DEVELOPMENT OF A RASTER DATABASE OF ANNUAL, MONTHLY AND DAILY RAINFALL FOR SOUTHERN AFRICA

Report to the
Water Research Commission

by

SD Lynch

School of Bioresources Engineering and Environmental Hydrology
University of KwaZulu-Natal, Pietermaritzburg, South Africa

WRC Report No. 1156/1/04 DECEMBER 2004

Set Number: 1-77005-249-6
ISBN Number: 1-77005-250-X

Disclaimer

This report emanates from a project financed by the Water Research Commission (WRC) and is approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the WRC or the members of the project steering committee, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

EXECUTIVE SUMMARY

INTRODUCTION

Rainfall measurements were recorded as early as 400 BC and, since its inception, both the principles and the purpose of rainfall measurement have remained unchanged (Biswas, 1967; Ward, 1975). The rainfall of an area helps to structure society in a geographical sense. Water is an essential element for life, thus the more water that exists in an environment, the more potential that environment has for sustenance of life. Early settlers settled in higher rainfall areas, and towns were built up in and around these areas, shaping settlement patterns as we experience them today (Antevs, 1938).

In order to obtain rainfall data the rainfall needs to be measured. Rainfall measurements can be undertaken by numerous different methods. The most common method is the use of a standard non-recording rain gauge, but nowadays estimation by radar and satellite is practised as well. While radar and satellite imaging for rainfall estimates are able to give real time, areal estimates of rainfall data, the primary source of rainfall data is still provided by the daily non-recording rain gauge (Seed and Austin, 1990). This is mainly so because rain gauges are cheap and generally reliable. Rain gauge data are also available for longer time periods, which is advantageous in many respects. It is anticipated that over time, the rain gauge network density will increase, so as to provide better areal estimates of rainfall from the use of point measurements.

The measurement of rainfall is a simple procedure provided that accuracy is not essential, as an exact measurement of rainfall is impossible to obtain owing to the random and systematic errors occur in measuring rainfall (Schultz, 1985). As no 'true rainfall amount' can be achieved, one can only attempt to improve the estimation of rainfall amounts by minimising the known errors, which are the systematic errors that are associated with a rain gauge used to measure rainfall amounts. Empirical equations have been derived which can be used to account for the systematic errors in point rainfall measurements (Schultz, 1985). Added to these errors, rainfall amounts are extrapolated to give an areal average of rainfall. Boughton (1981) stated that deficiencies of 10-20% could be expected in point measurements of rainfall and that a further 10-20% error is likely in extrapolating data from a point measurement to an areal average. To aid in decreasing this error, a sufficiently dense and suitably spaced rain gauge network needs to be used (Schultz, 1985).

Mean annual precipitation (MAP) is a widely used variable in hydrological design, water resources planning and agrohydrology. The basic sources for MAP in South Africa up to the end of the 1980s were the 1:250 000 average rainfall map series compiled and drawn in the early 1960s by the erstwhile Hydrological Research Division of what was then called the Department of Water Affairs. After that, with rainfall records having lengthened by more than 20 years and techniques of analysis and computerised mapping having increased dramatically, the next series of maps using data up to the mid-1980s, and the first set of digital raster rainfall information, became available towards the end of the 1980s (Dent, Lynch and Schulze, 1989). Since then a further 15 years of data have been collected.

OBJECTIVES

Project K5/1156 is titled "Research on the Development of a Revised Spatial Database of Annual, Monthly and Daily Rainfall and Other Hydroclimatic Variables for South Africa". This report covers the rainfall component of the project; a separate report (Schulze, 2003) deals with the other hydroclimatic variables. In addition to the development of revised spatial and temporal databases of annual, monthly and daily rainfall values and related rainfall statistics, this research was also to provide for:

- quality control and provision of standardised daily rainfall datasets for input into hydrological simulation models, using GIS and other techniques, where appropriate,
- development and application of new regionalised infilling/data extension techniques for rainfall, and for
- support of the *ACRU* modelling system and other WRC projects.

RAINFALL DATASETS

There are many organisations and private individuals that record daily and monthly rainfall data in South Africa and in its neighbouring countries. The majority of the rainfall data are recorded at a daily time-step, but there are a number of sites that only record rainfall figures at a monthly time-step. The rainfall data for the Kingdoms of Lesotho and Swaziland have historically been included in the South African dataset and the *status quo* will remain. The current rainfall database, assimilated in this project, consists of data from South Africa, Lesotho, Swaziland, Namibia, Botswana, Zimbabwe and Moçambique.

