%	0.18	3.25%	0.10	1.01%
	February	-0.26	6.55%	0.66	43.26%	0.14	2.02%	-0.01	0.01%
	March	-0.12	1.48%	0.61	37.44%	0.20	4.12%	0.06	0.38%
	April	-0.18	3.26%	0.55	30.44%	0.32	10.07%	0.32	9.99%
	May	-0.19	3.69%	0.48	23.04%	0.30	8.93%	0.36	12.82%
	June	0.22	5.04%	0.44	19.75%	0.31	9.72%	0.38	14.08%
	July	0.16	2.48%	0.39	15.33%	0.38	14.11%	0.41	16.43%
	August	0.10	0.94%	0.44	19.16%	0.33	10.96%	0.39	14.95%
	September	0.19	3.60%	0.50	25.32%	0.37	13.38%	0.42	17.36%
	October	-0.22	4.76%	0.56	31.46%	0.31	9.86%	0.28	7.89%
	November	0.02	0.05%	0.50	24.81%	0.27	7.04%	0.16	2.67%
	December	-0.30	9.16%	0.57	32.96%	0.14	2.02%	0.10	1.00%
	Annual	*-0.06*	*3.51%*	*0.53*	*28.68%*	*0.27*	*7.96%*	*0.25*	*8.22%*
Maximum Temperature	January	-0.09	0.74%	-0.40	16.36%	-0.11	1.27%	-0.24	5.68%
	February	-0.01	0.01%	-0.41	16.82%	-0.11	1.27%	-0.20	3.96%
	March	-0.07	0.51%	-0.43	18.22%	-0.14	1.89%	-0.20	4.18%
	April	-0.07	0.43%	-0.34	11.83%	-0.21	4.31%	-0.20	3.94%
	May	-0.05	0.24%	-0.09	0.85%	-0.21	4.54%	-0.15	2.36%
	June	-0.03	0.09%	0.08	0.58%	-0.26	6.83%	-0.10	0.97%
	July	-0.01	0.01%	0.00	0.00%	-0.26	6.55%	-0.16	2.65%
	August	-0.09	0.79%	-0.21	4.24%	-0.20	4.02%	-0.26	6.51%
	September	-0.11	1.14%	-0.35	12.23%	-0.19	3.78%	-0.30	8.95%
	October	-0.14	1.92%	-0.40	15.95%	-0.18	3.27%	-0.27	7.24%
	November	-0.11	1.32%	-0.38	14.39%	-0.14	2.00%	-0.23	5.50%
	December	-0.11	1.16%	-0.39	14.87%	-0.13	1.68%	-0.25	6.13%
	Annual	*-0.07*	*0.70%*	*-0.28*	*10.53%*	*-0.18*	*3.54%*	*-0.21*	*4.43%*
Minimum Temperature	January	-0.36	12.92%	0.16	2.41%	-0.20	3.82%	-0.20	4.13%
	February	-0.38	14.55%	0.17	2.93%	-0.24	5.59%	-0.20	4.18%
	March	-0.40	15.98%	0.23	5.21%	-0.24	5.79%	-0.17	2.82%
	April	-0.40	15.71%	0.39	14.84%	-0.24	5.95%	-0.04	0.20%
	May	-0.37	13.82%	0.43	18.19%	-0.11	1.16%	0.09	0.88%
	June	-0.26	6.56%	0.43	18.69%	-0.04	0.15%	0.17	2.92%
	July	-0.30	8.71%	0.42	17.27%	-0.01	0.01%	0.17	2.78%
	August	-0.39	15.02%	0.39	15.46%	-0.01	0.02%	0.12	1.39%
	September	-0.43	18.65%	0.33	10.88%	-0.05	0.29%	0.02	0.05%
	October	-0.46	21.19%	0.26	6.67%	-0.14	1.92%	-0.09	0.80%
	November	-0.36	12.94%	0.22	5.04%	-0.28	7.69%	-0.18	3.29%
	December	-0.40	16.08%	0.16	2.48%	-0.26	6.62%	-0.21	4.48%
	Annual	*-0.38*	*14.34%*	*0.30*	*10.01%*	*-0.15*	*3.25%*	*-0.04*	*2.33%*

Note: Values where $p \leq 0.05$ are printed in grey blocks.

Using illumination value to investigate the relationship between a station's climate and its surrounding topography indicated that the relationship between an area's temperature and topography is negligible. This is not the case for rainfall where correlation analysis indicated that south-facing slopes receive more rain than north-facing slopes. The next step is to investigate the relationship between a station's distance to nearest ocean and its climate.

3.5.2 Correlation with distance to oceans

Table 3.3 presents the results of the correlation analysis of weather station's distance to the nearest ocean and rainfall and temperature conditions experienced at that station. For precipitation a negative *r*-value indicates that the closer a station is to the ocean, the higher the rainfall experienced (i.e. inverse relationship), while a positive value for temperature indicates that the farther a station is situated from an ocean, the higher the temperature is for that region.

The *South Coast* region (all-year rainfall) displays substantial relationship between the climate variables and distance from ocean. For all the months of the year this region experiences an increase in rainfall with a decrease in distance from ocean. During September to April this region experiences higher maximum temperatures for stations farther away from the ocean. There is an inverse relationship for minimum temperatures, that is, stations nearer to the oceans experience higher minimum temperatures from April to October. This highlights the effect of the Indian Ocean with its warm ocean currents which moderates temperature extremes.

The *Karoo* (late summer rainfall) region of the study area experiences moderate correlation between distance to coastline and climate. Rainfall data indicates a moderate correlation with a substantial direct relationship for January, February and April and a moderate correlation with a substantial inverse relationship for May, July and August. This information points to higher precipitation experienced during summer months by stations farther away from the ocean, while stations closer to the coastline experience higher rainfall during the winter months.

In the *"Mediterranean"* region (winter rainfall) there is no significant correlation between rainfall and distance to the ocean. The temperature variables for this region indicate moderate correlation during November to March when higher maximum and minimum temperatures are experienced by stations farther away from the coast.

Table 3.3: Standard Pearson's coefficient (r) and the coefficient of determination (r^2) for distance to nearest ocean and rainfall and temperature

		Karoo Late Summer		South Coast All-year		Mediterranean Winter		All the Stations	
		r	r^2 x 100	R	r^2 x 100	r	r^2 x 100	r	r^2 x 100
Rainfall	January	0.47	21.77%	-0.54	29.54%	-0.09	0.83%	0.11	1.14%
	February	0.62	37.85%	-0.56	31.08%	-0.14	2.09%	0.36	13.28%
	March	0.34	11.89%	-0.50	24.52%	-0.06	0.33%	0.19	3.67%
	April	0.42	18.00%	-0.53	27.70%	-0.08	0.72%	-0.27	7.09%
	May	-0.50	24.87%	-0.71	49.86%	0.00	0.00%	-0.43	18.29%
	June	-0.33	10.69%	-0.58	33.59%	0.00	0.00%	-0.40	15.61%
	July	-0.55	30.15%	-0.60	36.13%	0.03	0.11%	-0.43	18.11%
	August	-0.50	25.37%	-0.72	51.88%	0.00	0.00%	-0.46	20.92%
	September	-0.30	9.18%	-0.60	35.45%	0.00	0.00%	-0.47	21.65%
	October	-0.10	0.91%	-0.67	44.77%	-0.06	0.37%	-0.32	10.42%
	November	-0.05	0.30%	-0.53	28.45%	0.10	0.93%	-0.16	2.57%
	December	0.00	0.00%	-0.46	21.13%	0.03	0.11%	-0.08	0.72%
	Annual	*-0.04*	*15.92%*	*-0.58*	*34.51%*	*-0.02*	*0.46%*	*-0.20*	*11.12%*
Maximum Temperature	January	0.46	21.59%	0.82	67.52%	0.56	31.10%	0.62	37.96%
	February	0.16	2.71%	0.83	68.96%	0.54	29.04%	0.48	22.71%
	March	0.14	1.86%	0.81	65.38%	0.50	24.51%	0.41	16.60%
	April	-0.02	0.03%	0.67	45.51%	0.34	11.62%	0.19	3.56%
	May	-0.01	0.01%	0.30	8.93%	0.11	1.20%	0.00	0.00%
	June	-0.13	1.80%	-0.13	1.65%	-0.14	2.09%	-0.27	7.50%
	July	0.02	0.04%	-0.01	0.00%	-0.14	2.06%	-0.07	0.53%
	August	0.19	3.76%	0.33	10.88%	0.00	0.00%	0.28	7.83%
	September	0.43	18.34%	0.68	46.21%	0.25	6.29%	0.57	32.16%
	October	0.47	22.01%	0.78	61.54%	0.39	15.12%	0.57	32.36%
	November	-0.16	2.50%	0.81	65.63%	0.47	22.08%	0.51	25.75%
	December	0.54	28.77%	0.80	64.00%	0.54	29.22%	0.63	39.49%
	Annual	*0.17*	*8.62%*	*0.56*	*42.18%*	*0.28*	*14.53%*	*0.32*	*18.87%*
Minimum Temperature	January	0.35	12.52%	-0.30	9.08%	0.56	31.10%	0.16	2.65%
	February	0.19	3.45%	-0.28	7.95%	0.54	29.04%	0.04	0.17%
	March	0.05	0.22%	-0.38	14.66%	0.50	24.51%	-0.14	1.85%
	April	-0.15	2.28%	-0.58	34.11%	0.34	11.62%	-0.46	21.21%
	May	-0.38	14.55%	-0.67	45.29%	0.11	1.20%	-0.63	40.27%
	June	-0.56	31.04%	-0.70	49.49%	-0.14	2.09%	-0.71	50.68%
	July	-0.51	26.49%	-0.70	48.33%	-0.14	2.06%	-0.71	49.87%
	August	-0.39	15.21%	-0.68	45.82%	0.00	0.00%	-0.65	42.30%
	September	-0.07	0.47%	-0.63	39.69%	0.25	6.29%	-0.45	20.13%
	October	0.13	1.65%	-0.52	26.80%	0.39	15.12%	-0.26	6.92%
	November	0.31	9.82%	-0.39	15.50%	0.47	22.08%	-0.45	20.43%
	December	0.35	12.03%	-0.31	9.72%	0.54	29.22%	0.07	0.55%
	Annual	*-0.06*	*10.81%*	*-0.51*	*28.87%*	*0.28*	*14.53%*	*-0.35*	*21.42%*

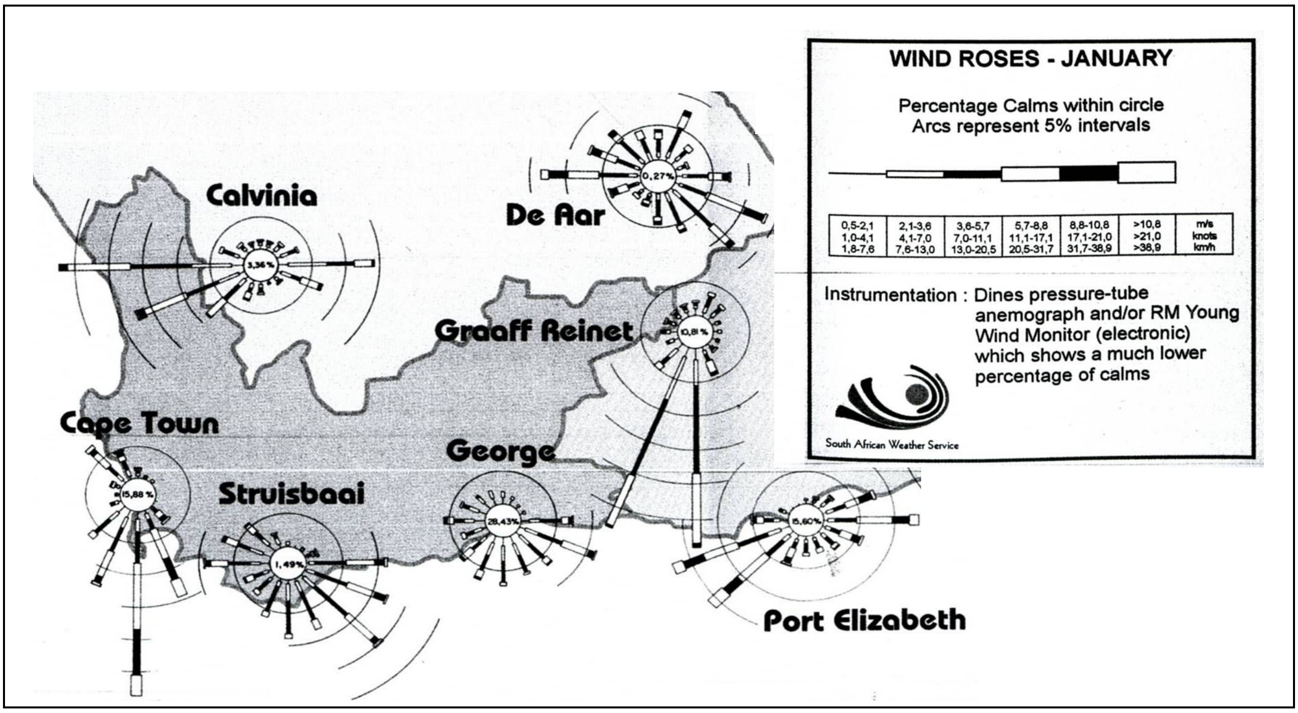
Note: Values where $p \leq 0.05$ are printed in grey blocks.

The picture for the *whole study area* displays a more substantial relationship between distance to ocean and the area's temperatures (both maximum and minimum) compared to rainfall. Analysis of maximums temperature shows a moderate correlation and substantial positive relationship from September until March and a slight negative correlation for June and minimum temperature indicates a moderate negative correlation for April to August. This highlights the moderating effect of the oceans since higher minimum temperatures are recorded near the coast during the winter months, while higher maximum temperatures are experienced farther away from the coastline during the summer months. Rainfall indicates a slight to low positive correlation for February and March during the interior's rainy season indicating that higher rainfall is experienced by areas farther away from the coastline. A moderate negative correlation from May to September indicates that areas near the oceans receive more rain during the winter months.

3.5.3 Summary of the correlation analyses

When the illumination was calculated using northwestern azimuth, no reliable evidence was found that the southern slopes receive less rainfall compared to the northern slopes, except for the South Coast region where the correlation analysis indicated that southern slopes receive more rain during October and December. The "Mediterranean" region, where rainfall is associated with cold fronts accompanied with northwesterly winds, had only a small percentage of weather stations located on the north/northeast/northwest-facing slopes (Table 3.4). Some stations located in this area indicate that certain parts of the region's northern-facing slopes receive more rainfall than the southern slopes, but the correlation analysis points to an almost negligible relationship between a station's illumination value and rainfall. Therefore the hillshade grid created using the northwestern azimuth was disregarded for the rest of this project.

When using the southern azimuth it was established that north-facing slopes receive less rain than southern slopes. The correlation analysis for the "Mediterranean" region indicates a small, but definitive relationship between illumination and rainfall during the rainy season (April to October) with southern slopes receiving more rain than north-facing slopes. Correlation analysis for the South Coast region shows a substantial relationship between direction of slope and rainfall during the summer months when the predominant wind blows on-shore (Figure 3.7). In this case the south-facing slopes receive more rainfall than north-facing slopes.



Source: Kruger 2002: no page

Figure 3.7: Wind-roses indicating the wind directions during January for the study area

The topographical data generated for the weather stations indicates that few stations are located on the north-facing slopes, the majority being on the southern slopes (Table 3.4). This results in to a few stations representing the entire range of illumination values which in turn has an effect on the accuracy of the correlation analysis.

Table 3.4: Aspect of weather stations

	North-facing	North east-facing	East-facing	South east-facing	South-facing	South west-facing	West-facing	North west-facing
"Mediterranean"	1.64%	9.84%	8.20%	22.95%	13.11%	29.51%	13.11%	1.64%
South Coast	4.17%	8.33%	4.17%	12.50%	45.83%	20.83%	0.00%	4.17%
Karoo	0.0 %	25.00%	12.50%	6.25%	12.50%	12.50%	6.25%	25.00%
Total	**1.98%**	**11.88%**	**7.92%**	**17.82%**	**20.79%**	**24.75%**	**8.91%**	**5.94%**

The correlation analysis of distance to the nearest ocean provides evidence that this additional variable does have an effect on the temperatures (both maximum and minimum) experienced at a station. The results support the contention that the oceans have a moderating effect on temperatures, since the weather stations closer to the oceans experienced higher minimum temperatures and lower maximum temperatures than stations farther away from the sea.

The distance to ocean and rainfall correlation analysis indicates that the rainfall experienced by a region increases or decreases depending on the season, but generally stations closer to the sea receive more rain compared to stations farther away. For the rainfall variable it is clear that both distance from the sea and topography play a role in the amount of precipitation recorded at a station.

Figure 3.4 shows the distribution of weather stations used during this study while Table 3.5 provides the maximum, minimum, average and median of the stations' distances from the nearest ocean for each of the three rainfall regions. Since distance was calculated to the closest ocean, the possibility that the cold Atlantic waters on the west coast could have more of an influence than the warmer Indian Ocean waters on a region which is bordered by both oceans (such as the southwestern part of the study area) or vice versa was not taken into account during this study and may require further investigation.

Table 3.5: Shortest distance to ocean from weather stations

	Maximum	**Minimum**	**Average**	**Median**
"Mediterranean "	14 8205	0	42 144	37 713
South Coast	96 149	0	25 377	13 951
Karoo	376 316	44 705	197 166	189 472
Total	**376 316**	**0**	**42 144**	**37 713**

Note: Calculated in metres.

Now that the correlation between the additional variables and the region's climate has been assessed, the next step will be to incorporate these factors into the interpolation process. In view of the fact that no suggestions of suitable transformation and units for these additional variables were provided in the literature, the following section will aim at identifying the most suitable parameterization for topography (aspect and slope) and distance to nearest ocean.

3.6 SUITABLE TRANSFORMATIONS AND UNITS FOR ADDITIONAL VARIABLES

ANUSPLIN version 4.3 makes provision for a range of units (e.g. metres, kilometres, degrees) and also for the transformation of independent and dependent variables which can have an overall influence on the accuracy of the interpolation process (Hutchinson 2004). In this study the independent variables are weather station location characteristics, while climate data are the dependent variables. Before the interpolation process can commence it is important to carry out the required variable transformations and to specify the correct units for the various additional variables. Research has been done elsewhere on the most suitable transformations and units of dependent variables such as rainfall and the standard independent variables longitude, latitude and elevation, but very little information is available for the most suitable units for and transformations of additional variables.

The fourth objective of this study is to investigate suitable transformations of and units for the additional variables discussed in sections 3.4 and 3.5. Before this is done, it is expedient to recall other studies in which the transformation of and units chosen for the *standard variables* are discussed to get clarity about how they are to be applied in the interpolation process of this project. Variables providing location information such as latitude and longitude are given in degrees and elevation in metres. The units for dependent variables are degrees for temperature and millimetres for rainfall.

ANUSPLIN permits two dependent variable transformations, namely the square root and natural logarithm which may be applied to the dataset before the spline function is calculated. When a transformation has been performed, ANUSPLIN allows for the back transformation of the variable before the final climate grid is created. Hutchinson (1998a) has concluded that applying the square root transformation to daily rainfall data can reduce the interpolation error by 10%. In the ANUSPLIN user's manual, Hutchinson (2004) provides evidence that superior interpolation performance over restricted areas can be achieved if elevation is scaled from metres to kilometres and included as an independent covariate instead of an independent variable (thus, "the coefficient of an elevation covariate would be an empirically determined temperature lapse rate") (Hutchinson 2004:12)).

With the purpose of identifying suitable transformations and units for the *additional variables*, various possibilities for the distance to ocean and topography (aspect and slope) variables were investigated. In order to test the accuracy of the range of options, ANUSPLIN provides a series of statistical outputs which can be used for performance analysis (Hutchinson 1998a, 1998b; Price *et al.* 2000).

The generalised cross-validation (GCV), the mean square residual (MSR - average difference between observed and predicted data) and an unbiased estimate of the 'true' mean square error (MSE) of the fitted function across the data points are written to an output log file. Other statistical information such as the square roots of the GCV, MSR and MSE (in the units of the data values), namely the root of the generalised cross-validation (RTGCV), the root of the mean square residual (RTMSR) and the root of the mean square error (RTMSE) are also provided.

The RTGCV values are conservative estimates of the overall standard prediction error as it includes the data error estimated by the procedure. The RTMSE value is a prediction of the standard error after the predicted data error has been removed (Hutchinson 2004).

Another important diagnostic statistic associated with a fitted spline function is the signal value. The signal value is an indication of the number of degrees of freedom of the fitted spline, i.e. it provides the effective number of parameters of the fitted spline model used after the interpolation has been completed. Hutchinson (1998a, 1998b, 2004) and Price *et al.* (2000) propose a signal value of approximately half the number of data points used for a second order splining function. A signal value higher than 80% of the number of data points indicates significant data errors, lack of data points or a short range correlation in the data values (Hutchinson 1998a). Since 116 suitable data points were available for this study an acceptable signal value will fall between 40 and 70 for a trivariate (only latitude, longitude and elevation are input variables) second order spline function. If ANUSPLIN's splining function calculator could not calculate a suitable spline function (i.e. the signal value is too low or high) the spline value is marked with an asterisk (*). The signal, RTGCV and RTMSE values are the diagnostic statistics most often used to assess the performance of ANUSPLIN.

In addition to using the output statistics produced by ANUSPLIN, another popular method of testing the accuracy of the interpolation process is by means of withholding a selected number of stations from the dataset used during the interpolation process. The interpolated climate information for the test stations are then compared to the actual measured information to produce error values (in the original variable units). In order to select the test stations to be extracted, the following simple methodology was employed.

Resembling other interpolation projects which followed the same procedure, it was decided to select 20% of the stations as test cases. In order to enforce a proportionally greater selection of test cases where weather stations are more densely concentrated a stratified random selection was implemented (Mouton 2001). The area in which stations were available was divided into the three concentration zones illustrated in Figure 3.8. The density zones were plotted using ArcMap's (version 9.1) spatial Analyst tool. The selection of test stations was done randomly within each density zone shown.

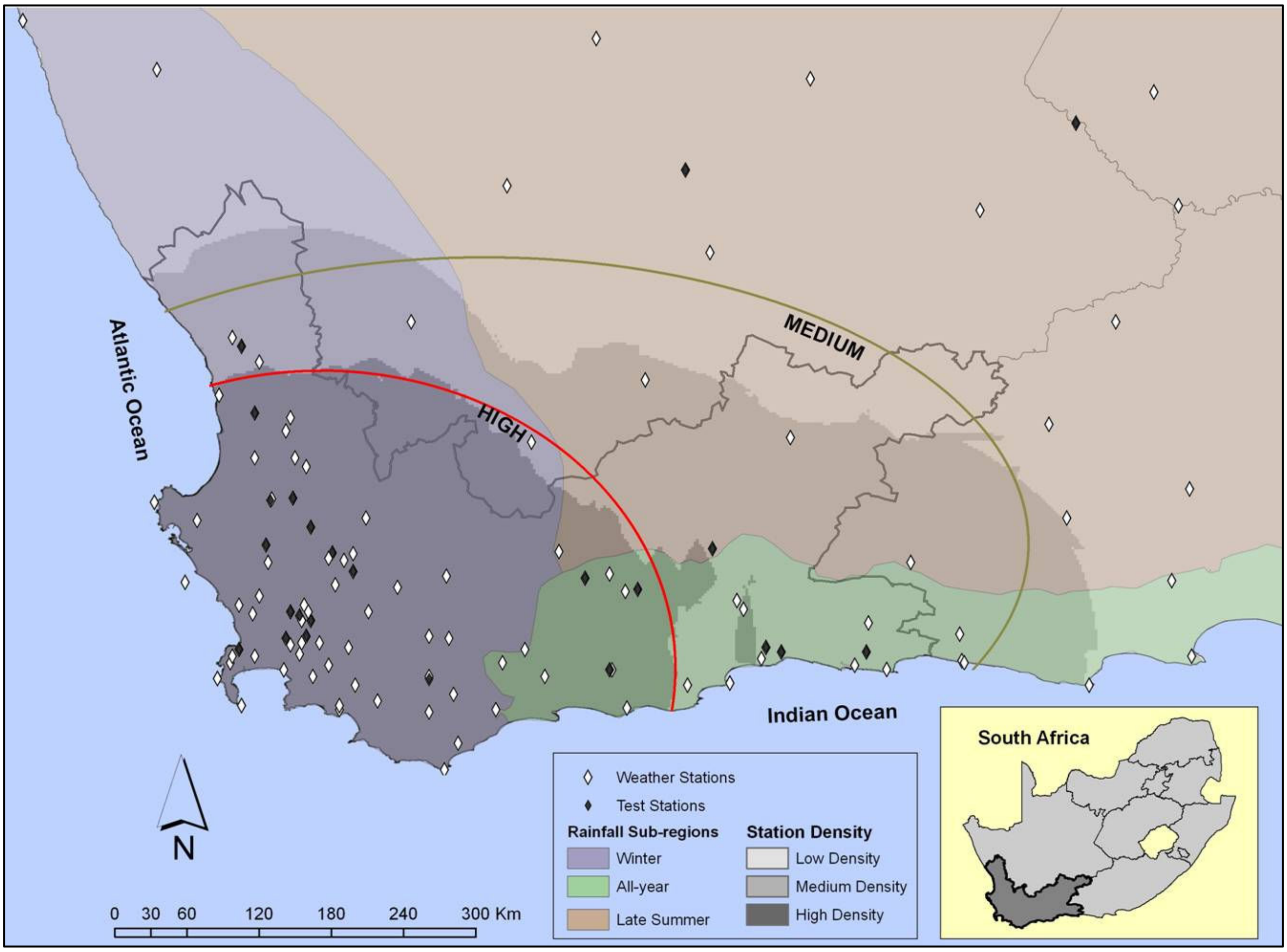


Figure 3.8: Distribution of weather stations selected for testing

Table 3.6 gives a breakdown of the population of stations and the sample of test stations in each density zone. 70% of the test stations are located in the high density zone which contains 65% of the weather stations. For the medium and high zones approximately 22% of the stations were selected as test stations while only 11.1% was selected in the low density zone.

Table 3.6: Number of weather stations and test stations in the three density zones

Station Density Zone	Weather Stations	Test Stations	Test stations as % of total in density zone
Low	18 (15.5%)	2 (8.4%)	11.1
Medium	22 (19.0%)	5 (20.8%)	22.7
High	76 (65.5%)	17 (70.8%)	22.4
Total	**116 (100%)**	**24 (100%)**	**20.6%**

The following section describes the selection of suitable units for and transformations of (i.e. parameterization) the additional variables for monthly mean daily maximum and minimum temperature and monthly mean rainfall. ANUPSLIN's statistical output (signal, RTGCV and RTMSE values) as well as the test stations residual values will be used to identify the most accurate model for the interpolation process.

3.6.1 Interpolation analysis: Monthly mean daily maximum temperature

The correlation analyses (Table 3.1 and Table 3.2) completed on the hillshade data for both the southern- and northwestern azimuth grids provided little evidence of a dependable relationship between the topography at the location of the station and a station's recorded temperatures. The relationships that do exist, is small to negligible. Due to these inadequate results hillshade's illumination values will not be used as an additional variable in the creation of maximum and minimum temperature grids.

In the case of the correlation analyses done on the relationship between a station's temperatures and its distance to the nearest ocean, moderate correlations were observed (Table 3.3). Therefore it is advisable to incorporate this variable into the interpolation process. This calls for determining suitable transformations of and units for this additional variable using the quadvariate (latitude, longitude, elevation and distance) models described in Table 3.7.

Table 3.7: Quadvariate model description for distance to ocean variable analysis for maximum temperature

Model	Description
A	Latitude, longitude, elevation and distance are all independent variables
B	Latitude and longitude as independent variables; elevation and distance as covariates with distance in metres
C	Latitude and longitude as independent variables; elevation and distance as covariates with distance scaled to kilometres
D	Latitude and longitude as independent variables; elevation and distance (in metres) as covariates with the logarithmtic transformation of distance (A x log (x + B))
E	Latitude and longitude as independent variables; elevation and distance (scaled to kilometres) as covariates, with the logarithmtic transformation of distance (A x log (x + B))

When using trivariate models the ANUSPLIN user's manual (Hutchinson 2004) suggests a second-order spline function by default. Adding an additional variable as an independent variable changes the default to a third-order spline function. This is not the case when adding the additional variable as a covariate; then the default remains a second-order spline function. Hutchinson (1998b) indicates that third-order splines produce better results when working with a quadvariate model while preliminary analysis of the results obtained for model B, C, D and E corroborate these findings (i.e. the error values were to high when using a second-order spline). Therefore, all the quadvariate models implements a third-order spline function in this project.

Model A incorporates all the variables as independent variables while Model B includes elevation and distance as covariates. When model A was run it was found that the maximum and minimum temperatures produced for the maximum temperature grids were significantly lower or higher (by ±10°C) than the original dataset's values, while model B's values were within the dataset's range. Consequently the rest of the analysis was done using model B's parameterization, but the distance to ocean variable units and transformation formulas were changed. Model C scales distance to kilometres while models D and E investigate the possibility of using a logarithmic transformation on the distance variable.

The logarithmic equation A x log(x + B), with A and B being variables determined by the user and x the distance value, will decrease the range of values for distance, therefore reducing the influence of the variable in the interpolation process forcing elevation to be more important. For the purpose of this study, A and B were both determined to be equal to 1. The reason for B being set to 1 is that the underlying grid used to create the climate grids has values of 0m (i.e. on the coastline). ANUSPLIN calculates the logarithm transformation for each cell on the grid, but when calculating the logarithm (log x) of cells whose values are 0m an error occurs since log(0) is not defined (therefore all the cells in the distance grid and all the station's distance values are increased by 1 when setting B to 1).

Table 3.8 presents the output statistics produced when ANUSPLIN calculates the spline function for the models described in Table 3.7. An overview of the statistics and their significance is provided in introduction paragraph of section 3.6. Model A's signal values are generally higher than those of models B, C and D, with model E producing the lowest signal values. For all the models run, the signal values for April to August are much lower compared to the rest of the months.

Table 3.8: Spline function statistics for the maximum temperature quadvariate models

	Signal					RTGCV (°C)					RTMSE (°C)				
Model	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Jan	72.2	62.8	62.8	64.5	59.8	1.06	1.04	1.04	**1.01**	1.04	0.51	0.52	0.52	**0.50**	0.52
Feb	70.9	64.9	64.9	66.0	32.9	1.15	1.11	1.11	**1.06**	1.13	0.56	0.55	0.55	0.52	**0.51**
Mar	74.6	65.4	65.4	66.2	28.4	1.08	1.05	1.05	**1.00**	1.02	0.52	0.52	0.52	0.49	**0.44**
Apr	69.8	65.2	65.2	25.7	23.0	0.94	0.90	0.90	0.87	**0.81**	0.46	0.44	0.44	0.36	**0.32**
May	60.0	23.3	23.3	22.9	22.8	0.77	0.76	0.76	**0.66**	0.67	0.39	0.31	0.31	**0.26**	0.27
Jun	38.7	25.1	25.1	25.6	26.8	0.62	0.62	0.62	**0.58**	0.59	0.29	0.26	0.26	**0.24**	0.25
Jul	37.9	25.5	25.5	25.2	25.1	0.60	0.58	0.58	**0.55**	**0.55**	0.28	0.24	0.24	**0.23**	**0.23**
Aug	56.0	24.8	24.8	24.7	23.7	0.70	0.70	0.70	**0.65**	0.66	0.35	0.29	0.29	**0.27**	**0.27**
Sep	70.9	66.7	66.7	66.8	24.4	0.79	0.77	0.77	**0.73**	0.79	0.39	0.38	0.38	0.36	**0.32**
Oct	69.1	63.6	63.6	63.1	56.2	0.91	0.90	0.90	**0.87**	0.90	0.45	0.45	0.45	**0.43**	0.45
Nov	71.4	62.2	62.2	64.2	38.2	0.94	0.92	0.92	**0.88**	0.91	0.46	0.46	0.46	0.44	**0.43**
Dec	71.4	62.7	62.7	63.8	33.9	1.03	1.03	1.03	**1.02**	1.03	0.50	0.51	0.51	0.51	**0.47**
Avg	*63.6*	*51.0*	*51.0*	*48.2*	*32.9*	*0.90*	*0.88*	*0.88*	***0.84***	*0.86*	*0.44*	*0.42*	*0.42*	*0.40*	***0.39***

Note: Bold values are the lowest errors.

The results for models B and C indicate that scaling the distance variable does not have any effect on the statistical output produced while calculating the spline function or the error statistics produced when the test stations were extracted from the dataset (Table 3.9). For models D and E more experimentation with variable A in the logarithmic transformation was done using multiples of ten, but it was found that there were no noticeable changes in the signal, RTGCV and RTMSE values of the spline function.

Model A's RTGCV and RTMSE values are higher than that of the other models. Model D and E's RTGCV and RTMSE values are predominantly lower than those of models A, B and C. Compared to model E, model D's signal values are higher, while its RTGCV values are often lower and its RTMSE values are mainly higher (where model E's RTMSE values are higher the difference is generally not more than 0.02^{o}C). Hutchinson (2004) and Price *et al*. (1999) indicate that a RTMSE error value of 0.5^{o}C or less is acceptable when calculating the spline function. This was achieved for almost all of the months of the year with only February being slightly higher.

Isolating the test stations from the dataset provided the error values indicated in Table 3.9 with the lowest values shown in bold. For most of the months of the year model E has the lowest mean, maximum and minimum errors while model D's values are also low for May to August.

Table 3.8 and 3.9 indicate that overall model E produced the best interpolation results. It is possible to decide on the most suitable model for each month, but for the purpose of this project, model E was selected to interpolate the maximum temperature grids for all twelve months of the year.

Table 3.9: Error values for the maximum temperature quadvariate models

	Mean Error (°C)					Minimum Error (°C)					Maximum Error (°C)				
Model	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Jan	0.52	0.60	0.60	0.48	**0.39**	0.38	0.32	0.32	0.29	**0.24**	**0.72**	1.50	1.50	0.92	0.74
Feb	0.54	0.62	0.62	0.48	**0.40**	0.40	0.33	0.33	0.29	**0.25**	0.76	1.70	1.70	0.90	**0.75**
Mar	0.50	0.58	0.58	0.42	**0.34**	0.37	0.31	0.31	0.25	**0.21**	0.72	1.60	1.60	0.79	**0.64**
Apr	0.46	0.41	0.41	0.31	**0.26**	0.33	0.24	0.24	0.19	**0.16**	0.62	0.78	0.78	0.59	**0.49**
May	0.35	0.27	0.27	**0.22**	**0.22**	0.24	0.16	0.16	**0.14**	**0.14**	0.47	0.52	0.52	**0.42**	**0.42**
Jun	0.27	0.23	0.23	**0.21**	0.22	0.18	0.14	0.14	**0.13**	0.14	0.37	0.44	0.44	**0.40**	0.41
Jul	0.26	0.22	0.22	**0.20**	**0.20**	0.18	0.13	0.13	**0.12**	**0.12**	0.35	0.42	0.42	**0.37**	**0.37**
Aug	0.32	0.25	0.25	**0.22**	**0.22**	0.22	0.15	0.15	**0.14**	**0.14**	0.42	0.48	0.48	0.42	**0.41**
Sep	0.37	0.34	0.34	0.28	**0.24**	0.27	0.20	0.20	0.17	**0.15**	0.52	0.63	0.63	0.53	**0.46**
Oct	0.35	0.44	0.44	0.34	**0.27**	0.27	0.24	0.24	0.21	**0.18**	0.54	1.40	1.40	0.64	**0.53**
Nov	0.46	0.52	0.52	0.49	**0.34**	0.34	0.28	0.28	0.27	**0.21**	0.64	1.20	1.20	1.20	**0.63**
Dec	0.47	0.57	0.57	0.47	**0.38**	0.35	0.31	0.31	0.28	**0.23**	**0.68**	1.50	1.50	0.88	0.72
Avg	*0.41*	*0.42*	*0.42*	*0.34*	***0.29***	*0.29*	*0.23*	*0.23*	*0.21*	***0.18***	*0.57*	*1.01*	*1.01*	*0.67*	***0.55***

Note: Bold values are the lowest errors.

The best performing quadvariate model's statistics (model E) was compared with those of the original trivariate model (Table 3.10 A and B) in order to determine whether adding the distance variable enhanced the accuracy for the maximum temperature grids. This comparison undoubtedly indicates that the quadvariate model has lower signal, RTGCV and RTMSE values while the mean and minimum, and most of the maximum error values, are also lower than that of the trivariate model.

Table 3.10: A and B: Spline function statistics (A) and error calculations (B) for maximum temperature for the trivariate model and the best-performing quadvariate model.

	Signal		RTGCV (°C)		RTMSE (°C)	
Model	**Triv**	**Quad**	**Triv**	**Quad**	**Triv**	**Quad**
Jan	72.2	59.8	1.06	**1.04**	**0.51**	0.52
Feb	70.9	32.9	1.15	**1.13**	0.56	**0.51**
Mar	74.6	28.4	1.08	**1.02**	0.52	**0.44**
Apr	69.8	23.0	0.94	**0.81**	0.46	**0.32**
May	60.0	22.8	0.77	**0.67**	0.39	**0.27**
Jun	38.7	26.8	0.62	**0.59**	0.29	**0.25**
Jul	37.9	25.1	0.60	**0.55**	0.28	**0.23**
Aug	56.0	23.7	0.70	**0.66**	0.35	**0.27**
Sep	70.9	24.4	**0.79**	**0.79**	0.39	**0.32**
Oct	69.1	56.2	0.91	**0.90**	**0.45**	**0.45**
Nov	71.4	38.2	0.94	**0.91**	0.46	**0.43**
Dec	71.4	33.9	**1.03**	**1.03**	0.50	**0.47**
Average	*63.6*	*32.9*	*0.90*	***0.86***	*0.44*	***0.39***

Mean Error (°C)		Min Error (°C)		Max Error (°C)	
Triv	**Quad**	**Triv**	**Quad**	**Triv**	**Quad**
0.52	**0.39**	0.38	**0.24**	**0.72**	0.74
0.54	**0.40**	0.40	**0.25**	0.76	**0.75**
0.50	**0.34**	0.37	**0.21**	0.72	**0.64**
0.46	**0.26**	0.33	**0.16**	0.62	**0.49**
0.35	**0.22**	0.24	**0.14**	0.47	**0.42**
0.27	**0.22**	0.18	**0.14**	**0.37**	0.41
0.26	**0.20**	0.18	**0.12**	**0.35**	0.37
0.32	**0.22**	0.22	**0.14**	0.42	**0.41**
0.37	**0.24**	0.27	**0.15**	0.52	**0.46**
0.35	**0.27**	0.27	**0.18**	0.54	**0.53**
0.46	**0.34**	0.34	**0.21**	0.64	**0.63**
0.47	**0.38**	0.35	**0.23**	**0.68**	0.72
0.41	***0.29***	*0.29*	***0.18***	*0.57*	***0.55***

Note: Bold values are the lowest errors.

The analysis of the maximum temperature data shows that the inclusion of the additional variable of distance to ocean produces better results compared to the trivariate model. The most appropriate parameterisation for this additional variable appears to be the logarithmic transformation of the distance variable which is scaled to kilometres.

3.6.2 Interpolation analysis: Monthly mean daily minimum temperature

Minimum temperature was considered to be a separate entity from maximum temperature and parameterization analysis was done independently. Although a few months did display a relationship between hillshade's illumination value and recorded minimum temperature, it was not strong and was very irregular (Table 3.1 and Table 3.2).

Correlation analyses performed on the data point towards a strong relationship between distance to ocean and minimum temperature for the entire study area from April until September, while there is no indication of correlation during January to March, or November and December (Table 3.3). The statistics produced during the correlation analysis show that higher minimum temperatures are experienced by areas closer to the coastline demonstrating the moderating effect of the oceans during the winter months. Consequently, the distance to ocean variable will be incorporated in the minimum temperature interpolation process using the same models as presented in Table 3.7.

Table 3.11 indicates that calculating a spline function for the minimum temperature data proved to be more difficult than for maximum temperature since the signal values are lower. As with the maximum temperature investigation, it was found that model A produced maximum and minimum temperatures that deviated a great deal (by ±10^{o}C) from the original dataset's values while model B's values were within the dataset's range. For some months (indicated with an asterisk), models A, D and E could not calculate a suitable spline function since too few data points could be used resulting in signal values that are too small. Therefore, no error analysis could be performed for these months. For these months, models B and C – using different parameterization, could calculate a suitable spline function from the dataset.

The best RTGCV values were obtained through logarithmic transformation of distance scaled to kilometres (model E). Once the known error value were incorporated into the statistics, model E's RTMSEs were the lowest for most of the months, while model B and C's values were close to the acceptable value of 0.5^{o}C or lower.