POINT AND RASTER RAINFALL DATA

The rainfall data described in this research fall into three categories, *viz*.

- point temporal rainfall data that are recorded at a site,
- infilled point temporal rainfall values that are estimated at a site, and
- spatial rainfall information that is stored in a raster.

The point rainfall data are normally recorded at a daily time-step and sometimes at a monthly time-step. These point rainfall values are converted onto a rectangular grid, or raster, using various regression and interpolation techniques that are discussed later in more detail.

RAINFALL DATA FOR SOUTH AFRICA

The initial daily and monthly rainfall datasets used in this study were acquired from the erstwhile Computing Centre for Water Research (CCWR) early in 2000. The datasets had initially been developed for a WRC funded project titled "*Mapping the Mean Annual Precipitation and Other Rainfall Statistics over Southern Africa*" (Dent *et al*., 1989) and had been updated annually until the start of this current WRC project.

The daily rainfall database developed in this project consists of data from a wide variety of organisations and individuals (Table ES1) that include, *inter alia*:

- the South African Weather Service (SAWS),
- the Agricultural Research Council (ARC),
- the South African Sugar Association (SASA), and
- a large number of municipalities, private companies and individuals (PVT).

All the recorders of rainfall data are hereby acknowledged gratefully for their co-operation and diligence, without which this project could not have succeeded.

Table ES1 Number of rainfall stations per organisation for South African rainfall data

Organisation	No. of stations
SAWS	8 281
ARC	2 661
SASA	161
PVT	1 050
Total	12 153

The Institute for Commercial Forestry Research (ICFR) supplied this project with a database of 445 stations (Figure ES1) at which accumulated monthly rainfall had been recorded. The ICFR is acknowledged gratefully for supplying these records.

RAINFALL DATA FOR THE NEIGHBOURING COUNTRIES

The datasets for Namibia, Botswana, Zimbabwe and Moçambique (Table ES2) that are in this current rainfall database are by no means complete. Some of the countries supplied daily data and these data are only to be used in this project and may not be distributed. The monthly totals, however, that are derived from these daily data may well be distributed freely since the daily values cannot be reproduced from the monthly totals.

Monthly rainfall data were obtained for 32 stations in Moçambique. It is not necessary to elaborate on the amount and time spent in obtaining these data, suffice it to say that a University of Natal Hydrology Honours student from Moçambique, Helio Banze, was of great assistance. Dr Jerry Ndamba supplied the project with daily rainfall data for 66 rainfall stations scattered across Zimbabwe and Dr Louis du Pisani made available rainfall data for Namibia.