Table 3.11: Spline function statistics for the minimum temperature quadvariate models

	Signal					RTGCV (°C)					RTMSE (°C)				
Model	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Jan	15.0*	27.5	27.5	26.4	25.7	–	1.17	1.17	–	**1.09**	–	0.50	0.50	**0.49**	**0.49**
Feb	16.6*	51.7	51.7	53.2	26.3	–	1.14	1.14	–	**1.10**	–	0.57	0.57	0.57	**0.50**
Mar	15.4*	49.7	49.7	51.0	22.0	–	1.22	1.22	–	**1.18**	–	0.60	0.60	0.61	**0.50**
Apr	23.5	8.3	8.3	18.2	9.3	1.32	1.38	1.38	1.39	**1.29**	0.53	**0.35**	**0.35**	0.51	0.39
May	23.7	8.6	8.6	8.0*	8.0	**1.52**	1.59	1.59	–	–	0.61	**0.42**	**0.42**	–	–
Jun	25.0	11.1	11.1	9.2	8.0 *	**1.62**	1.72	1.72	1.60	–	0.67	0.51	0.51	**0.43**	–
Jul	26.5	11.0	11.0	9.7	8.0 *	**1.59**	1.72	1.72	1.59	–	0.67	0.50	0.50	**0.44**	–
Aug	26.8	10.7	10.7	8.0*	8.0 *	**1.36**	1.50	1.50	–	–	0.57	**0.44**	**0.44**	–	–
Sep	28.5	10.1	10.1	8.0*	8.4	1.19	1.34	1.34	–	**1.17**	0.51	0.38	0.38	–	**0.34**
Oct	27.0	9.1	9.1	21.6	9.2	1.15	1.26	1.26	1.21	**1.15**	0.49	**0.34**	**0.34**	0.47	0.35
Nov	34.4	21.0	21.0	22.4	20.8	1.10	1.14	1.14	1.12	**1.06**	0.50	**0.44**	**0.44**	**0.44**	**0.44**
Dec	15.0*	58.1	58.1	62.3	23.9	–	1.02	1.02	**1.01**	1.10	–	0.51	0.51	0.50	**0.49**
Avg	*23.1*	*23.1*	*23.1*	*24.8*	*14.8*	*1.30*	*1.37*	*1.37*	*1.31*	***1.26***	*0.52*	*0.47*	*0.47*	*0.47*	***0.43***

Notes: Bold values are the lowest errors. Asterisks indicate signal values that are too high or too low.

Table 3.12 provides the error statistics after the test stations were extracted from the original dataset. During the months that model E and D produced acceptable signal values, they produced the smallest mean, minimum and maximum errors, while model A values are the highest. The mean error values for models B and C were exactly the same. All the mean error values for model B, C, D and E are below 0.5^{o}C.

Table 3.12: Error values for the minimum temperature quadvariate models

	Mean Error (°C)					Minimum Error (°C)					Maximum (°C)				
Model	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Jan	–	0.47	0.47	**0.39**	**0.39**	–	0.27	0.27	**0.24**	**0.24**	–	0.85	0.85	0.73	**0.72**
Feb	–	0.48	0.48	0.47	**0.39**	–	0.28	0.28	0.28	**0.24**	–	0.87	0.87	0.85	**0.74**
Mar	–	0.50	0.50	0.41	**0.39**	–	0.29	0.29	**0.25**	**0.25**	–	0.91	0.91	0.78	**0.74**
Apr	0.47	0.32	0.32	**0.30**	**0.30**	0.30	0.21	0.21	**0.20**	0.22	1.00	0.65	0.65	0.63	**0.62**
May	0.58	**0.37**	**0.37**	–	–	0.38	0.24	0.24	–	–	1.30	0.75	0.75	–	–
Jun	0.64	0.43	0.43	**0.33**	–	0.42	0.27	0.27	**0.23**	–	1.40	0.85	0.85	0.72	–
Jul	0.63	0.43	0.43	**0.32**	–	0.41	0.27	0.27	**0.22**	–	1.40	0.85	0.85	0.71	–
Aug	0.55	**0.38**	**0.38**	–	–	0.36	0.24	0.24	–	–	1.20	0.75	0.75	–	–
Sep	0.48	**0.32**	**0.32**	–	0.27	0.31	0.20	0.20	–	0.19	1.00	0.64	0.64	–	**0.55**
Oct	0.48	0.30	0.30	0.26	0.27	0.31	0.19	0.19	0.**18**	0.19	1.00	0.60	0.60	**0.55**	**0.55**
Nov	0.39	0.36	0.36	**0.35**	**0.35**	0.25	**0.22**	**0.22**	**0.22**	**0.22**	0.83	0.69	0.69	**0.66**	**0.65**
Dec	–	0.40	0.40	**0.38**	**0.38**	–	0.24	**0.24**	**0.23**	0.24	–	0.76	0.76	0.72	**0.71**
Avg	*0.53*	*0.40*	*0.40*	***0.34***	***0.34***	*0.34*	*0.24*	*0.24*	***0.22***	*0.23*	*1.14*	*0.76*	*0.76*	*0.69*	***0.67***

Note: Bold values are the lowest errors.

Model E's low values indicate that the logarithmic transformation produces the best results, but it can not be used for all twelve months. Model C, which produced suitable results for all the months of the year, was selected as the most suitable quadvariate model to compare with the trivariate model (Table 3.13). The trivariate model has higher signal and lower RTGCV values, but once the known standard error was removed the quadvariate model had lower RTMSE values. Except for the months May to August, the mean and minimum error values of the quadvariate model are lower than those of the trivariate model, but the trivariate model has lower maximum errors throughout the year.

Table 3.13: A and B: Spline function statistics (A) and error calculations (B) for minimum temperature for the trivariate model and the best-performing quadvariate model.

	Signal		RTGCV (°C)		RTMSE (°C)	
Model	**Triv**	**Qua**	**Triv**	**Qua**	**Triv**	**Qua**
Jan	47.3	27.5	**1.14**	1.17	0.56	**0.50**
Feb	53.2	51.7	**1.11**	1.14	**0.55**	0.57
Mar	54.7	49.7	**1.17**	1.22	**0.59**	0.60
Apr	53.9	8.3	**1.32**	1.38	0.66	**0.35**
May	11.4	8.6	**1.59**	**1.59**	0.47	**0.42**
Jun	14.1	11.1	**1.71**	1.72	0.56	**0.51**
Jul	14.5	11.0	**1.71**	1.72	0.57	**0.50**
Aug	14.5	10.7	**1.50**	**1.50**	0.50	**0.44**
Sep	13.9	10.1	**1.34**	**1.34**	0.43	**0.38**
Oct	58.3	9.1	**1.17**	1.26	0.58	**0.34**
Nov	57.8	21.0	**1.05**	1.14	0.53	**0.44**
Dec	58.5	58.1	**1.01**	1.02	**0.51**	**0.51**
Avg	*37.0*	*23.1*	***1.34***	*1.37*	*0.55*	***0.47***

Mean Error		Min Error (°C)		Max Error (°C)	
Triv	**Quad**	**Triv**	**Quad**	**Triv**	**Quad**
0.52	**0.47**	0.38	**0.27**	**0.72**	0.85
0.54	**0.48**	0.40	**0.28**	**0.76**	0.87
0.50	**0.50**	0.37	**0.29**	**0.72**	0.91
0.46	**0.32**	0.33	**0.21**	**0.62**	0.65
0.35	0.37	**0.24**	**0.24**	**0.47**	0.75
0.27	0.43	**0.18**	0.27	**0.37**	0.85
0.26	0.43	**0.18**	0.27	**0.35**	0.85
0.32	0.38	**0.22**	0.24	**0.42**	0.75
0.37	**0.32**	0.27	**0.20**	**0.52**	0.64
0.35	**0.30**	0.27	**0.19**	**0.54**	0.60
0.46	**0.36**	0.34	**0.22**	**0.64**	0.69
0.47	**0.40**	0.35	**0.24**	**0.68**	0.76
0.41	***0.40***	*0.29*	***0.24***	***0.57***	*0.76*

Note: Bold values are the lowest errors.

From the discussion above, it can be stated that better results were obtained for the minimum temperature interpolation procedure for November to March compared to the ‘cooler’ months of April to October. This contrasts with the correlation analyses which indicated that minimum temperature and distance to ocean have a more substantial relationship during the ‘cooler’ months (April to September). Although the results indicate that minimum temperature interpolation is less successful than maximum temperature interpolation, it is apparent that by adding the distance to ocean parameter better results for both maximum and minimum temperature were obtained.

3.6.3 Interpolation analysis: Mean monthly rainfall

Compared to the temperatures experienced in a region, rainfall displays complex spatial patterns, resulting in less accurate rainfall grids being produced during interpolation projects. This problem has been given attention by Hutchinson (1998a, 1998b) where the most efficient parameterization for the various input variables were analysed for the production of rainfall grids.

One of the conclusions drawn by Hutchinson (1998a) is that more accurate results were obtained when the square root transformation was applied to the rainfall data before the spline function is calculated. There is also evidence that better results are obtained when elevation is included as a covariate instead of as an independent variable and scaled to kilometres (Hutchinson 2004). Before any analysis could be performed on the additional variables, it was necessary to identify the correct parameterization for elevation and to establish if any transformation should be done on the rainfall data as suggested by previous rainfall interpolation projects. The four trivariate models which were run in this project are summarized in Table 3.14.

Table 3.14: Trivariate model description for the standard variables for rainfall

Model	Description
A	Latitude, longitude and elevation as independent variables; rainfall no transformation
B	Latitude, longitude and elevation as independent variables; rainfall data square root transformation
C	Latitude and longitude as independent variables; elevation as covariate; rainfall data no transformation
D	Latitude and longitude as independent variables; elevation as covariate; rainfall data square root transformation

The test runs of all the models were also done with scaling elevation to kilometres or leaving it in metres, but no differences in output statistics were observed. Table 3.15 gives the output statistics from calculating the spline function for the trivariate models. The RTGCV and RTMSE values produced by models A and C are too high which provides confirmation that rainfall values need to undergo square root transformation, but note that these values are transformed back to original values when generating the final grids. Model D generated the best signal value (should be around half the number of data points, from 40 to 70) but its RTGCV and RTMSE values are higher than those of model B.

Table 3.15: Spline function statistics for the rainfall trivariate models

	Signal				**RTGCV ($mm^{1/2}$)**				**RTMSE ($mm^{1/2}$)**			
Model	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Jan	105.5	98.1	96.0	91.7	–	**0.83**	9.27	0.85	–	**0.30**	3.50	0.35
Feb	104.7	90.7	34.0	32.2	–	**0.93**	11.60	1.03	–	**0.39**	5.29	0.46
Mar	85.4	71.9	76.8	73.4	11.60	**0.95**	11.70	0.96	5.13	**0.46**	5.55	**0.46**
Apr	64.8	80.9	19.9	26.9	15.50	**1.00**	16.10	1.07	7.71	0.46	6.07	**0.45**
May	55.5	77.3	24.0	46.9	22.60	**1.18**	24.40	1.31	11.30	**0.56**	9.90	0.64
Jun	57.9	82.3	20.6	31.2	32.10	**1.45**	34.40	1.61	16.00	**0.66**	13.10	0.71
Jul	77.2	98.4	22.6	61.7	27.30	**1.26**	30.00	1.41	12.90	**0.45**	11.90	0.70
Aug	42.3	66.6	19.0	35.0	27.40	**1.35**	28.80	1.46	13.20	**0.67**	10.70	**0.67**
Sep	72.5	99.1	22.2	57.4	18.40	**1.13**	19.60	1.24	8.89	**0.40**	7.72	0.62
Oct	105.1	88.1	18.3	26.0	–	**1.21**	17.90	1.29	–	**0.52**	6.51	0.54
Nov	97.4	102.0	59.7	68.1	8.88	**0.76**	9.89	0.83	3.25	**0.25**	4.95	0.41
Dec	78.4	82.2	56.6	70.6	10.40	**0.88**	11.00	0.90	4.85	**0.40**	5.50	0.44
Avg	*78.9*	*86.5*	*39.1*	*51.8*	*19.10*	***1.10***	*20.50*	*1.19*	*8.94*	***0.47***	*8.11*	*0.55*

Notes: Bold values are the lowest errors. Asterisks indicate signal values that are too high or too low.

Table 3.16 indicates error values in millimetres as well as percentages only for models B and D since models A and C could not be used. The percentage value, calculated using mean surface values, is provided to illustrate the magnitude of the error since 5mm in the dry season is much more significant than 5mm in a wet season. Like the results obtained during the temperature analyses, it was found that all the interpolated rainfall grids' maximum values are much higher (200mm) than the dataset's maximum values when elevation is incorporated as an independent variable.

Table 3.16: Error values for the rainfall trivariate models

	Mean Error				Minimum Error		Maximum Error	
Model	**B**		**D**		**B**	**D**	**B**	**D**
	mm	**%**	**Mm**	**%**	**mm**	**mm**	**mm**	**mm**
Jan	3.0	14.2	2.7	**12.9**	**1.5**	**1.5**	**5.7**	7.2
Feb	3.0	14.2	2.8	**13.0**	**1.4**	1.5	**7.2**	9.0
Mar	3.8	12.7	3.2	**10.7**	2.4	**2.0**	**7.5**	8.1
Apr	4.0	**8.9**	4.7	10.9	**2.6**	3.4	**6.0**	7.6
May	5.9	**10.1**	7.1	12.5	**2.4**	4.1	**10.0**	**10.0**
Jun	7.7	**11.3**	8.7	13.2	**2.7**	4.4	15.0	**14.0**
Jul	4.2	**6.3**	8.2	12.7	**1.0**	2.9	**7.5**	14.0
Aug	7.3	**12.0**	7.6	12.5	**2.4**	3.1	13.0	**11.0**
Sep	5.2	**12.8**	5.8	14.4	**2.0**	3.1	**8.7**	**8.7**
Oct	3.4	**10.7**	5.0	16.0	**1.9**	3.2	**5.9**	7.9
Nov	3.0	**11.1**	3.8	14.0	**1.7**	2.2	**5.1**	6.9
Dec	3.2	**12.3**	4.2	15.8	**1.8**	2.3	**5.1**	6.8
Average	4.5	**10.8**	5.3	13.1	**2.0**	2.8	**8.1**	9.3

Note: Bold values are the lowest errors.

The parameterization of elevation suggested by the literature was confirmed by the results reported in this section, thus model D's parameterization (i.e. latitude and longitude as independent variables; elevation as covariate; rainfall data square root transformation) will be used for the standard variables. The following sections will investigate the most suitable units for and transformations of additional variables to be incorporated in the rainfall interpolation process.

3.6.3.1 Topography

The slope and aspect (combined using the relative radiance formula) correlation analyses for the rainfall dataset indicated a definite, but small relationship between a station's topography and the rainfall experienced at the location for April to September when calculating the hillshade grid using a southern azimuth. The next step investigated the possibility of incorporating hillshade's illumination value as an additional variable in the calculation of the rainfall splining function using the models described in Table 3.17.

Table 3.17: Quadvariate model description for the topography (slope and aspect) variable analysis for rainfall

Model	Description
A	Latitude, longitude, elevation and distance (m) are all independent variables
B	Latitude and longitude – independent variables; elevation and hillshade – covariates
C	Latitude and longitude – independent variables; elevation and hillshade – covariates and scale hillshade by multiples of 10
D	Latitude and longitude – independent variables; elevation and hillshade – covariates and logarithmic transformation of hillshade variable

Table 3.18 recounts the results of the spline function calculations when using the models described above. When scaling hillshade's illumination value (with multiples of 10), the spline function calculations produced the same signal, RTGCV and RTMSE output statistics.

The signal values of models B, C and D do not differ significantly, while model A's signal values are considerably higher. When model A was used, ANUSPLIN's grid-creator program the calculated grids' maximum values exceeded the dataset's maximum values by more than 150mm so that model A was deemed unsuitable for the interpolation process. The RTGCV values for models B and C are most often the lowest (April to September to September). Once the known error was removed, the RTMSE values of model D were frequently similar to those of models B and C.

Table 3.18: Spline function statistics for the rainfall quadvariate (topography) models

	Signal				RTGCV ($mm^{1/2}$)				RTMSE ($mm^{1/2}$)			
Model	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Jan	27.4	20.4	20.4	20.2	**0.79**	0.93	0.93	0.94	**0.34**	0.35	0.35	0.36
Feb	24.5	20.1	20.1	20.2	**0.91**	1.00	1.00	1.01	**0.37**	0.38	0.38	0.38
Mar	21.6	22.2	22.2	22.1	**0.93**	0.98	0.98	0.98	**0.36**	0.38	0.38	0.39
Apr	43.9	20.9	20.9	20.9	1.03	**1.02**	**1.02**	1.03	0.50	**0.39**	**0.39**	0.40
May	26.1	24.2	24.2	24.2	1.49	**1.35**	**1.35**	1.36	0.62	**0.55**	**0.55**	**0.55**
Jun	29.3	22.2	22.2	22.2	1.88	**1.64**	**1.64**	1.65	0.82	**0.65**	**0.65**	**0.65**
Jul	81.9	23.7	23.7	23.7	1.58	**1.45**	**1.45**	1.46	0.72	**0.59**	**0.59**	**0.59**
Aug	65.7	21.4	21.4	21.4	1.55	**1.44**	**1.44**	1.45	0.77	**0.56**	**0.56**	**0.56**
Sep	68.0	23.8	23.8	23.7	1.28	**1.23**	**1.23**	1.24	0.63	**0.50**	**0.50**	**0.50**
Oct	18.5	19.4	19.4	19.3	**1.23**	1.24	1.24	1.25	**0.45**	0.46	0.46	0.47
Nov	77.3	34.4	34.4	34.5	**0.84**	0.87	0.87	0.87	**0.40**	**0.40**	**0.40**	**0.40**
Dec	28.9	37.8	37.8	38.2	**0.96**	0.99	0.99	0.99	**0.41**	0.46	0.46	0.47
Average	*42.8*	*24.2*	*24.2*	*24.2*	*1.25*	***1.20***	***1.20***	*1.21*	*0.56*	***0.48***	***0.48***	***0.48***

Note: Bold values are the lowest errors.

Table 3.19 indicates the error statistics produced when the test stations were extracted from the original dataset. The percentage values were calculated to indicate the magnitude of difference between the real values and the interpolated errors since high residuals during the dry season are more significant than during the wet season. Model C most often has the lowest mean, minimum and maximum errors, while model D's error values were slightly higher than those of model C, but lower than models A and B values.

Table 3.19: Error values for the rainfall quadvariate (topography) models

	Mean Error							
Model	**A**		**B**		**C**		**D**	
	mm	**%**	**mm**	**%**	**mm**	**%**	**mm**	**%**
Jan	3.1	13.4	2.6	12.3	2.5	**11.9**	2.5	12.0
Feb	3.3	14.0	2.9	13.7	2.8	**13.3**	2.8	13.3
Mar	4.1	13.1	3.7	12.3	3.3	**11.4**	3.3	**11.4**
Apr	6.7	15.5	4.2	9.8	3.9	**9.2**	3.9	**9.2**
May	10.0	19.6	6.3	11.3	5.9	**10.8**	5.9	**10.8**
Jun	14.0	21.4	7.6	11.8	7.0	**11.1**	7.0	**11.1**
Jul	9.0	12.9	7.0	11.0	6.5	**10.4**	6.5	**10.4**
Aug	12.0	20.3	6.5	10.8	6.1	**10.3**	6.1	**10.3**
Sep	6.5	14.3	5.0	12.4	4.7	**11.9**	4.7	**11.9**
Oct	4.6	14.6	4.3	13.8	4.1	**13.3**	4.1	**13.3**
Nov	4.0	15.0	3.3	**12.3**	3.3	**12.3**	3.3	12.4
Dec	4.2	15.7	3.7	**14.6**	3.7	14.7	3.7	14.7
Average	*6.8*	*15.8*	*4.8*	*12.2*	*4.5*	***11.7***	*4.5*	***11.7***

	Minimum Error (mm)				**Maximum Error (mm)**			
Model	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
	mm	**mm**	**mm**	**mm**	**mm**	**mm**	**mm**	**mm**
Jan	1.1	1.3	**1.2**	**1.2**	9.9	7.5	**7.3**	**7.3**
Feb	1.2	1.3	**1.3**	**1.3**	11.0	10.0	**9.8**	**9.8**
Mar	1.8	2.1	**1.9**	**1.9**	12.0	9.8	**8.8**	**8.9**
Apr	3.1	2.9	**2.7**	**2.7**	17.0	8.2	**7.6**	**7.7**
May	3.4	4.9	**4.5**	4.6	21.0	9.2	**8.4**	8.5
Jun	4.0	4.8	**4.4**	**4.4**	44.0	12.0	**11.0**	**11.0**
Jul	1.9	3.2	**2.9**	**2.9**	31.0	12.0	**10.0**	**11.0**
Aug	2.8	3.2	**2.9**	3.0	31.0	10.0	9.6	**9.5**
Sep	1.9	3.4	**3.1**	**3.1**	27.0	8.9	8.3	**8.2**
Oct	2.2	2.7	**2.6**	**2.6**	14.0	7.9	**7.7**	**7.7**
Nov	1.9	1.8	**1.8**	**1.8**	12.0	6.5	**6.4**	**6.4**
Dec	1.7	2.2	**2.1**	**2.1**	12.0	6.9	**6.7**	**6.8**
Average	***2.3***	*2.8*	*2.6*	*2.6*	*20.2*	*9.1*	***8.5***	*8.6*

Note: Bold values are the lowest errors.

Notice that the mean error values for all the models are less than 14%, except for December. These results are better than those for rainfall interpolation analyses done in other studies (Hutchinson 1998b, Price *et al*. 2000) which state that error values of up to 35% for rainfall grids can be expected. From Tables 3.18 and 3.19 it can be deduced that model C (with the illumination value scaled to 1000) most often provided the best results while model D (logarithmic transformation of hillshade variable) error values were often also low. For this study, model C was selected as the most suitable model for the entire year.

3.6.3.2 Distance to ocean

The correlation analyses indicate a definite, but small positive relationship between a weather station's recorded rainfall and its distance to the nearest ocean for December, February and March and a low to moderate negative correlation from April to September. Table 3.20 describes the models used to investigate the suitable parameterization of the distance to ocean as an additional variable in the rainfall grid interpolation process, while the parameterization for latitude, longitude and elevation were established during the introduction of section 3.6.3.

Table 3.20: Quadvariate model description for the distance to ocean parameter analysis for rainfall

Model	Description
A	Latitude, longitude, elevation and distance (m) are all independent variables
B	Latitude and longitude – independent variables; elevation and distance (m) – covariates.
C	Latitude and longitude – independent variables, elevation and distance – covariates, scale distance to kilometres
D	Latitude and longitude – independent variables; elevation and distance – covariates, logarithmic transformation of distance variable
E	Latitude and longitude – independent variables; elevation and distance – covariates, distance scaled to kilometres and then the logarithmic transformation of the distance variable

Model A only implements independent variables while model B takes note of recommendations in the literature that whenever a quadvariate model is used, all variables in addition to latitude and longitude should be incorporated as covariates. It was found that model A over-estimates the maximum rainfall values for the study region, thus corroborating the recommendations in the literature (Hutchinson 1998a, 1998b).

The spline function statistics produced by the models in Table 3.20 are presented in Table 3.21 with the lowest error values printed in bold. The signal, RTGCV and RTMSE values for models B and C are identical – an indication that scaling the additional distance to ocean parameter to kilometres has no effect on the interpolation process. Notice that the differences between the RTMSE and RTGCV values are insignificant for models B, C and D (generally not more than $0.02mm^{1/2}$). The RTGCV and RTMSE values of model E are most often higher than those of models B, C and D. Thus unlike, the temperature grids where smaller error values were obtained when logarithmic transformation was applied to the distance parameter, this is not always the case when applied to the rainfall spline calculations.

Table 3.21: Spline function statistics for the rainfall quadvariate (distance to ocean) models

	Signal					RTGCV (mm$^{1/2}$)					RTMSE (mm$^{1/2}$)				
Model	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Jan	16.9	19.1	19.1	18.8	18.9	1.08	1.01	1.01	**1.00**	1.02	0.38	**0.37**	**0.37**	**0.37**	0.38
Feb	17.6	18.7	18.7	17.6	17.9	1.05	**1.05**	**1.05**	**1.05**	1.07	**0.38**	0.39	0.39	**0.38**	0.39
Mar	16.3	20.0	20.0	19.7	20.6	1.07	**1.03**	**1.03**	**1.03**	1.05	**0.37**	0.39	0.39	0.39	0.40
Apr	17.8	18.6	18.6	18.0	19.1	1.16	1.08	1.08	**1.07**	1.09	0.42	0.40	0.40	**0.39**	0.41
May	17.6	21.8	21.8	22.9	23.5	1.60	**1.37**	**1.37**	1.39	1.39	0.57	**0.54**	**0.54**	0.55	0.56
Jun	17.6	20.9	20.9	21.5	21.7	1.97	**1.67**	**1.67**	1.68	1.69	0.71	**0.64**	**0.64**	0.65	0.66
Jul	16.8	21.8	21.8	23.4	23.5	1.88	**1.49**	**1.49**	**1.49**	1.51	0.66	**0.58**	**0.58**	0.60	0.61
Aug	16.5	20.5	20.5	20.8	21.2	1.77	**1.49**	**1.49**	1.50	1.51	0.62	**0.57**	**0.57**	0.58	0.58
Sep	20.4	21.0	21.0	21.9	24.4	1.50	**1.30**	**1.30**	1.29	1.32	0.57	**0.50**	**0.50**	0.51	0.54
Oct	15.2	18.4	18.4	17.6	18.8	1.36	**1.29**	**1.29**	**1.29**	1.32	0.46	0.47	0.47	**0.46**	0.49
Nov	15.0*	65.2	65.2	62.7	68.3	–	0.89	0.89	0.89	**0.87**	–	**0.44**	**0.44**	**0.44**	0.43
Dec	15.0*	74.2	74.2	71.3	78.7	–	0.95	0.95	**0.89**	**0.89**	–	0.46	0.46	0.43	**0.42**
Avg	*16.9*	*28.4*	*28.4*	*28.0*	*29.7*	*1.42*	***1.24***	***1.24***	***1.24***	*1.25*	*0.50*	***0.49***	***0.49***	***0.49***	*0.50*

Notes: Bold values are the lowest errors. Asterisks indicate signal values that are too high or too low

By extracting the predetermined test stations the error values tabulated in Table 3.22 were produced. In this case the average mean error for models B, C, D and E are all less than 12%. Model D has the lowest mean errors for ten out of the twelve months of the year. The minimum and maximum errors for models B and C are often lower than those of the other models. Model E's error values were generally higher than those of models B, C and D. Notice that the highest maximum error for models B, C, D and E is no more than 12mm for all twelve months of the year.

Table 3.22: Error values for the rainfall quadvariate (distance to ocean) models

	Mean Error									
Model	**A**		**B**		**C**		**D**		**E**	
	mm	**%**	**mm**	**%**	**mm**	**%**	**mm**	**%**	**mm**	**%**
Jan	2.8	13.9	2.3	**11.2**	2.3	**11.2**	2.5	11.8	2.4	11.7
Feb	2.9	13.4	2.1	10.3	2.1	10.3	2.0	**9.5**	2.0	9.5
Mar	3.6	12.2	2.7	**9.7**	2.7	**9.7**	3.0	10.5	2.3	8.2
Apr	4.3	10.4	4.1	9.8	4.1	9.8	3.9	**9.1**	4.1	9.7
May	6.4	12.7	6.2	11.4	6.2	11.4	6.3	**11.3**	6.5	11.6
Jun	8.0	13.7	7.6	12.0	7.6	12.0	7.6	**11.7**	7.8	12.1
Jul	7.3	12.9	6.8	11.0	6.8	11.0	7.0	**10.9**	7.2	11.3
Aug	6.7	12.5	6.6	11.2	6.6	11.2	6.6	**10.9**	6.8	11.3
Sep	4.9	13.7	4.8	12.5	4.8	12.5	4.8	**12.0**	5.2	13.0
Oct	3.9	13.0	4.0	13.3	4.0	13.3	4.0	**12.8**	4.4	14.2
Nov	–	–	3.5	13.4	3.5	13.4	2.9	**11.0**	3.7	13.9
Dec	–	–	3.9	15.7	3.9	15.7	3.9	**15.2**	4.0	15.7
Average	*4.7*	*12.8*	*4.6*	*11.7*	*4.6*	*11.7*	*4.5*	***11.2***	*4.7*	*11.8*

Note: Bold values are the lowest errors.

Table 3.22 (Continued): Error values for the rainfall quadvariate (distance to ocean) models

	Minimum Error (mm)					Maximum Error (mm)				
Model	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Jan	1.4	**1.3**	**1.3**	1.4	1.4	9.9	**7.2**	**7.2**	7.9	7.6
Feb	1.3	1.5	1.5	**1.1**	**1.1**	12.0	8.6	8.6	**8.1**	8.3
Mar	1.8	1.8	1.8	1.7	**1.5**	12.0	**8.2**	**8.2**	8.5	7.3
Apr	**2.7**	2.9	2.9	2.8	3.1	10.0	8.5	8.5	**8.3**	8.7
May	4.5	**4.4**	**4.4**	4.7	4.9	12.0	9.3	9.3	**9.1**	9.3
Jun	4.7	**4.3**	**4.3**	4.7	4.8	17.0	13.0	13.0	**12.0**	13.0
Jul	4.0	**2.7**	**2.7**	3.1	3.2	16.0	**12.0**	**12.0**	**12.0**	**12.0**
Aug	3.8	**2.9**	**2.9**	3.2	3.3	13.0	**10.0**	**10.0**	**10.0**	**10.0**
Sep	3.6	**2.8**	**2.8**	3.2	3.5	9.3	**7.5**	**7.5**	8.4	8.9
Oct	2.4	2.9	2.9	**2.8**	2.9	8.6	**7.4**	**7.4**	8.0	8.5
Nov	–	**2.1**	**2.1**	1.9	2.2	–	**7.5**	**7.5**	6.7	7.7
Dec	–	**2.3**	**2.3**	**2.3**	2.4	–	7.9	7.9	**7.6**	8.0
Average	*2.8*	***2.6***	***2.6***	*2.7*	*2.9*	*11.4*	***8.9***	***8.9***	***8.9***	*9.2*

Note: Bold values are the lowest errors.

From the results presented in this section, model D (i.e. latitude and longitude – independent variables; elevation and distance – covariates, logarithmic transformation of distance variable) can be singled out as the most appropriate quadvariate model when incorporating distance as the additional variable since it produced the lowest error values overall. The next step is to determine if it is possible to incorporate topography together with distance as two additional variables in the interpolation process.

3.6.3.3 Combining topography and distance to ocean variables

After identifying the best parameterization for the two additional variables the next step is to analyse the possibility of incorporating both the variables using a five-variable model. The five-variable model comprises latitude, longitude, elevation, distance and the illumination values using the parameterization of the best two quadvariate models established in sections 3.6.3.1 and 3.6.3.2 and is designated as the quinvariate model in the following tables. Table 3.23 sets out the spline function statistics for the various models available for the rainfall interpolation process. The first quadvariate model (Quad-1) incorporates distance to nearest coastline, while the second quadvariate model (Quad-2) incorporates topography.

Since third-order spline functions were used for the quadvariate models it was necessary to investigate the possibility of using a fourth-order spline function for the five-variable model. The output error statistics produced during this investigation indicated that a third-order spline function is most suitable for this model since unsuitable signal values were obtained and the maximum values of the climate grids produced are over-estimated.

Table 3.23: Spline function statistics for the trivariate (Triv) model, quadvariate models (distance: Quad-1 and topography: Quad-2) and the quinvariate model (Quin) for rainfall

	Signal				RTGCV (mm^2)				RTMSE (mm^2)			
Model	**Triv**	**Quad-1**	**Quad-2**	**Quin**	**Triv**	**Quad-1**	**Quad-2**	**Quin**	**Triv**	**Quad-1**	**Quad-2**	**Quin**
Jan	91.7	18.8	20.4	90.4	**0.83**	1.00	0.93	0.91	**0.30**	0.37	0.35	0.38
Feb	32.2	17.6	20.1	19.2	**0.93**	1.05	1.00	0.99	0.39	0.38	0.38	**0.37**
Mar	73.4	19.7	22.2	21.1	**0.95**	1.03	0.98	**0.95**	0.46	0.39	0.38	**0.37**
Apr	26.9	18.0	20.9	19.4	**1.00**	1.07	1.02	**1.00**	0.46	0.39	0.39	**0.37**
May	46.9	22.9	24.2	23.8	**1.18**	1.39	1.35	1.36	0.56	**0.55**	**0.55**	**0.55**
Jun	31.2	21.5	22.2	22.4	**1.45**	1.68	1.64	1.64	0.66	**0.65**	**0.65**	**0.65**
Jul	61.7	23.4	23.7	23.9	**1.26**	1.49	1.45	1.44	**0.45**	0.60	0.59	0.58
Aug	35.0	20.8	21.4	21.3	**1.35**	1.50	1.44	1.44	0.67	0.58	**0.56**	**0.56**
Sep	57.4	21.9	23.8	21.8	**1.13**	1.29	1.23	1.20	**0.40**	0.51	0.50	0.47
Oct	26.0	17.6	19.4	18.0	**1.21**	1.29	1.24	**1.21**	0.52	0.46	0.46	**0.44**
Nov	68.1	62.7	34.4	35.1	**0.76**	0.89	0.87	0.86	**0.25**	0.44	0.40	0.40
Dec	70.6	71.3	37.8	74.9	**0.88**	0.89	0.99	0.91	**0.40**	0.43	0.46	0.43
Avg	*51.8*	*28.0*	*24.2*	*32.6*	***1.10***	*1.24*	*1.20*	*1.18*	***0.47***	*0.49*	*0.48*	***0.47***

Note: Bold values are the lowest errors.

The signal values for the trivariate model were generally higher than those of the other models. The trivariate model's RTGCV values were the lowest for all the months of the year, but this was not the case for its RTMSE values where the quinvariate model often had the lowest values. Between the two quadvariate models the model incorporating topography had better signal values (closer to 40) as well as lower RTGCV and RTMSE error values, and when higher, it was not by more than $0.1mm^{1/2}$.

The error statistics obtained when extracting the test stations is provided in Table 3.24. This provides a useful indication of the accuracy of the various models. Unlike the quinvariate model which predominantly had the lowest RTMSE values, no model could be identified as having the lowest error values for a predominant part of the year. Both the quadvariate models produced lower error values than the quinvariate model for seven out of the twelve months when the test stations were withheld from the dataset.

The RTMSE values of the quadvariate-2 model are frequently equal to or lower than those of the quadvariate-1 model, while the former's signal values were also higher. In addition to the lower RTMSE values the quadvariate-2 model has lower mean, minimum and maximum error values for May to August after removal of the selected test stations.

Table 3.24: Error values for the trivariate (Triv) model, quadvariate models (distance: Quad-1 and topography: Quad-2) and the quinvariate model (Quin) for rainfall

	Mean Error							
Model	**Triv**		**Quad-1**		**Quad-2**		**Quin**	
	mm	**%**	**mm**	**%**	**mm**	**%**	**mm**	**%**
Jan	3.0	14.2	2.5	**11.8**	2.5	11.9	2.7	12.2
Feb	3.0	14.2	2.0	**9.5**	2.8	13.3	2.8	13.1
Mar	3.8	12.7	3.0	**10.5**	3.3	11.4	3.6	11.8
Apr	4.0	**8.9**	3.9	9.1	3.9	9.2	4.1	9.4
May	5.9	**10.1**	6.3	11.3	5.9	10.8	6.5	11.5
Jun	7.7	11.3	7.6	11.7	7.0	**11.1**	7.9	12.1
Jul	4.2	**6.3**	7.0	10.9	6.5	10.4	7.2	11.1
Aug	7.3	12.0	6.6	10.9	6.1	**10.3**	6.8	11.1
Sep	5.2	12.8	4.8	12.0	4.7	11.9	4.8	**11.7**
Oct	3.4	**10.7**	4.0	12.8	4.1	13.3	4.0	12.5
Nov	3.0	11.1	2.9	11.0	3.3	12.3	2.9	**10.7**
Dec	3.2	**12.3**	3.9	15.2	3.7	14.7	4.0	15.4
Average	*4.5*	***10.8***	*4.5*	*11.2*	*4.5*	*11.7*	*4.7*	*11.3*

	Minimum Error (mm)				**Maximum Error (mm)**			
Model	**Triv**	**Quad-**	**Quad-2**	**Quin**	**Triv**	**Quad-1**	**Quad-2**	**Quin**
Jan	1.4	1.4	**1.2**	1.3	**5.7**	7.9	7.3	7.8
Feb	**1.1**	**1.1**	1.3	1.3	**7.2**	8.1	9.8	9.8
Mar	1.7	**1.5**	1.9	2.0	**7.5**	8.5	8.8	9.5
Apr	2.8	3.1	**2.7**	2.8	**6.0**	8.3	7.6	8.0
May	4.7	4.9	**4.5**	4.6	10.0	9.1	**8.4**	9.9
Jun	4.7	4.8	**4.4**	4.6	15.0	12.0	**11.0**	13.0
Jul	3.1	3.2	**2.9**	3.0	**7.5**	12.0	10.0	12.0
Aug	3.2	3.3	**2.9**	3.1	13.0	10.0	**9.6**	12.0
Sep	3.2	3.5	3.1	**2.9**	8.7	8.4	**8.3**	9.6
Oct	2.8	2.9	**2.6**	**2.6**	**5.9**	8.0	7.7	8.4
Nov	1.9	2.2	1.8	**1.7**	**5.1**	6.7	6.4	6.4
Dec	2.3	2.4	**2.1**	2.3	**5.1**	7.6	6.7	7.7
Average	*2.7*	*2.9*	**2.6**	**2.6**	***8.1***	*8.9*	*8.5*	*9.3*

Note: Bold values are the lowest errors.

The maximum values produced after the grids were created for the quinvariate model significantly exceeded the dataset's maximum values (by more than 100% compared to the maximum recorded value). Although various combinations of the parameterization of the two additional variables were investigated using models from Tables 3.17 and 3.20, this over-estimation problem could not be solved. The over-estimation is not due to the order of the spline function since this was also the case when the model was run using second- and fourth-order spline functions. Evidently it is not possible to combine all the variables to calculate the rainfall grids using the models described in this project.

3.6.4 Summary of the parameterization investigations

The parameterization investigation of the additional variable for minimum and maximum temperatures found that adding distance to ocean as a variable increased the accuracy of the interpolation process compared to the trivariate models. It was established that elevation and distance had to be incorporated as covariates, but that scaling the distance unit (calculated in metres) to kilometres had no effect on the accuracy. Logarithmic transformation of the distance variable provided lower error values. In the maximum and minimum temperature analyses, mean errors of less than 0.5^{o}C were obtained when the selected test stations were isolated from the dataset.

As with the application of logarithmic transformation to the distance variable in the temperature grids, improved results were also obtained when this was done for the rainfall calculations. This confirms that the additional distance to ocean parameter can be successfully incorporated in the interpolation process when logarithmic transformation is applied as an additional variable.

Adding topography, using the hill-shading to combine aspect and slope, to the rainfall interpolation was also done successfully. In this case the best results were obtained when the hillshade parameter was scaled by 1000, and not when logarithmic transformation was applied. Comparison of the two rainfall quadvariate models indicated that the model incorporating topography is more accurate.

The error analyses showed that the two quadvariate models were not always more accurate than the trivariate model except for the RTMSE values of the hillshade model which was generally lower than the trivariate model during the cooler months of the year.

The quinvariate model (incorporating both hillshade and distance to ocean as additional parameters), produced low error values, but the final grids' maximum values exceeds the dataset's maximum values by far (more than 50%) and could not be used, therefore indicating that combining hillshade and distance to ocean using various models was not done successfully during this project.

Table 3.25 indicates the most appropriate units of and transformations for latitude, langitude, elevation and the additional variables (distance and topography). Notice that the models described for each climate variable is used for all the months of the year.

Table 3.25: Parameterization for the interpolation process for the three climate variables

Maximum Temperature	Latitude and longitude (degrees) as independent variables; elevation and distance (both scaled to kilometres) as covariates, with the logarithmtic transformation of distance (A x log (x + B))
Minimum Temperature	Latitude and longitude (degrees) as independent variables; elevation and distance (both scaled to kilometres) as covariates
Rainfall	Latitude and longitude (degrees) as independent variables; elevation (scaled to kilometres) as covariate and hillshade as covariate and scaled hillshade by a 1000.

Now that the most suitable models have been selected from the models described in Table 3.17 and Table 3.20, the final objective of this study is to compare existing climate grids with climate grids created by ANUSPLIN where the additional variables have been incorporated (using all the available data points).

3.7 COMPARISON OF ANUSPLIN-CREATED GRIDS WITH EXISTING GRIDS

The next step is to compare the monthly mean daily maximum and minimum temperatures and the mean monthly rainfall grids created using the models selected in the previous section with the appropriate existing datasets using all the available weather station's data. The most relevant climate data for the Western Cape is contained in the *South African atlas of agrohydrology and climatology* produced under leadership of Prof. Schulze in 1997 and subsequently updated (Lynch 1999; 2003; Shulze 1997; Schulze & Maharaj 2004).

The *South African atlas of agrohydrology and climatology* climate data (from now on called the ATLAS grids) are available in digital format as grids at a resolution of 1.6km x 1.6km and can be viewed and analysed in a GIS environment enabling comparison with the ANUSPLIN grids. Due to the difference in resolution, the ATLAS data were first re-sampled to the resolution of the ANUSPLIN grids (90m x 90m) making use of ArcInfo's nearest neighbour re-sampling function. The ATLAS grids were then subtracted from the ANUSPLIN grids in order to calculate the differences. The product is a grid illustrating the range of differences between the two different datasets.