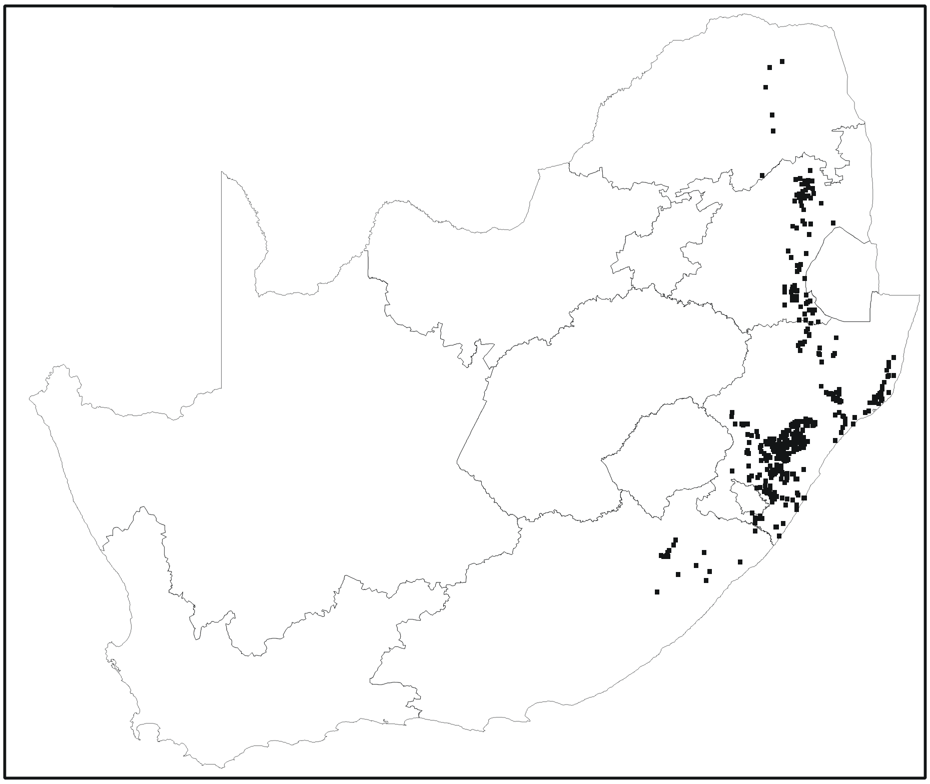


Figure ES1 Locations of the 445 monthly stations supplied by the ICFR

Table ES2 Number of rainfall stations from the neighbouring countries

Country	No. of stations
Namibia	640
Botswana	306
Zimbabwe	69
Moçambique	83
Total	1 098

There are many tales of unanswered requests for rainfall data, but the people who responded are acknowledged gratefully. Without their support this project would have been less successful. The rainfall database therefore consists of daily and monthly values for more than 13 000 stations (Figure ES2). These data were collected up to 21 November 2002.

QUALITY CONTROL

A rainfall database that has in excess of 300 million values is bound to have some errors.

Fotheringham *et al*. (2000) suggest that one needs to find out how *useful* the data are and not if they are *completely free of errors*.

Some of the errors that are known to exist in the data only become apparent when the rainfall data are used in hydrological modelling exercises and the model results are tested against measured streamflow values. The most common of these errors is the incorrect recording of the time and date at which the gauge is read, which in South Africa should be at eight o'clock each morning for the previous day's rainfall. These phasing errors, however, have hardly any effect on the monthly and annual totals.

A rainfall amount of 597 mm was recorded at St Lucia Lake for the day of 31 January 1984 when the cyclone Domoina struck the northeastern part of South Africa. It is the largest daily total of rainfall recorded in South Africa to date. One of the first quality control procedures that was performed on the daily rainfall data was to flag all rainfall amounts greater than 597 mm as suspect data.