The spline interpolation method creates a surface whereby the exactness of fit is determined by the degree of freedom of the splining function. The grids created by ANUSPLIN are therefore true to the actual data provided for a station. In the foregoing analysis the accuracy of the ANUSPLIN grids was calculated to be within 0.5^{o}C for both maximum and minimum temperatures using the quadvariate models. This ensures that the ANUSPLIN grids are all within half a degree Celsius of accuracy using the dataset provided by SAWS and ARC. The rainfall interpolation process analysis noted that the topography quadvariate model had average error values of 4.5mm or 11.7% for the twelve months of the year. Thus ANUPSLIN's rainfall grid-creation process is reasonably accurate compared to the results of other rainfall interpolation studies. The purpose of this section is to assess how much the ATLAS grids differ from SAWS and ARC datasets provided for this study and, in turn, from the grids created by ANUSPLIN. Monthly mean daily maximum temperature, monthly mean daily minimum temperature and monthly mean rainfall will be considered in turn in the following three subsections.

3.7.1 Monthly mean daily maximum temperature

Table 3.26 presents the mean value of the grids produced when the ANUSPLIN and ATLAS monthly mean daily maximum temperature grids were subtracted from each other for each month of the year. The percentage values were calculated using the difference value and the ANUSPLIN grid's mean temperature value in order to illustrate the magnitude of the deviation. Notice that the October to February (i.e. 'warmer' months) have the lowest variation while March to September (i.e. 'cooler' months) recorded the largest differences.

Table 3.26: Mean values of grids produced when the ATLAS grid was subtracted from the ANUSPLIN grid for maximum temperature

Month	Difference		Month	Difference		Month	Difference	
	(°C)	(%)		(°C)	(%)		(°C)	(%)
Jan	0.29	1.00	**May**	0.36	1.94	**Sep**	0.89	3.70
Feb	0.28	1.05	**Jun**	0.69	4.01	**Oct**	0.28	1.01
Mar	0.58	2.42	**Jul**	0.78	4.50	**Nov**	0.20	0.67
Apr	0.48	2.23	**Aug**	0.47	2.31	**Dec**	0.31	1.03
						Averag	**0.47**	**1.97**

Figure 3.9 shows the spatial variation of temperature for January which represents the spatial pattern for the warmer months. For the period September to February ATLAS produced much higher temperatures along the coastal region of the western part of the of the study area. This can be due to a paucity of weather stations or the distance to nearest ocean variable having less weight during the interpolation process implemented in the ATLAS calculations.

During the ten months August to May, higher temperatures were calculated by the ATLAS project for the higher parts of the South Coast mountain range (Outeniqua) and the escarpment area west of Sutherland. The stations situated in the Outeniqua Mountains are fairly low in altitude, thus few data points were available for the calculation of the temperature lapse rate for this region, while it is evident from the dataset that Sutherland's temperature is lower than its surrounding areas.

The ATLAS calculations produced much lower temperatures for the region southwest of Beaufort West and the low-lying parts of the areas farther from the coast in the southwestern part of the study region during October to January. ATLAS also calculated lower temperatures throughout the year for the northern regions of the study area.

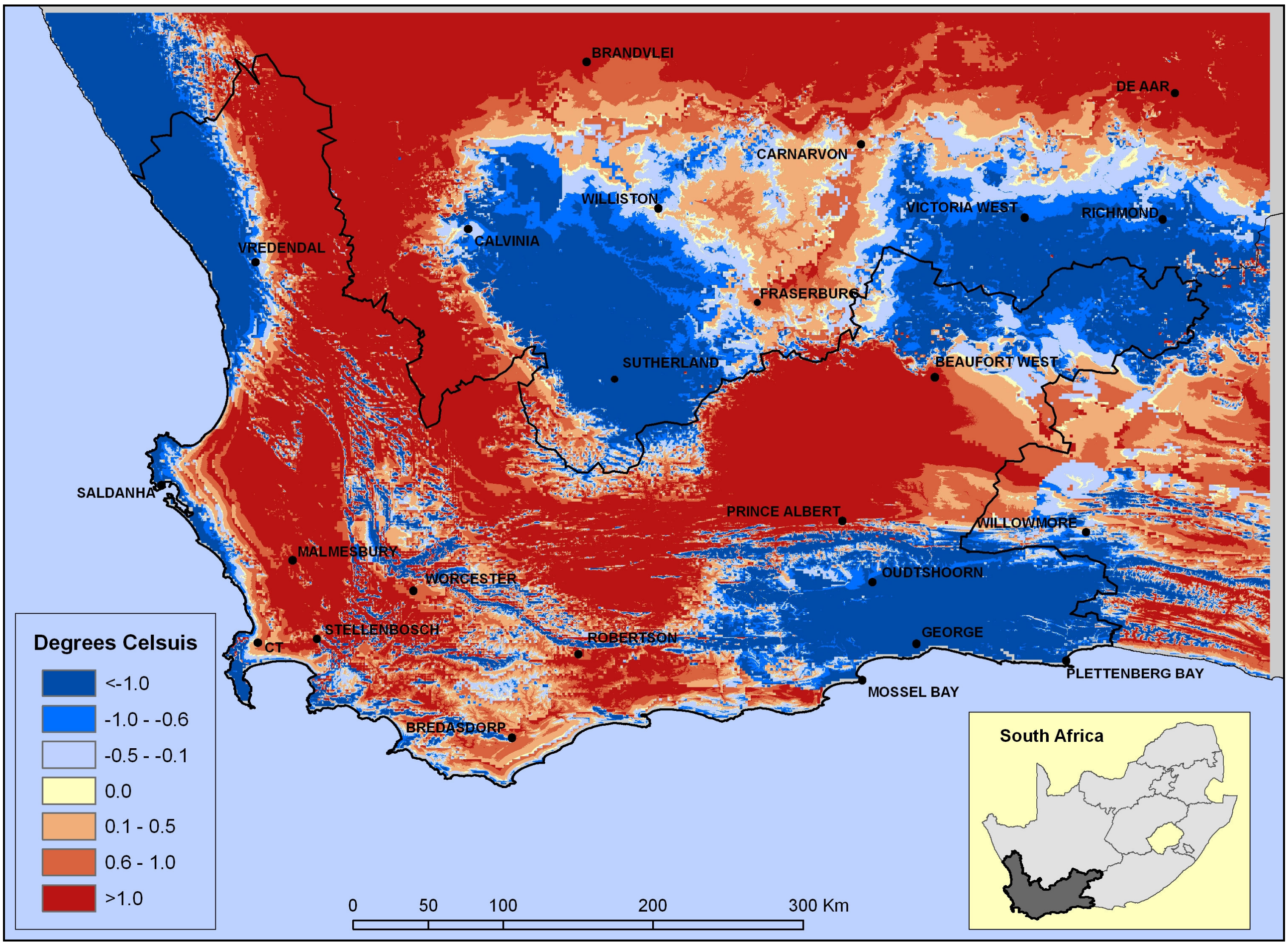


Figure 3.9: Difference in maximum temperatures between ANUSPLIN and ATLAS grids during January

Although the mean surface value differs by less than 0.5°C, the maps demonstrate large areas where the difference between the two grids for the 'warmer' months is more than 1°C. For the worst case scenario this indicates that the ATLAS grids deviate from the dataset used to create the ANUSPLIN grids by more than ± 1.5°C.

Figure 3.10 shows the comparisons for June which represents the spatial pattern of minimum temperature for the 'cooler' months of the year. Compared to the 'warmer' months, the 'cooler' months feature less spatial variation, while the differences between the mean surface temperatures are actually higher according to Table 3.26.

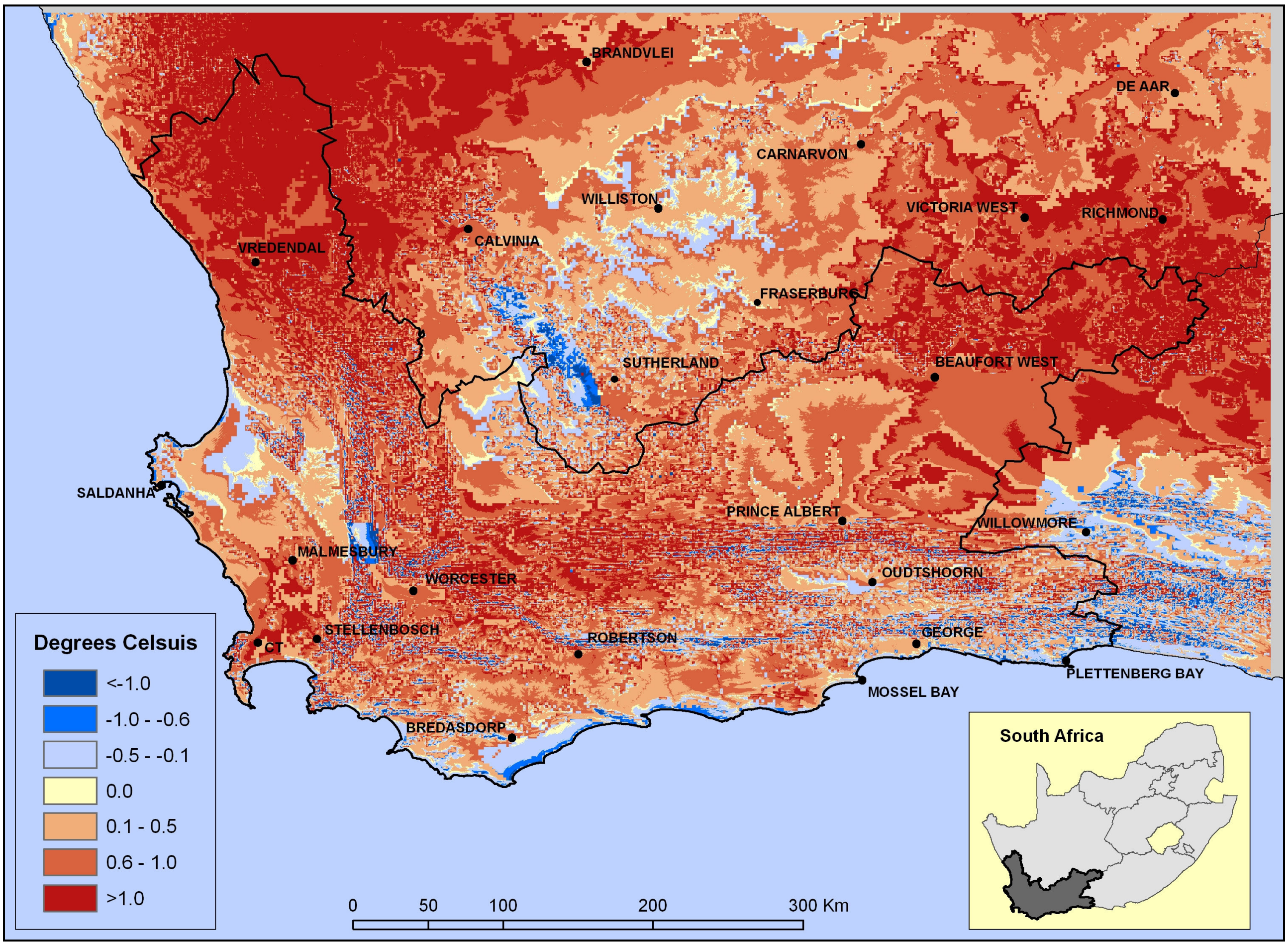


Figure 3.10: Difference in maximum temperatures between ANUSPLIN and ATLAS grids during June

The most prominent area where ATLAS calculated lower temperatures is in the northwestern region where few weather stations occur. Apart from small areas throughout the study region where ATLAS calculate lower temperatures (>1.0°C) compared to the ANUSPLIN maps, the temperature variation over the majority of the study area is positive, most often less than 1.0°C. To a lesser degree, higher temperatures were interpolated by ATLAS around the Sutherland plateau and the mountain ranges running north-south and east-west parallel to the coastlines.

The discrepancies between the two datasets can be attributed to a number of reasons. Firstly, the weather station dataset used for the ANUSPLIN project ranged over a period of 10 to 35 years, while the ATLAS datasets made use of a suite of infilling techniques in order to work with a dataset that spanned 50 years (Lynch 2003). Secondly, different methods to calculate temperature lapse rates since few weather stations are located high up in the mountains, thus providing little data from which more accurate equations can be calculated. In addition, the difference in spatial resolution between the two underlying DEMs can have an impact on the results. The underlying elevation grid used to create the ATLAS grids has a lower resolution and is therefore "smoother" than the grid used during the ANUSPLIN calculations.

3.7.2 Monthly mean daily minimum temperature

Table 3.27 records the monthly mean daily minimum temperature differences between the two datasets. Compared to the maximum temperature differences, it manifests that ANUSPLIN's minimum temperatures are considerably higher than the ATLAS values. Similar to the comparison of maximum temperatures, the greatest differences occur during the 'cooler' months of the year, while the lesser differences are recorded during the 'warmer' months. Percentage-wise the differences are also larger, since the mean surface temperatures are lower.

Table 3.27: Mean values of grids produced when the ATLAS grid was subtracted from the ANUSPLIN grid for minimum temperature

Month	Difference		Month	Difference		Month	Difference	
	(°C)	(%)		(°C)	(%)		(°C)	(%)
Jan	0.85	2.17	**May**	0.62	8.06	**Sep**	0.85	8.76
Feb	0.55	2.45	**Jun**	0.62	19.54	**Oct**	0.86	2.13
Mar	0.69	6.41	**Jul**	0.69	18.57	**Nov**	0.76	1.37
Apr	1.07	7.24	**Aug**	0.66	6.98	**Dec**	0.78	2.12
						Averag	**0.75**	**5.04**

Figure 3.11 illustrates the spatial differences of minimum temperature during the 'warmer' months as characterized by January's variation. Except for the Outeniqua Mountain range area between Oudtshoorn and George, the November to March values display the same pattern as the maximum temperature variations illustrated in Figure 3.9. Compared to the ANUSPLIN grids, the minimum temperature grids created by ATLAS reveal higher temperatures for the coastal region of the western part of the study area and for the region surrounding Sutherland. ATLAS also produced higher temperatures along the low- and high-lying mountain areas (as opposed to the mountain slopes) of the southwestern part of the Western Cape and the area north and northeast of Plettenberg Bay.

The ATLAS grids evince lower minimum temperatures over the large part of the study area, especially in the northeastern and northwestern regions and the area north of Prince Albert. According to Table 3.27 the ATLAS and ANUSPLIN minimum temperature grids differ considerably as confirmed by Figure 3.11. Major parts of the map exhibit minimum temperature differences above 1^{o}C.

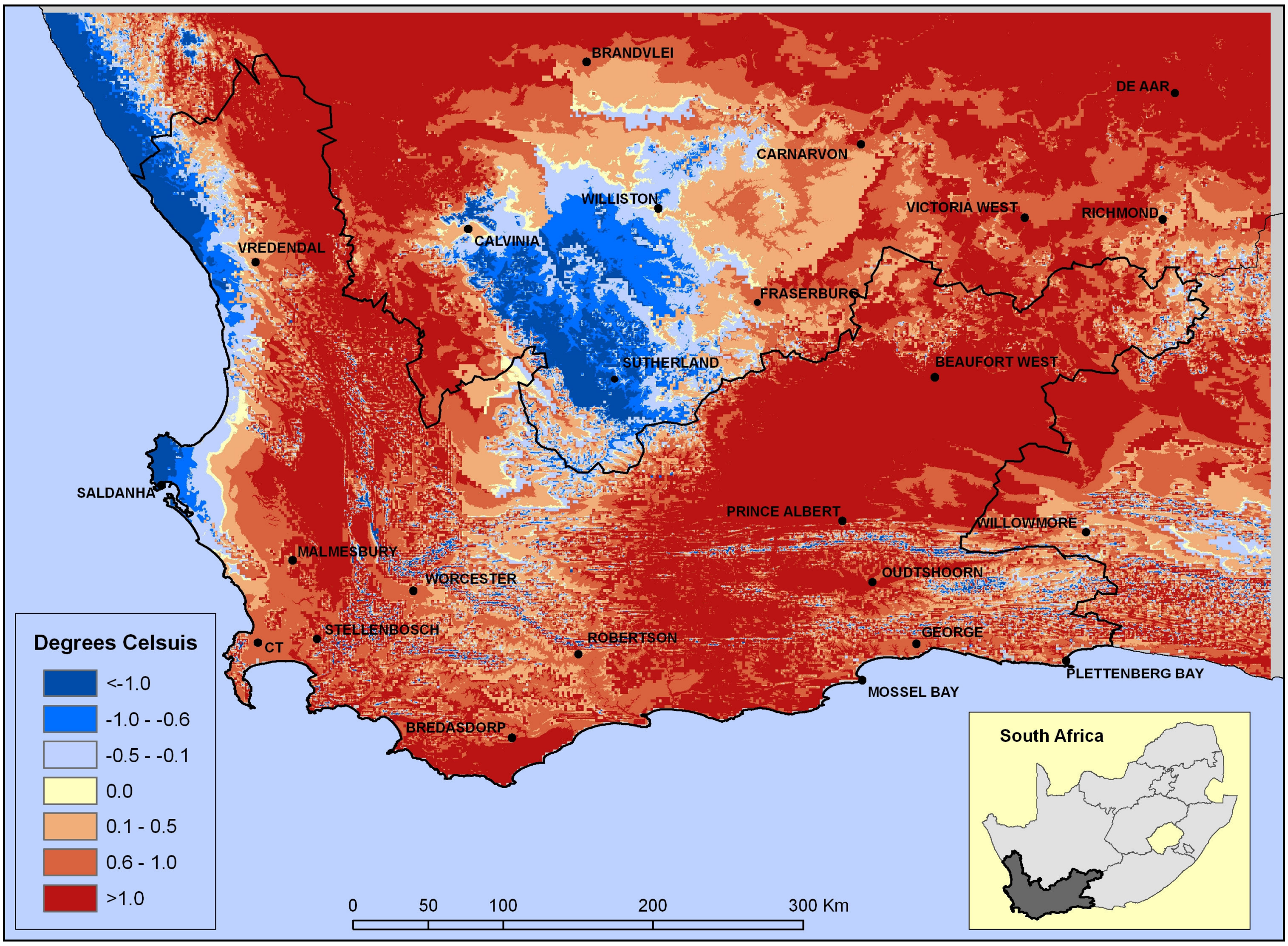


Figure 3.11: Difference in minimum temperatures between ANUSPLIN and ATLAS grids during January

Figure 3.12 reveals the variation of the minimum temperature between the two datasets for the ‘cooler’ months of the year as represented by June temperature variations. During the ‘cooler’ months ATLAS displays notably lower minimum temperature values (>1.0^{o}C) along the western coastal areas, as well as the southeastern and northeastern parts of the study area.

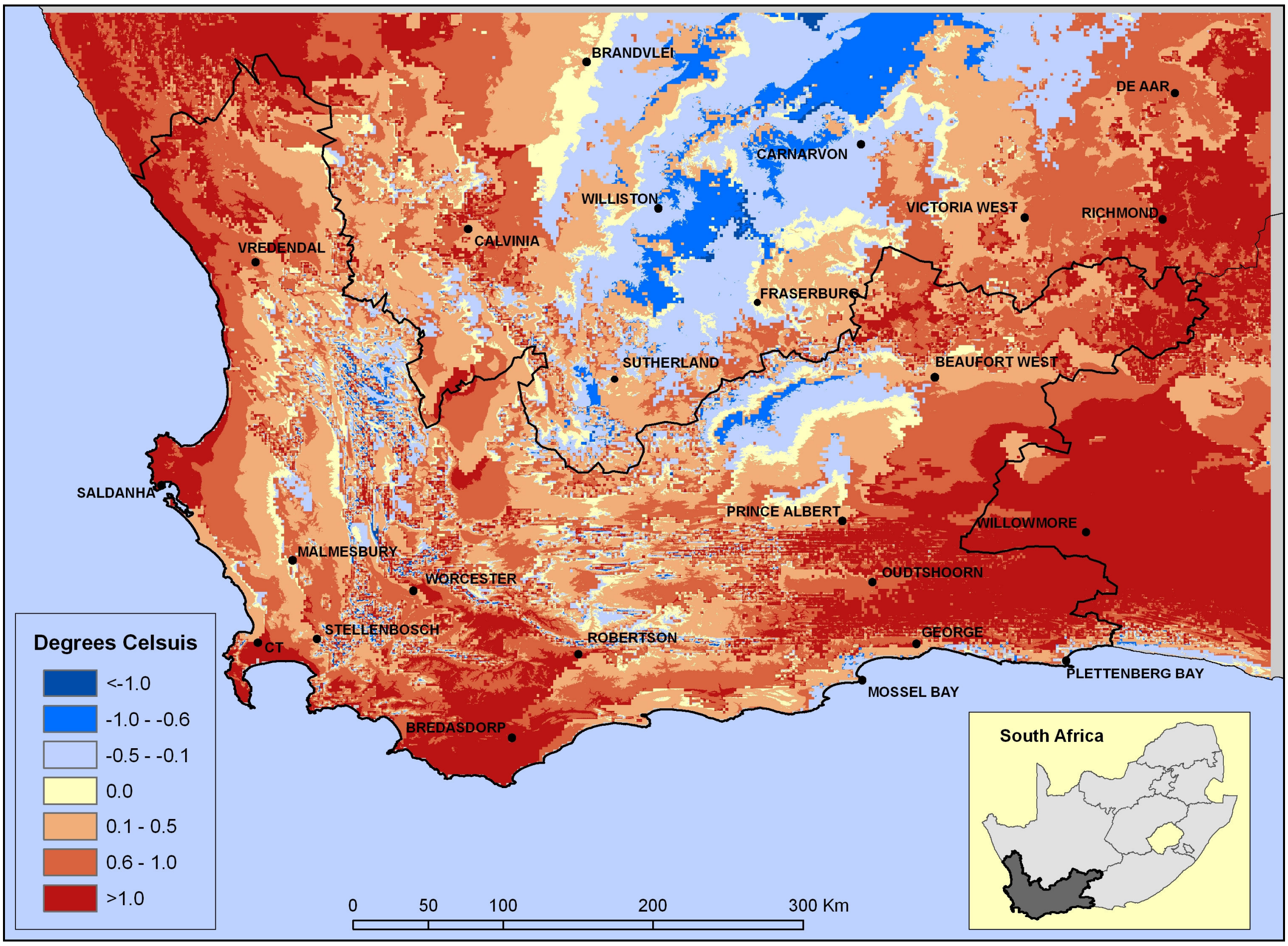


Figure 3.12: Difference in minimum temperatures between ANUSPLIN and ATLAS grids during June

The minimum temperature values were higher for the ATLAS dataset over a large part of the northern reaches of the study area. The differences were generally less than 1.0^{o}C, but greater than 0.5^{o}C for this area. ATLAS also produced higher minimum temperatures along the high- and low-lying areas of the southwestern Cape, emphasizing the different temperature lapse rates calculated by the two different interpolation processes or because of the different underlying elevation grids. This phenomenon also manifested in the maximum temperature comparisons.

3.7.3 Mean monthly rainfall

The model selected for the rainfall interpolation process is the quadvariate model which includes topography as an additional variable. Table 3.28 presents the mean value of the grids produced when the ATLAS grids are subtracted from the ANUSPLIN grids. On average the ANUSPLIN grids have a higher rainfall than the ATLAS grids while January has the greatest difference and May has the smallest difference percentage-wise.

Table 3.28: Mean values of grids produced when the ATLAS grid was subtracted from the ANUSPLIN grid for monthly total rainfall

Month	Difference		Month	Difference		Month	Difference	
	(°C)	(%)		(°C)	(%)		(°C)	(%)
Jan	4.13	18.12	**May**	3.36	11.60	**Sep**	2.99	14.24
Feb	4.13	15.95	**Jun**	7.26	22.13	**Oct**	6.16	24.53
Mar	5.68	16.47	**Jul**	6.98	23.27	**Nov**	4.64	19.59
Apr	9.04	26.68	**Aug**	6.57	21.75	**Dec**	5.17	24.27
						Average	**4.13**	**15.02**

Figure 3.13 illustrates the spatial variation for rainfall for November to March as represented by January. The effect of aspect is evident in the north-facing slopes registering less rain than the southern slopes. Thus ANUSPLIN is able to successfully incorporate topography by using hillshading's illumination value as an additional variable.

The areas subject to rain during the summer months show greater differences compared to the winter rainfall areas. For the region surrounding Willowmore, ATLAS estimated lower rainfall (>10mm) than did ANUSPLIN. Higher rainfall were interpolated for the mountainous areas north of Plettenberg Bay and George, and those east of Robertson and southwest of Stellenbosch by the ATLAS interpolation process.

Apart from the regions described in previous paragraph, Figure 3.13 shows that large parts of the study area's deviation in rainfall are between 0mm and 10mm. For the worst case scenario, this tells one that the ATLAS maps differ approximately 15mm from the dataset used during the project (in the warmer months of the year).

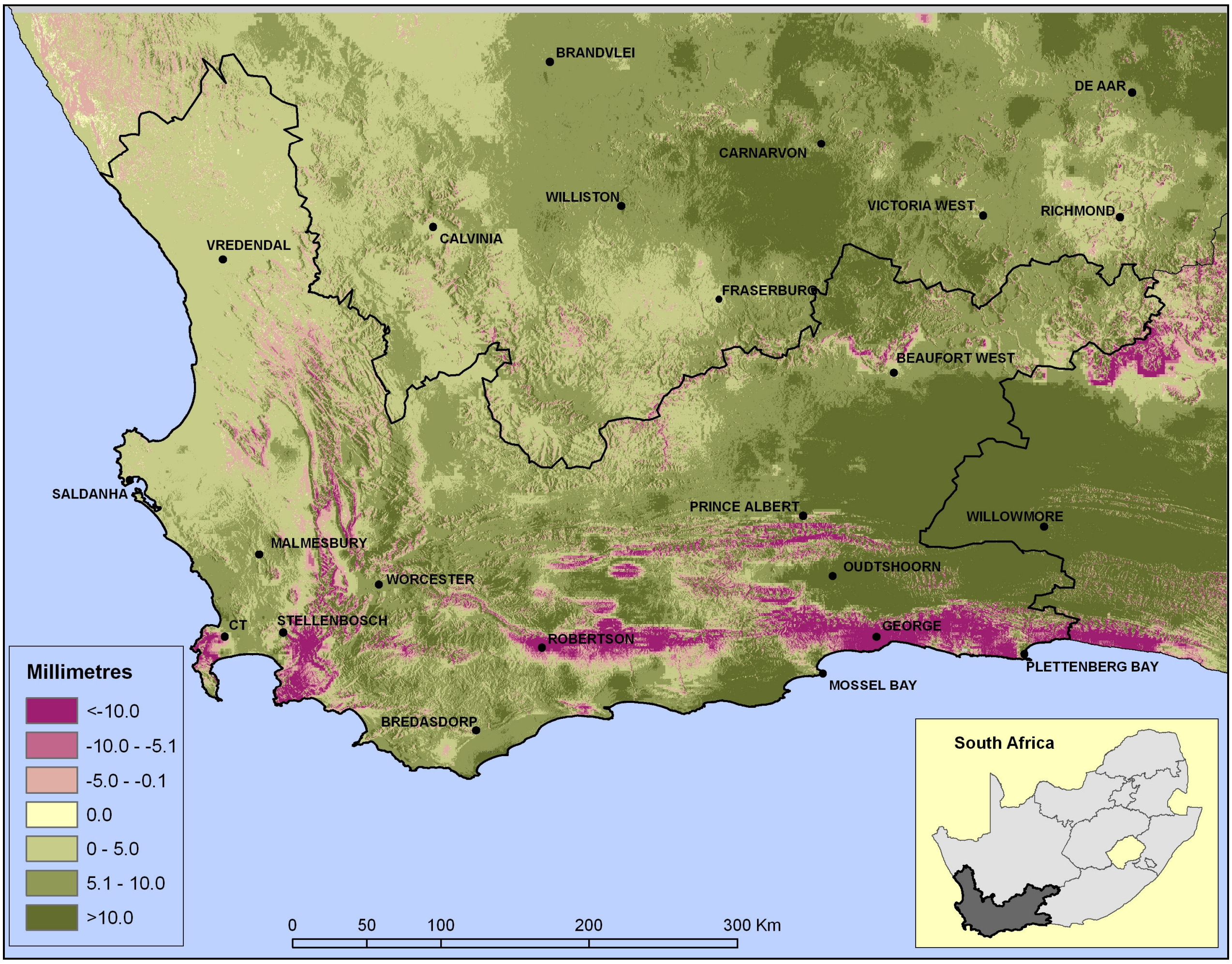


Figure 3.13: Difference in rainfall between ANUSPLIN and ATLAS grids during January

Figure 3.14 portrays the spatial variations between the ANUSPLIN and ATLAS datasets for April to October (cooler months) as represented by June. The patterns of deviations are similar to that of November to March featured in Figure 3.13. For the cooler months the southern part of the study area indicates noteworthy variation between the two datasets especially in the region which receives rainfall during the winter. ATLAS predicts more rainfall for the northern slopes of the Cape Peninsula mountains and the mountains southeast of Stellenbosch.

Over a large part of the interior, which does not receive rain during the winter, the ATLAS grid's rainfall values are lower than the ANUSPLIN values by less than 5mm. Apart from the areas described in the previous paragraph, ATLAS produced lower rainfall values (more than 10mm) for the rest of the study area.

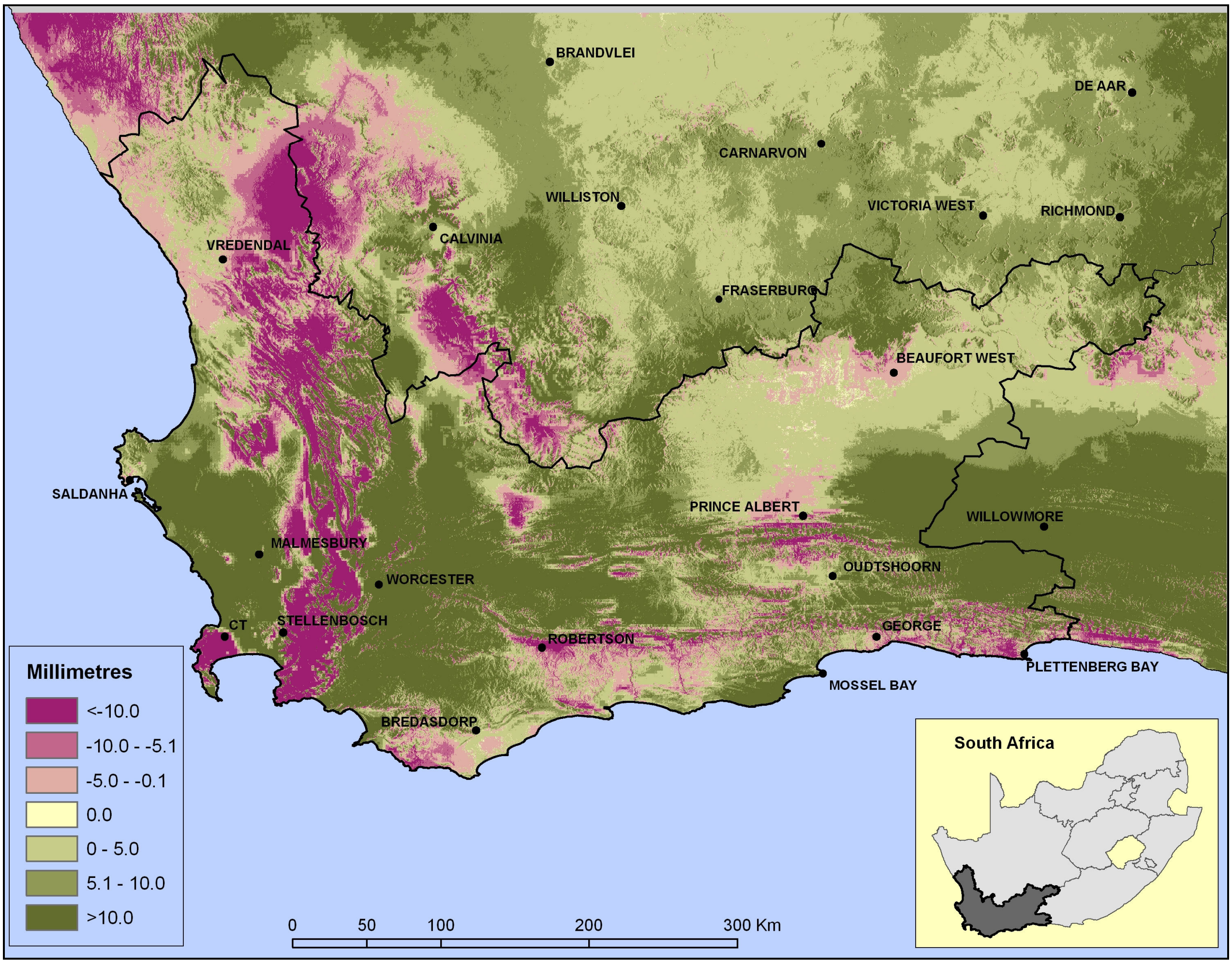


Figure 3.14: Difference in rainfall between ANUSPLIN and ATLAS grids during June

For all 12 months of the year large differences (>10mm) between the two rainfall grids are evident for significant parts of the study area. For the worst case scenario this indicates a difference of 15mm or more between the ATLAS grids and the dataset provided by the SAWS and the ARC. The variation in rainfall is higher during the cooler seasons.

3.7.4 Conclusion of climate grid comparison

In general, ATLAS calculated lower minimum temperatures (>1.0^{o}C) than ANUSPLIN, indicating that the ATLAS grids differ from the monthly mean minimum temperature dataset used for the ANUSPLIN interpolation process. This was also the case for the monthly mean maximum temperature grids. In the mountainous regions it appears that ATLAS interpolated higher temperatures compared to ANUSPLIN. This is probably attributed to the fact that the weather stations in these regions are all situated at lower altitudes and that different methods were used to calculate the temperature lapse rates. According the deviation grids, the maximum and minimum temperatures differ more during the 'warmer' months compared to the 'cooler' months.

The rainfall deviation grids for all the months of the year indicate that the ATLAS values are lower compared to the ANUSPLIN interpolated values. Opposed to this, ATLAS produced higher rainfall values for the mountain ranges running north-south in the study area. The areas where higher rainfall values were produced by the ATLAS calculations most often coincide with areas located on the rain shadow side of the mountains (northeastern facing). Although the ATLAS maps do take orographic lifting, aspect and terrain roughness into consideration, the effect of rain shadow cannot be clearly seen on the grids produced and are therefore of less importance during the ATLAS grid-creation process (Schulze 1997).

The third to last objectives set out during the project description have now been completed. The final maps of the monthly mean daily maximum and minimum temperature and the mean monthly rainfall for each month of the year can be found on the compact disk in appendix D. The next section will aim at providing a summary of the findings made in this chapter. It will also provide recommendations on how to improve the accuracy of the interpolation process and offer further recommendations for future projects which aims to produce and use interpolated climate datasets.

CHAPTER 4: CONCLUSION AND RECOMMENDATIONS

4.1 INTRODUCTION

The aim of this study was to create high-resolution climate grids for the Western Cape by means of a suitable interpolation technique. The climate variables selected where the monthly mean daily maximum and minimum temperatures and the mean monthly rainfall for the twelve months of the year.

The first and second objectives of the study were to investigate climate interpolation techniques and programs to assist in identifying a suitable climate interpolation program for the study region. ANUSPLIN, an interpolation software package which makes use of thin plate smoothing splines was selected. ANUSPLIN has mostly incorporated latitude, longitude and elevation as suitable variables, but undoubtedly additional factors also influences a region's climate. The third objective of the study was to identify such additional factors and explore the degree to which these variables influence the study area's climate.

Correlation analyses provided evidence that distance to nearest ocean and slope and aspect – the latter in combination using a hillshading method to provide a cell's illumination value, do influence the climate of the Western Cape. It was concluded that the two oceans bordering the study region have a moderating effect on the maximum and minimum temperatures experienced in the region. The farther away a station is located from the ocean, the higher its maximum temperature during summer months, while stations closer to the oceans have higher minimum temperatures during the winter. Correlation analysis also demonstrated that regions closer to the coast experienced higher rainfall in some parts.

It was established that the predominant winds come from a southern and northern direction for the study area, depending on the season. Therefore, two hillshade grids were created, one with an azimuth from a northwestern direction and one from a southern direction. In the one hillshade analysis it was found that there was no significant relationship between rainfall and topography (aspect and slope) when the illumination value is calculated using an azimuth from a northernwestern direction. In the other hillshade analysis it was found that there was a definitive, but small relationship between rainfall and topography (aspect and slope) when the illumination value is calculated using an azimuth from a southern direction. The analysis established that slopes facing a southern direction received higher rainfall compared to north-facing slopes. This result can be attributed to the small number of weather stations situated on the northwestern slopes of the western escarpment. This investigation of the influence of topography provided little evidence that aspect and slope influence an area's temperature.

Since the literature study provided relatively few examples of projects where these two additional variables were incorporated in the interpolation process, the next object of the study was to investigate the suitable transformations and units for these additional variables. The accuracy of the climate grids produced with these additional variables was examined in order to assess whether the selected additional variables enhance the accuracy of the interpolation process.

4.2 SUCCESSFUL INCORPORATION OF ADDITIONAL VARIABLES

Before the interpolation process could be started it was necessary to establish the correct parameterization for the standard dependent (temperature and rainfall) and independent (latitude, longitude, elevation) variables. The literature available on ANUSPLIN provides useful information on the suitable parameterization for these variables in order to obtain the best results during the interpolation process. The suggestions were investigated in order to derive the most apt parameterization for this project before the additional variables could be incorporated. The following conclusions were made regarding the parameterization of standard variables for this project:

- It was found that the final climate grids (for all three climate variables) over and under estimated the climate conditions compared to the weather station dataset when all the location variables were incorporated as independent variables. This confirms advice that all variables in addition to latitude and longitude should be incorporated as covariates (Hutchinson 1998b).

- The literature also suggests that elevation should be scaled to kilometres, but when this was done no significant changes to the results occurred.

- This study furthermore confirmed Hutchinson's (1998a) suggestions that the square root transformation of rainfall data will provide better results.

The next step (objective four) was to attempt adding the two additional variables to the interpolation process by means of identifying appropriate parameterization for each of the new variables. Distance to nearest ocean was incorporated in the monthly mean daily maximum temperature, monthly mean daily minimum temperature and monthly mean rainfall while topography (aspect and slope) was only incorporated in the monthly mean rainfall interpolation process.

In order to examinate the accuracy of this process, objective five aims at testing the precision of the interpolation process with and without the additional variables by means of using the output statistics produced by ANUSPLIN. Another popular method of testing the accuracy of the interpolation process is by means of withholding a selected number of stations from the dataset used during the interpolation process to produce residual values for the test stations. The results after the completion of objective four and five is summarized as follows:

- The incorporation of distance to ocean revealed to be most successful for the monthly mean daily maximum temperature when distance was scaled to kilometres and underwent logarithmic transformation. After the selected test stations were extracted from the original dataset, the average mean residual for the quadvariate model was less than 0.3^{o}C compared to the trivariate model (only latitude, longitude and elevation) which had an error of 0.4^{o}C. This additional variable also produced better unbiased estimates of the 'true' mean square error (MSE) values which are estimations of the accuracy of the spline function after the known error is removed.

- In the monthly mean daily minimum temperature analysis, the distance to ocean variable was more difficult to incorporate for the April to October period since the spline function's signal values were significantly lower than the acceptable value of half the number of data points, indicating that only a few stations' data could be used during the interpolation process. The winter months were identified as those where proximity to ocean does have a strong effect on a region's climate. This means that ANUSPLIN, using the models described in this project, is not extremely successful in incorporating this additional variable when the correlation between a region's temperature and distance to nearest ocean is moderate. It was found that the model which employs the logarithmic transformation of the distance to ocean parameter produced the lowest error with average mean residual of 0.4^{o}C. Compared to the trivariate model, the quadvariate model produced lower errors. Although less successful in contrast with the maximum temperature interpolation process, incorporating the effect of large water bodies provided better results when calculating both monthly mean daily maximum and minimum temperatures.

- The monthly mean rainfall interpolation analysis showed that the logarithmic transformation of the distance to ocean variable also provides better results compared to the other distance to ocean quadvariate models. In the investigation of the most suitable quadvariate model for topography (aspect and slope) variable, the best results were obtained when the illumination value was scaled by 1 000. Comparison of two quadvariate models to the trivariate model indicated that the additional hillshade variable provided lower RTMSE values (square root of

MSE). The error analysis produced an average residual of 4.5mm (or 11.7%) for the grids where hillshade was used. Compared to other rainfall interpolation projects done throughout the world (Hutchinson 1998b; Price et al. 2000) these results attest to an accurate interpolation of rainfall. The possibility of incorporating both distance to nearest ocean and topography into a quinvariate model was investigated. In this case the maximum value calculated by ANUSPLIN for the rainfall grids by far (>50%) exceeded the dataset's maximum value and therefore could not be incorporated successfully.

During the interpolation process for all three the climate variables it was noticed that the signal values of all the models which incorporated the additional variables were lower than those suggested by the literature. These suggestions are only for second-order trivariate (latitude, longitude and elevation variables) models and not for quadvariate/quinvariate models. From the results in this study it is apparent that when the order of the spline function is higher than two, lower signal values are obtained.

Using the best models identified for all three the climate variables the final climate grids were created (objective six).

For objective seven, the newly created grids were compared with existing grids. Comparisons between the ANUSPLIN output and the grids of the *South African atlas of agrohydrology and climatology* (referred to as the ATLAS dataset) provided a range of results with the two datasets differing over large areas. Since ANUSPLIN interpolates grids with values relatively close the dataset's values (within a range of 0.5^{o}C for temperature and 4.5mm (11.7%) for rainfall), the study reveals how much the ATLAS grids deviate from the data provided by SAWS and ARC.

- The comparison makes it clear that for all the months of the year both the maximum and minimum temperature grids differ more than 1.0^{o}C over a significant part of the study area, but depending on the season these regions varies.

- The rainfall deviation grids for all the months of the year indicate that the ATLAS values are often lower compared to the dataset values used for this project over large parts of the study area. Opposed to this, ATLAS produced higher rainfall values for the mountain ranges running north-south in the study area. The areas where higher rainfall values were produced by the ATLAS calculations most often coincide with areas located on the rain shadow side of the mountains (northeastern facing).

The results indicate a large deviation between the ANUSPLIN dataset and the ATLAS grids. The reason for this can be attributed to various factors ranging from the time period over which the climate data was collected to the effect of the difference in resolution of the underlying elevation grids. A more accurate comparison can be made if a temperature lapse rate adjustment is made when interpolating the ATLAS grids to the resolution of the ANUSPLIN grids.

4.3 RECOMMENDATIONS FOR FURTHER RESEARCH

The pursuit for the best interpolation of climate grids requires further investigation into the enhancement of the accuracy process. The following section will provide some possible means of accomplishing this and will also make suggestions on how climate interpolation can be modified to integrate different climate data other than traditional weather stations.

3.7.5 Increasing the accuracy of the interpolation process

During this project, the overall best performing model was selected to interpolate the climate variable for all the months of the year. For some months of the year, this model did not always produce the most accurate results. During a more specialized study, the model selection can be done monthly or seasonally.

Introducing distance to nearest ocean as an additional parameter was done successfully during this study, but the influence of the two different ocean currents were not taken into consideration. For an area such as the southwestern Cape where both the cold Benguela and warm Agulhas currents border the region, further studies are required to establish which current has the greatest influence. From this the distance of the more significant ocean current can be incorporated instead of the nearest ocean. During this study other large water bodies, such as dams or lakes, were not considered. Investigating the influence of large dams can be done when sufficient weather station data is available for regions surrounding such water bodies. According to Engelbrecht (2007, pers com.) the influence of dams will be very localized and insignificant.

The literature on climate interpolation underlines the fact that rainfall interpolation is consistently more difficult than temperature interpolation. This study supports this contention. This study also confirmed that topography (aspect and slope) plays an important role in a region’s climate and it established that the hillshading method, which combines aspect and slope into one variable, can be incorporated in the interpolation process. Although the final results indicated that the error of the interpolation process was lower than interpolation projects performed elsewhere, further investigation of other methods incorporating topography is necessary.

In this project the angle of the "light source" was set to 45^o when calculating the hillshade grid, but it has not been verified as to be the most optimum angle, therefore calling for further investigation. Correlation analysis of the dataset provided by SAWS and ARC revealed that only one hillshade (with a southern component) could be used. The reason is that there is insufficient data to provide evidence that correlation between aspect and slope and the climate variables exists for other azimuths. Further study is needed using data which represents a wider range of the different angle and slope components in order to seek relationships between a station's topography and the climate conditions experienced at that station. Pathetic

The data provided by a few stations located on northern slopes of mountains indicate that some of the north-facing slopes in the southwestern part of the study region receive more rainfall than the southern slopes. This suggests that this sub-region's rain shadow is not on the same side as the rest of the study area. While ANUSPLIN makes use of the splining function which is a deterministic with a local stochastically interpolator (see section 2.1 for definitions), one climate region may have an influence on another during the interpolation process, even where the difference in climate experienced by the two regions is fairly severe. In the analyses seasonal variability was also emphasized while the predominant wind directions change regionally. By implementing hillshade with different azimuths, compensation can be made for the changes in wind directions accompanying the change in seasons. In order to incorporate local phenomena, better results could be obtained when the study region is divided into sections described by Schulze & Maharaj (2004) and to interpolate these regions separately. Once the grids have been created they can be combined to cover the entire study region.

A profitable next step will be to compare the results of the hillshading method to those of the method described in Hutchinson (1998b) where slope and aspect are combined to form two additional variables. Other ways of incorporating slope and aspect should also be sought and compared with these two methods in order to establish the optimum way of incorporating topography into the interpolation process.

The comparisons of the final product created by ANUSPLIN with the maps created by the *South African atlas of agrohydrology and climatology* alerts one to the fact that there are large differences between the two datasets. Since the ANUPSLIN grids are true to weather station data provided by SAWS and ARC it is the ATLAS grids that deviate from SAWS and ARC data. For a more accurate method of comparing the ANUSPLIN data and the ATLAS data with each other (and not with the dataset provided by SAWS and ARC), it will be necessary to identify weather stations with long-term data that were not used in either of the interpolation projects. The residuals of these stations will give a better appraisal of the accuracy of the two interpolation methods employed by ANUSPLIN and ATLAS. The effect of using underlying elevation grids with different spatial resolution will also have to be taken into consideration and an adjustment in temperature lapse will have to be made.

3.7.6 Taking climate interpolation further

To assess the performance of ANUSPLIN as an interpolator, another possibility is to use climate data from climate modelling as opposed to weather stations and to compare the outcome with grids created using conventional weather station information. Furthermore, to establish the accuracy of ANUSPLIN, natural vegetation can be used as an indicator of climate and climate zones. Various factors such as soil types and natural vegetation can be incorporated into a multi-criteria evaluation to estimate climate zones in a region like the Western Cape with its diverse flora.

Until now ANUSPLIN has been used to interpolate climate data using the averages of weather conditions recorded at weather stations over a long period. A next step will be to investigate the possibility of using ANUSPLIN to interpolate weather conditions recorded on a daily basis in the study region. This data can then be compared with climate modelling programmes such as RAMS, a software package which simulates a region's atmospheric conditions on a specific time-scale.

4.4 CONCLUSION

This study has aimed to create high-resolution climate grids for the Western Cape by means of a suitable interpolation technique. Using ANUSPLIN, this aim has been achieved. The additional variables identified for the interpolation process were successfully incorporated once the optimum units and transformations of the variables had been identified. This demonstrates that it is possible to enhance the accuracy of the interpolation process by using distance to the nearest ocean and aspect and slope (in the form of hillshading) as independent variables in addition to latitude, longitude and elevation.

Although ANUSPLIN provided climate grids with acceptably low errors, the possibility of enhancing the accuracy of this climate interpolator requires further investigation. Currently the trend is to move away from statistical packages and to use human-expert software programs, but until these programs are commercially available the accuracy of this method cannot be compared to that of ANUSPLIN. An important advantage of statistical packages is that they are objective, therefore can be used universally. For the present, ANUSPLIN is easy to use, requires little computational speed and memory and provides accurate climate grids which can be used in research projects focusing on the same the study area used in this project.

REFERENCES

Anderson S 2004. An evaluation of spatial interpolation methods on air temperature in Phoenix AZ. [Online] Department of Geography, Arizona State University. Available from: http://www.cobblestoneconcepts.com/ucgis2summer/anderson/anderson.htm [Accessed 1 December 2005].

Antonić O, Križan J, Marki A & Bukovec D 2001. Spatio-temporal interpolation of climatic variables over large a region of complex terrain using neural networks. *Ecological Modeling* 138: 225 – 264.

ArcGIS 1999. ArcMap 9.1. ESRI ®. Online available: http://www.esri.com

Barringer JRF & Lilburne LF 2000. Developing fundamental data layers to support environmental modeling in New Zealand: progress and problems. Canada [Online] 4th International Conference on Integrating GIS and Environmental Modeling. Available from: http://www.colorado.edu/research/cires/banff/pubpapers/221/ [Accessed 22 December 2005].

Bonnardot V, Planchon O & Cautenet S 2005. Sea breeze development under an offshore synoptic wind in the South-Western Cape and implications for the Stellenbosch wine-producing area. *Theoretical and Applied Climatology* 81, 3: 203 -218.

Bootsman A, Anderson D & Cameda S 2004. *Potential impacts of climate change on agroclimatic indices in southern regions of Ontario and Quebec* (Technical Bulletin no. 03 – 284). Ottawa: Eastern Cereal and Oilseed Research Centre.

Borga M & Vizzaccarco A 1997. On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology* 195: 160 – 171.

Bryan BA & Adams JM 2002. Three-dimensional neurointerpolation of a annual mean precipitation and temperature surfaces for China. *Geographical Analysis* 34, 2: 93 – 111.

Centre for Resource and Environmental Studies 2004. CRES software products [Online]. Available from http://cres.anu.edu.au/outputs/software.php [Accessed on 15 June 2006].

Centro Internacional de Agricultura Tropical 2001. Solutions that cross frontiers [Online]. Colombia: CIAT. Available from: http://www.ciat.cgiar.org/ [Accessed 15 January 2006].

Chapman AD 2003. The Case for a 3 minute climate surface for South America (Internal report No. 3). Toowoomba South: Australian Biodiversity Information Services.

Chapman AD, Muńoz ES & Koch I 2005. Environmental information: Placing biodiversity phenomena in an ecological and environmental context. *Biodiversity Informatics* 2: 24-41.

Chang K 2006. *Introduction to geographic information systems*. 3rd ed. Boston: McGrawHill

Climate Source 2000. Climate mapping with PRISM [Online] Oregon: The Climate Source Inc. Available from: http:// www ocs.orst.edu/pub/prism/docs/prisguid.pdf [Accessed 9 June 2004].

Climate Systems Analysis Group 2006. CSAG publications [Online] Cape Town: University of Cape Town. Available from: http://www.csag.uct.ac.za/research.html [Accessed 31 January 2006].

Collins FC & Bolstad PV 1996. *A comparison of spatial interpolation techniques in temperature estimation*. In proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe: New Mexico.

Cramer W, Doherty R, Hulme M & Viner D (eds) 2000. *A concerted action towards the improved understanding and application of results from climate model experiments in European climate change impacts research: climate scenarios for agricultural, forest and ecosystem impacts.* (Report no. 2) ECLAT-2 Workshop, Postad: Germany.

Crane RG & Hewitson 2003. Clustering and upscaling of station precipitation records to regional patterns using self-organizing maps (SOMs). *Climate Research* 25: 95 – 107.

Daly C, Gibson WP, Taylor GH, Johnson GL & Pasteris P 2002. A knowledge-based approach to statistical mapping of climate. *Climate Research* 22: 99 – 113.

Daymet U.S. Data Centre 2005. Daily surface weather climatalogical summaries [Online] Available from: http://www.daymet.org/default.jsp [Accessed 16 January 2006].

Department of Environmental Affairs and Tourism (DEAT) 2005. Shape of the provinces [Online]. South Africa. Available from http://www.environment.gov.za/WesternCape/shape of the provinces.htm [Accessed 13 July 2005].

Dirks DN, Hay JE, Stow CD & Harris D 1998. High-resolution studies of rainfall on Norfolk Island Part II. Interpolation of rainfall data. *Journal of Hydrology* 208: 187 – 193.

Dubois G 1998. Spatial interpolation comparison 97: foreword and introduction. *Journal of Geographic Information and Decision Analysis* 2, 2: 1 – 10.

Engelbrecht FA 2005. Simulations of climate and climate change over southern and tropical Africa with the conformal-cubic atmospheric model. In *Climate Change and Water Resources in Southern Africa: Studies on Scenarios, Impacts, Vulnerabilities and Adaptation*, ed. R.E. Schulze, chap. 4, pp. 57-74. WRC Report 1430/1/05. Water Research Commission: Pretoria.

Engelbrecht FA, Rautenbach CJ de W, McGregor JL & Katzfey JJ 2002. January and July climate simulations over the SADC region using the limited-area model DARLAM. Water SA 28: 361 – 374.

Funk VA & Richardson KS 2002. Systematic Data in Biodiversity: Use it or lose it. System Botany [Online] 51, 2: 303-316. Available from: http://lsvl.la.asu.edu/bio470/jfouquette /FunkRichardsn.SB51(2).pdf [Accessed 26 Nov. 2003]

Hartkamp AD, De Beurs K, Stein A & White JW 1999. *Interpolation techniques for climate variables*. (Report No. 99-01). Mexico City: International Maize and Wheat Improvement Centre.

Hewitson B 2003. Developing perturbations for climate change impact assessments. *EOS* 84, 35: 337 – 341.

Hewitson BC & Crane RG 1996. Climate downscaling: techniques and application. *Climate Research* 7, 2 : 85-95.

Hijmans RJ, Cameron SE, Parra JL, Jones PG & Jarvis A 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965 – 1978.

Hudson DA & Jones RG 2002. *Simulations of present-day and future climate over Southern Africa using HadAM3H* (Technical note 38). Bracknell: Hadley Centre for Climate Prediction and Research, Met Office.

Hutchinson MF 1998a. Interpolation of rainfall data with thin plate smoothing splines: Part I. Two-dimensional smoothing of data with short range correlation. *Journal of Geographic Information and Decision Analysis* 2, 2: 152-167.

Hutchinson MF 1998b. Interpolation of rainfall data with thin plate smoothing splines – Part 2: Analysis of topographic dependence. *Journal of Geographic Information and Decision Analysis* 2, 2: 152 – 167.

Hutchinson MF 2004. ANUSPLIN Version 4.3. Canberra. Centre of Resource and Environmental Studies: Australian National University.

Hutchinson MF, Nix HA, McMahon JP & Ord KD 1996. The developing of a topographic and data climate database for Africa. Third Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Barbara.

Intergovernmental Panel on Climate Change 2001. Appendix I – Glossary [Online]. Available from: http://www.grida.no/climate/ipcc_tar/wg1/518.htm [Accessed 16 December 2005]

Jakob D, Taylor BF & Xuereb KC 2005. *A pilot study to explore methods for deriving design rainfalls for Australia - Part 1* (Report No.10). Melbourne: Hydrometeorological Advisory Service, Bureau of Meteorology.

Jeffrey SJ, Carter JO, Moodie KB & Beswick AR 2001. Using spatial interpolation to construct a comprehensive archive of Australia climate data. *Environmental Modeling & Software* 16: 309 – 330.

Johnson GL, Daly C, Taylor GH & Hanson CL 1999. Spatial variability and interpolation of stochastic weather simulation model parameters. *Journal of Applied Meteorology* 39: 778 – 796.

Joubert L 2000. Hot wine in a changing world [Online]. Winelands. Available from: http://www.wineland.co.za/0502hot.php3 [Accessed 10 March 2005]

Joubert SJ 2004. Enhancement of climate data in the Western Cape region. Literature Review. Department of Geography and Environmental Studies: University of Stellenbosch.

Kruger AC 2002. *Climate of South Africa. Surface winds*. (Report No. WS43). Pretoria: South African Weather Service.

Kurtzman D & Kadmon R 1999. Mapping of temperature variables in Israel: a comparison of different interpolation methods. *Climate Research* 13: 33 – 43.

Linacre E 1992. *Climate data and resources: a reference and guide*. London: Routledge.

Lutgens FK & Tarbuck EJ 1998. *The atmosphere*. 7th ed. New Jersey: Prentice Hall.

Lynch SD 1999. Converting point estimates of daily rainfall onto a rectangular grid. ESRI User Conference '98, San Diego.

Lynch SD 2003. *Development of a raster database of annual, monthly and daily rainfall for Southern Africa* (Report No. 1156/1/03). Pretoria: Water Research Commission.

Maneveldt GW 2005. UWC's enviro-facts guide to Fynbos [Online]. University of Western Cape: Cape Town. Available from: http://www.botany.uwc.ac.za/envfacts/fynbos/ [Accessed on 5 February 2006].

McGuffie K & Henderson-Sellers A 2005. A climate modelling Primer. 3rd ed. Toronto: John Wiley & Sons.

McKenney D 2000. Development of gridded climate data for Canada and North America using thin plate splines. Workshop: Methods and Techniques for Constructing Regional Climate Scenarios. Sault Ste. Marie: Canadian Forest Service.

Mouton J 2001. *How to succeed in your Master's and Doctoral studies: A South African guide and resource book* . Pretoria: Van Schaik.

Murray AE 1998. Development of a rainfall surface for the Herbert River catchment [Online]. Integrated Resource Use & Management Group, CSIRO Tropical Agriculture: Townsville. Available from: http://www.aims.gov.au/pages/research/herbert/herbert-rainfall-01.html [Accessed 1 June 2005]

Naoum S & Tsanis IK 2004. Rankining spatial interpolation techniques using a GIS-based DSS. *Journal of Global NEST* 6, 1: 1-20.

Natural Resources Conservation Service 2004. United States Climate Mapping [Online] United States Department of Agriculture. Available from: http://www.ncgc.nrcs.usda.gov/products/datasets/ climate/index.html [Accessed on 5 June 2005].

New M 2002. Climate change and water resources in the southwestern Cape, South Africa. South *African Journal of Science* 98: 1 – 8.

NOAA Coastal Services Center 2004. Analyzing Benthic Data: Spatial Analysis [Online] NOAA. Available from: http://www.csc.noaa.gov/benthic/mapping/analyzing/spatial.htm [Accessed on 1 December 2005]

Olwoch JM, Rautenbach CD de W, Erasmus BFN, Engelbrecht FA & van Jaarsveld AS 2003. Simulating tick distributions over sub-Saharan Africa: the use of observed and simulated climate surfaces. *Journal of Biogeography* 30: 1221 – 1232.

Price DT, McKenney DW, Nalder IA, Hutchinson MF & Ketsteven JL 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agriculture and Forest meteorology* 191: 81 – 94.

Price DT, Mckenney DW, Papadopol P, Logan T & Hutchinson MF 2004. High resolution future scenario climate data for North America. Paper (delivered) at the 26th Conference on Agricultural and Forest Meteorology, Vancouver.

Rabus B, Einededer M, Roth A & Bamler R 2003. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. *Journal of Photogrammetry & Remote Sensing* 57: 241 – 262.

Rautenbach H 2005. Laboratory for Research in Atmospheric Modelling [Online] University of Pretoria: Pretoria. Available from: http://www.up.ac.za/academic/geog/meteo/events.htm [Accessed on 20 February 2006].

Schulze RE 1997. *South African atlas of agrohydrology and climatology* (Report No. TT82/96). Pretoria: Water Research Commission.

Schulze RE & Maharaj M 2004. *Development of a database of gridded daily temperatures for Southern Africa* (Report No. 1156/2/04). Pretoria: Water Research Commission.

Silva AP 2003. Neural networks applications to spatial interpolation of climate variables. Paper (delivered) at the 3rd Conference on Artificial Intelligence Applications to Environmental Sciences, Long Beach.

South African Weather Bureau 1996. *Weather and climate of the extreme south-western Cape*. Pretoria: Department of Environmental Affairs and Tourism.

Thornton PE, Running SW & White MA 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* 190: 214 251.

STATISTICA 2005. STATISTICA version 7.1. StaSoft Inc. Online available: http://www.statsoft.com/

United States Geological Survey (USGS) 2006. Shuttle Radar Topography Mission DTED [Online]. US Department of Interrior. Available from http://edc.usgs.gov/products/elevation/srtmdted.html [Accessed on 10 February 2006]

University of Montana 2005. Single-Point Bioclimatology: MT-CLIM (Mountain Climate Simulator) [Online] Numerical Terradynamic Simulation Group. Available from: http://www.ntsg.umt.edu/ [Accessed on 5 December 2005]

Van Zyl D 2003. *South African weather and atmospheric phenomena*. Pretoria: Briza.

Wilby RL, Dawson CW & Barrow EM 2002. SDSM — a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software* 17: 147 – 159.

Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J & Wilks DS 1998. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research* 34, 11: 2995–3008.

Wilson J P, H Mitasova and D S Wright (2000) Water resource applications of geographic information systems. *Journal of the Urban and Regional Information Systems Association* 12, 2: 61-79

Winter K 2002. *Oxford intermediêre atlas vir Suider-Afrika*. Cape Town: Oxford University Press.

Zuo H, Hutchinson MF, McMahon JP & Nix HA 1996. Developing a mean monthly climatic database for China and Southeast Asia. In Booth TH (ed.) Matching trees and sites (Report No. 63). ACIAR Proceedings.

PERSONAL COMMUNICATION

Cautenet S 2006: Professor in Meteorology University of Blaise Pascal. Clermont-Ferrand. Interview on 21 July. Regional Atmospheric Modeling System.

Daly C (daly@coas.oregonstate.edu) 2005. Re: Commercial availability of PRISM. E-mail to SJ Joubert (13083031@sun.ac.za) (15 August).

Engelbrecht FA 2007: Lecturer (Doctor) in Meteorology University of Pretoria. April 2007. Notes on thesis: SJ Joubert - High-Resolution climate variable Generalisation for the Western Cape.

Jones P (p.jones@cgiar.org) 2006. Re: Climate interpolation software packages. E-mail to S Joubert (13083031@sun.ac.za) (15 January).

APPENDICES

APPENDIX A

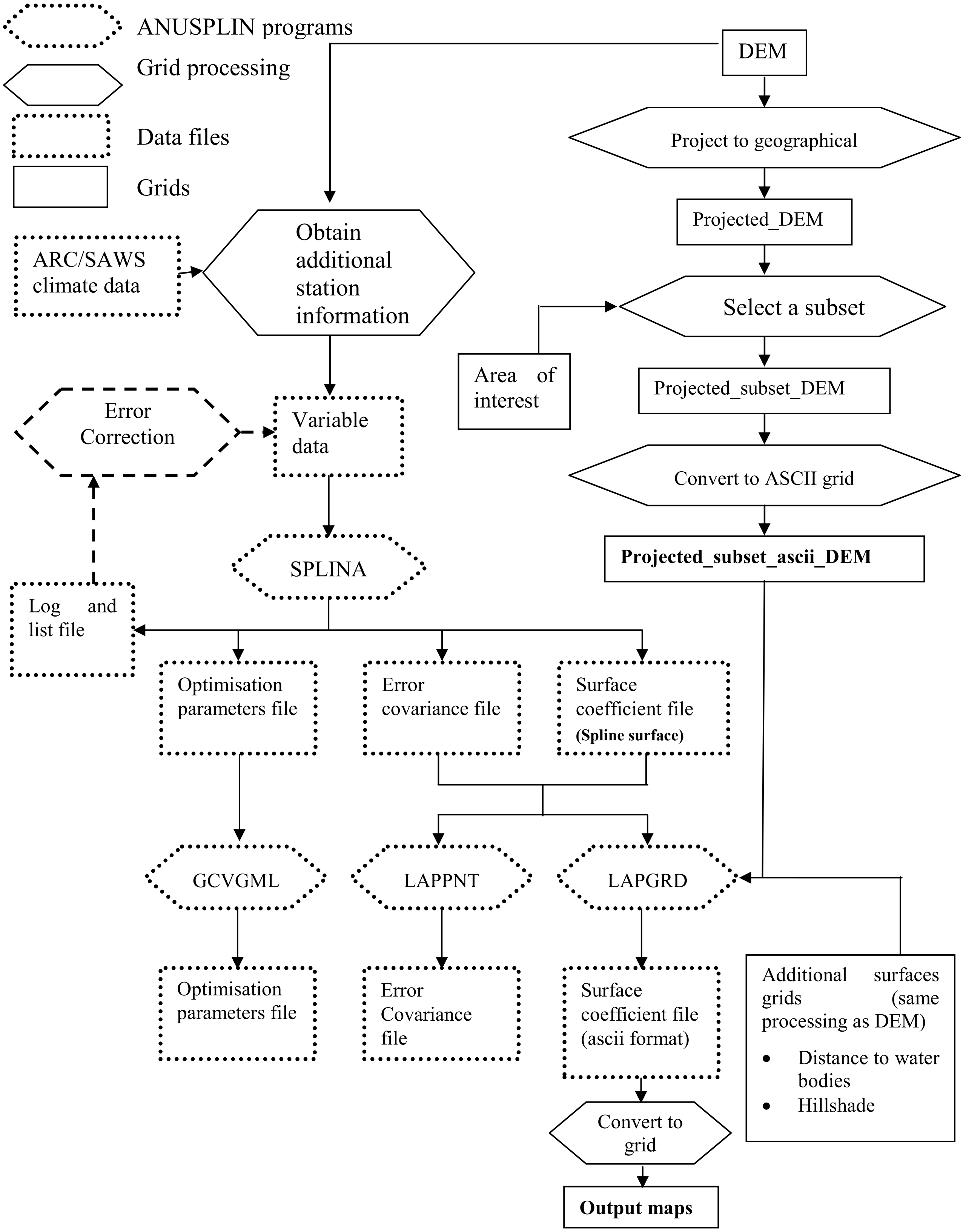
Brief summary of the programs which make up the ANUSPLIN package

PROGRAM	DESCRIPTION
SPLINA	A program which fits an arbitrary number of (partial) thin plate smoothing spline functions of one or more independent variables. Suitable for data sets with up to about 2000 points although data sets can have arbitrarily many points. The degree of data smoothing is normally determined by minimising the generalised cross validation (GCV) or the generalised maximum likelihood (GML) of the fitted surface.
SPLINB	An approximate version of SPLINA designed for larger data sets. It uses knots which are initially selected by SELNOT and updated by ADDNOT. Suitable for data sets with up to about 10,000 data points, with up to about 2000 knots, although data sets can have arbitrarily many points.
SELNOT	Selects an initial set of knots for use by SPLINB.
ADDNOT	Updates knot index file when additional knots are selected from the ranked residual list produced by SPLINB.
DELNOT	Adjusts knot index file when points are removed from the data file to be used by SPLINB.
GCVGML	Calculates the GCV or GML for each surface and the average GCV or GML for a range of values of the smoothing parameter for surfaces fitted by SPLINA. It can be applied to surfaces fitted by SPLINA or SPLINB. The values are written to a file for inspection and for plotting.
LAPPNT	Calculates values, and Bayesian standard error estimates, of partial thin plate smoothing spline surfaces at points supplied in a file.
LAPGRD	Calculates values, and Bayesian standard error estimates, of partial thin plate smoothing spline surfaces on a regular rectangular grid.

Source: Hutchinson 2004: 3

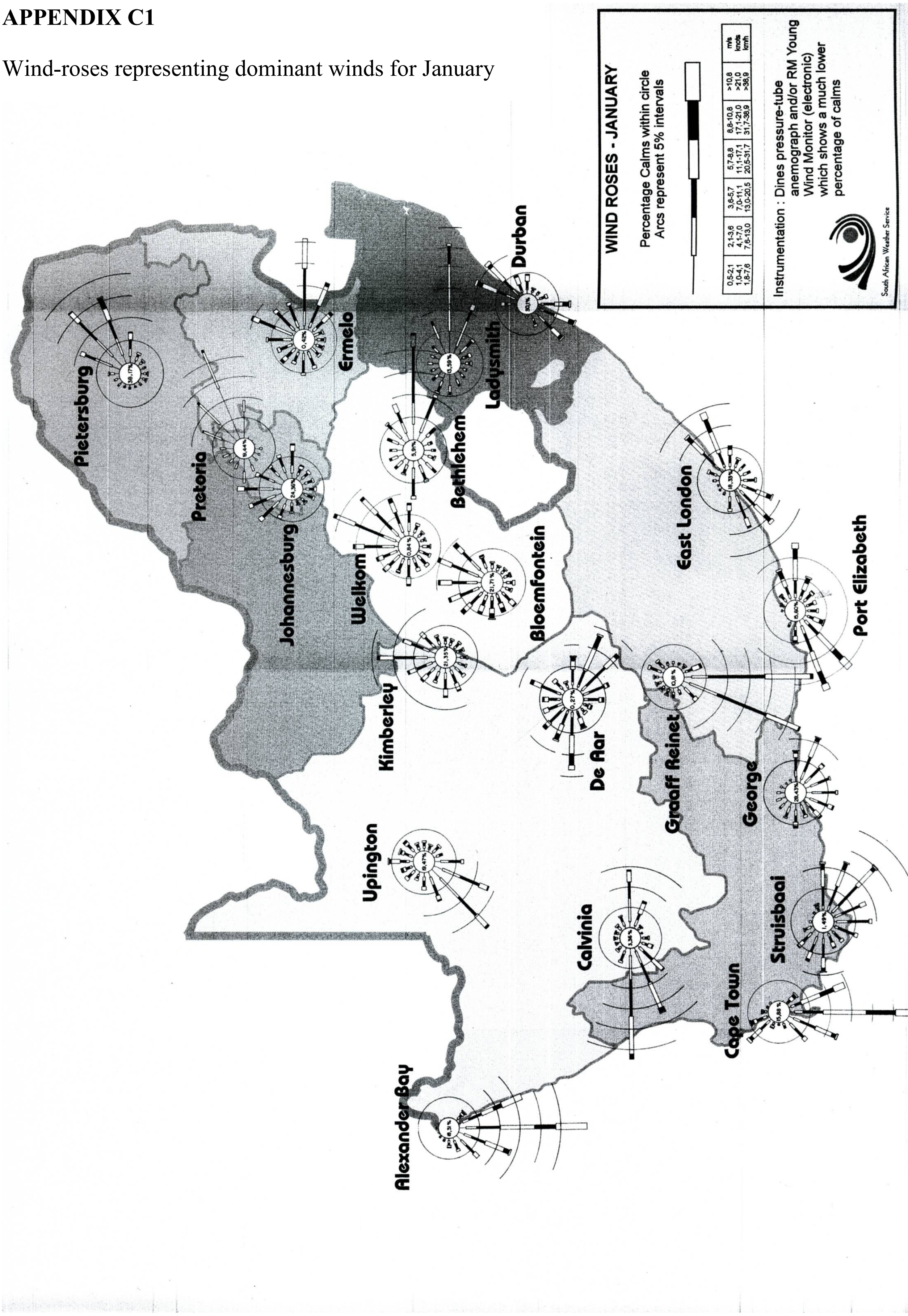
APPENDIX B

Data flow diagram for creating climate grids



APPENDIX C1

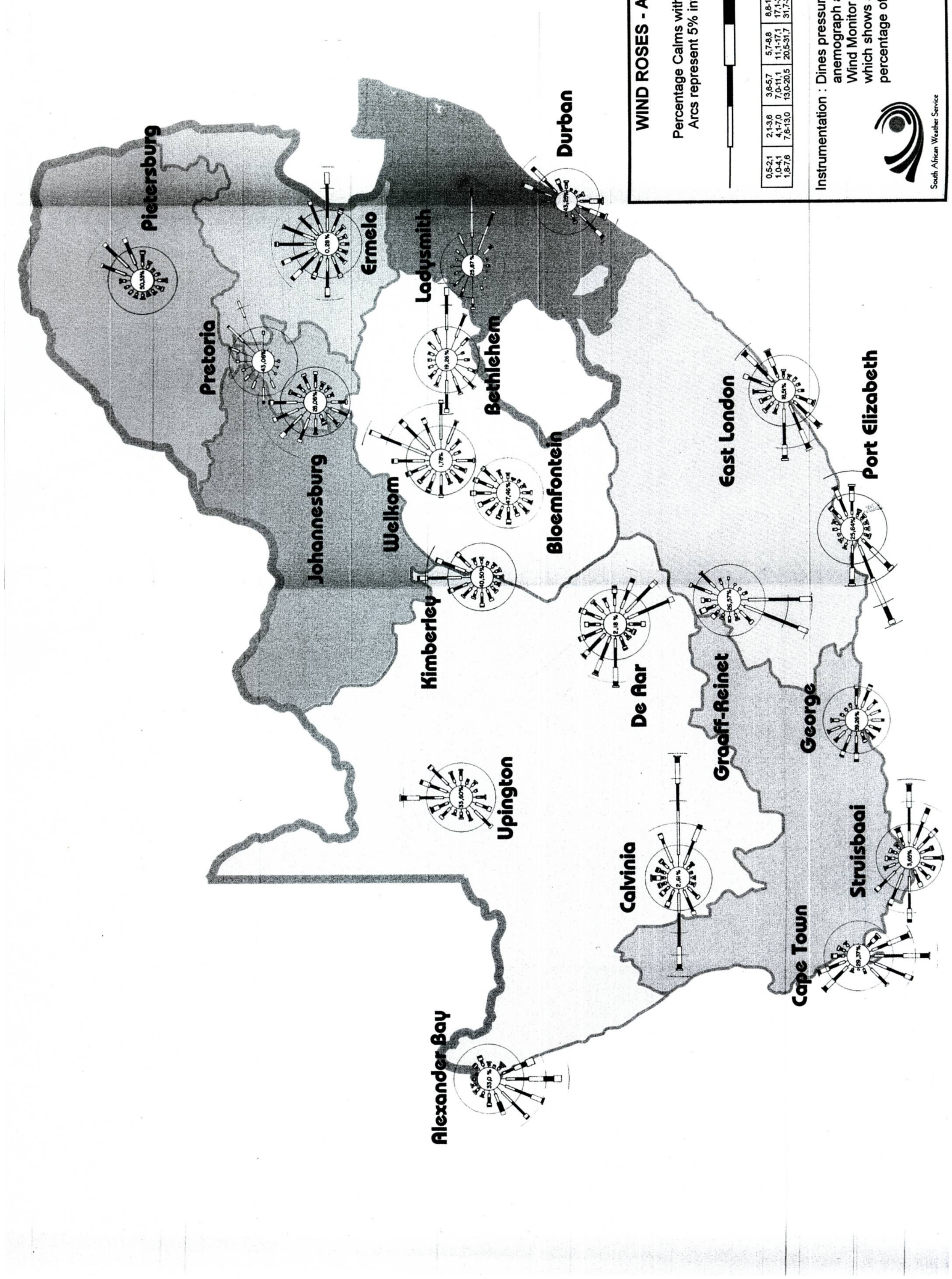
Wind-roses representing dominant winds for January



Source: Kruger 2004: no page

APPENDIX C2

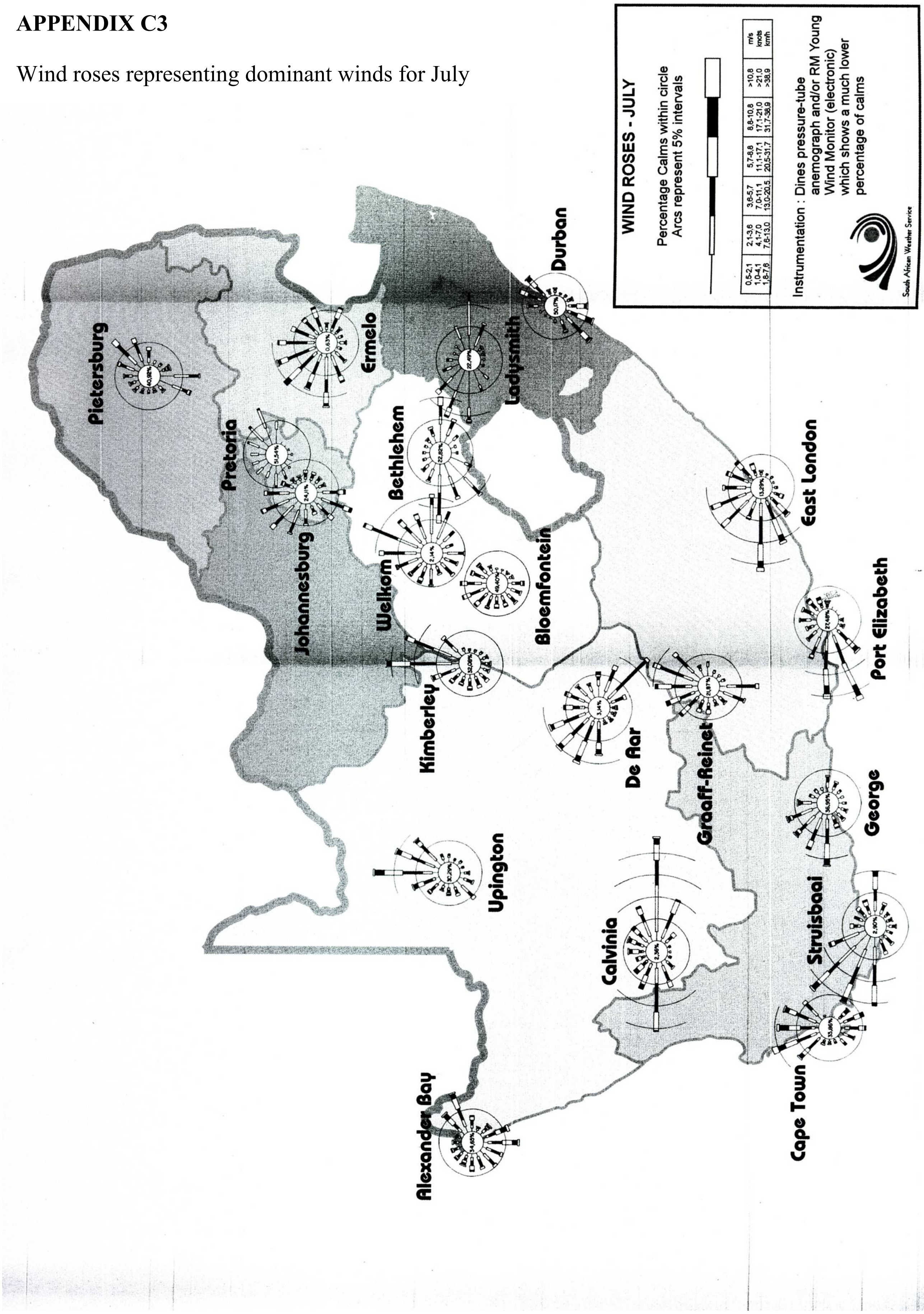
Wind roses representing dominant winds for April



Source: Kruger 2004: no page

APPENDIX C3

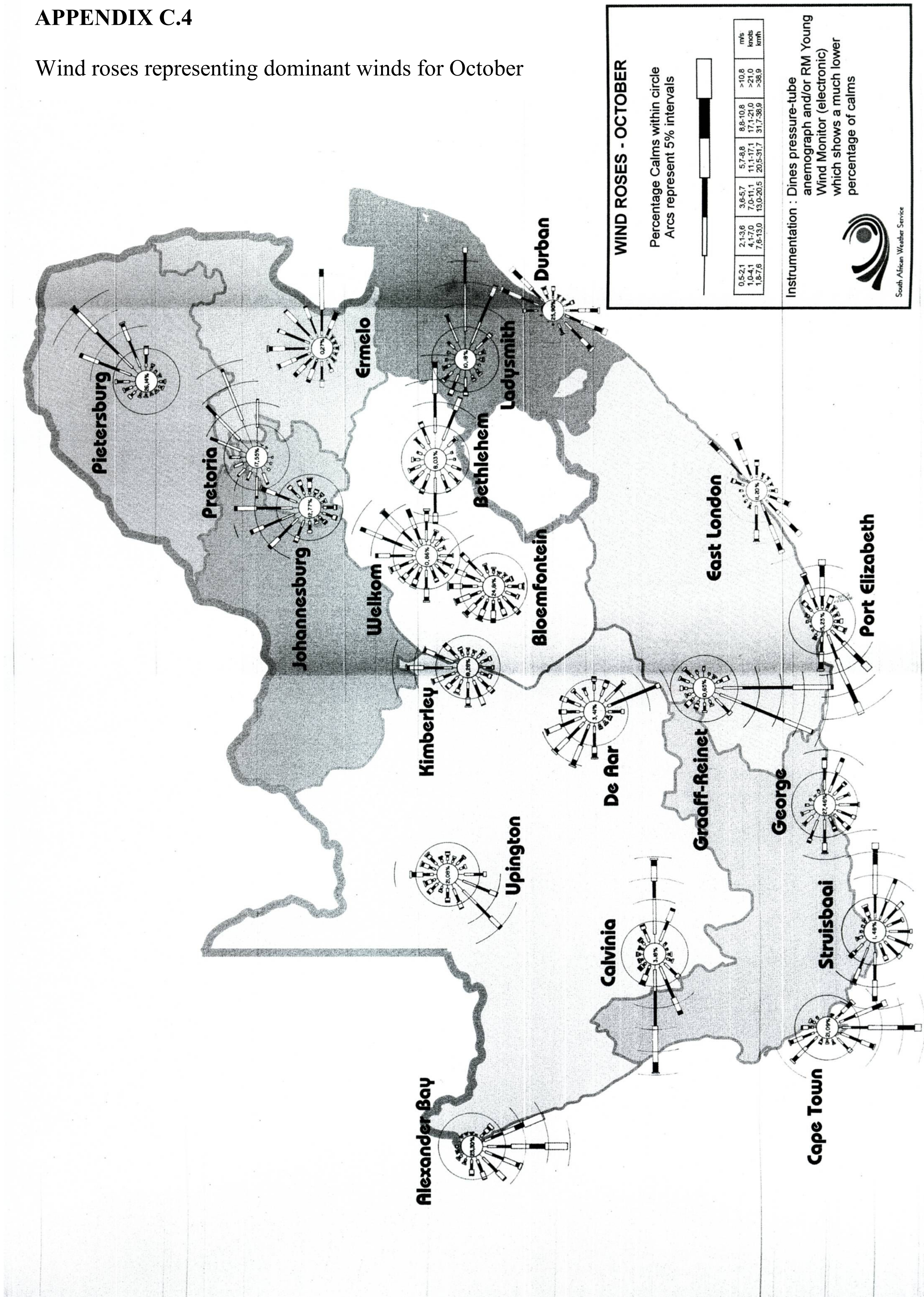
Wind roses representing dominant winds for July



Source: Kruger 2004: no page

APPENDIX C.4

Wind roses representing dominant winds for October



Source: Kruger 2004: no page

APPENDIX D

Monthly mean daily maximum and minimum temperature and mean monthly rainfall grids created by ANUSPLIN for each month of the year