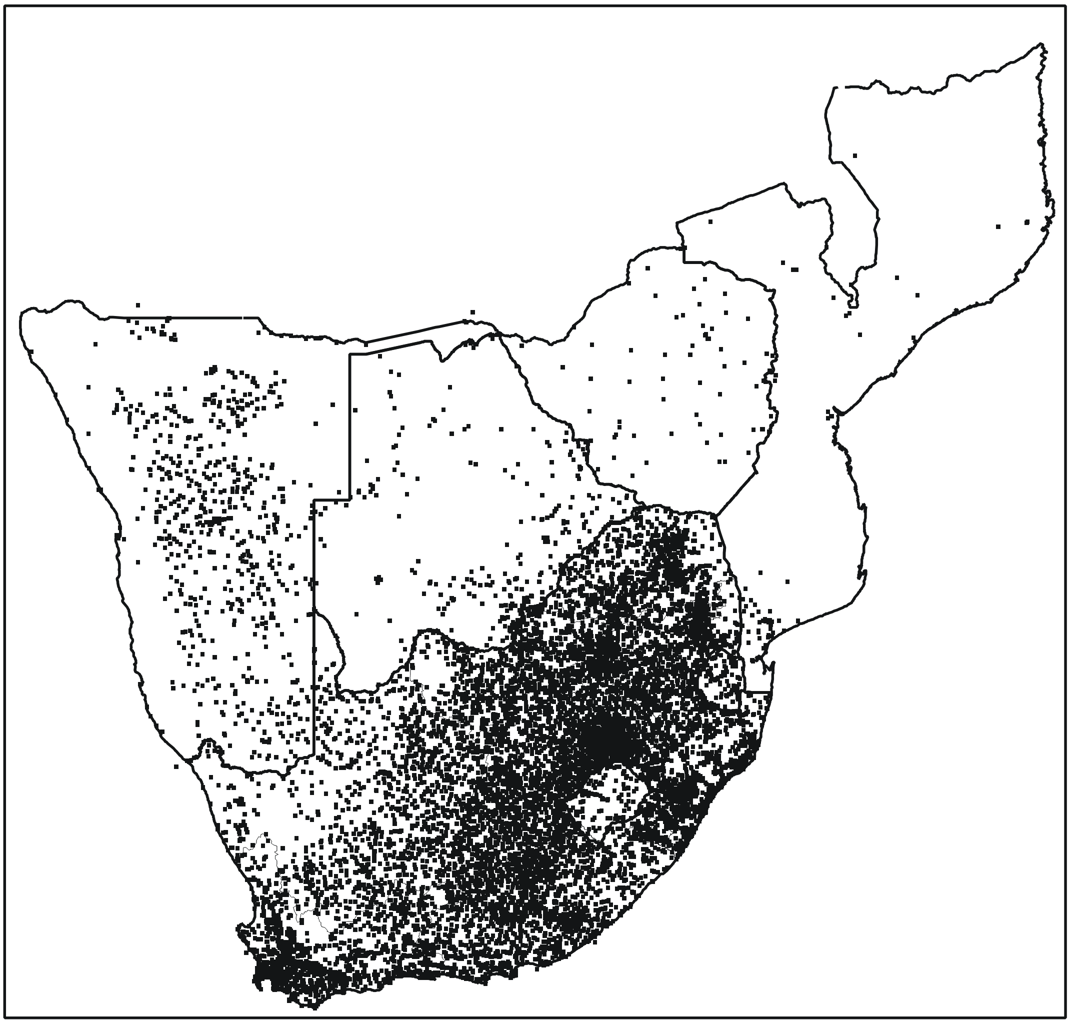


Figure ES2 Location of all the rainfall stations available to this project

Approximately 80% of the daily rainfall database consists of zero rainfall amounts while amounts less than 5 mm per day account for 10% of the database (Table ES3). It appears from the results in Table ES3 that infilling procedures are generating more 0-5 mm amounts, but retaining the less than 10 mm cumulative total of approximately 90% of the database.

One of the aims of this project was to collate and populate a daily and monthly rainfall database for southern Africa. Much time was spent detecting and correcting anomalies in the data. However, owing to the size of the databases it is impossible to produce one without any errors. It is hoped that these two rainfall databases (daily and monthly) will be used extensively and that future observed data and infilled rainfall values be added to maintain an up-to-date rainfall database.

Table ES3 Frequency analysis of the daily rainfall database

Rainfall (mm)	% of observed values	% of infilled values	% of observed and infilled values
zero	84.9093	77.4160	79.5348
0-5	6.7346	12.4999	10.7158
5-10	3.2739	5.9959	5.2536
10-25	3.5895	3.2651	3.4655
25-50	1.2073	0.6985	0.8559
50-100	0.2569	0.1129	0.1575
100-200	0.0261	0.0108	0.0155
200-300	0.0019	0.0007	0.0011
300-600	0.0005	0.0002	0.0003

DECLINE OF THE SOUTH AFRICAN RAIN GAUGE NETWORK

One of the first accounts, from Jan van Riebeeck's journal, of heavy rain in South Africa dates back to 22-23 July 1652 when the garden at the fort was washed away and the packing shed in the fort was 150 mm under water (The Chief Director, 1990). The Royal Observatory (SAWS Station No. 0020866 W) in Cape Town, however, is the earliest systematic recording station, with records dating back to 1850. The second active rainfall station, Ufumba (SAWS Station No. 0375383 W), which is located approximately 20 km north of Hluhluwe in northern KwaZulu-Natal, has records that date back to 1865, while an active rainfall station opened in Clanwilliam, in the Western Cape, around 1869. By 1880 the region which now comprises South Africa had more than 100 active daily recording stations and this number increased to a maximum of 3 841 in 1938, with a steady decline in the number of rainfall stations since then (Figure ES3). The current daily rainfall database was augmented around 1960 with data from neighbouring countries, but also shows a decline in the number of active rainfall stations since the mid-1980s (Figure ES3).

The question is often asked as to whether the decreasing number of active rain gauges is going to cause problems in the future. The number of active rainfall stations covering South Africa started to decline in 1938, but this decline has become more pronounced since 1980 (Figure ES3). An analysis was performed on the daily rainfall database to determine the number of active rainfall stations per year per 15 arc minute square in an attempt to determine if the closure of the stations affected the spatial coverage of the rainfall monitoring network. Initially a test for at least one active station per 15 arc minute square was performed (Figure ES4) and the trend is similar to the total number of active stations. The next tests, using at least 2 and then at least 5 stations per 15 arc minute square (Figure ES4), all display the same trend, which highlights the declining trend in the spatial coverage of the rainfall monitoring network.

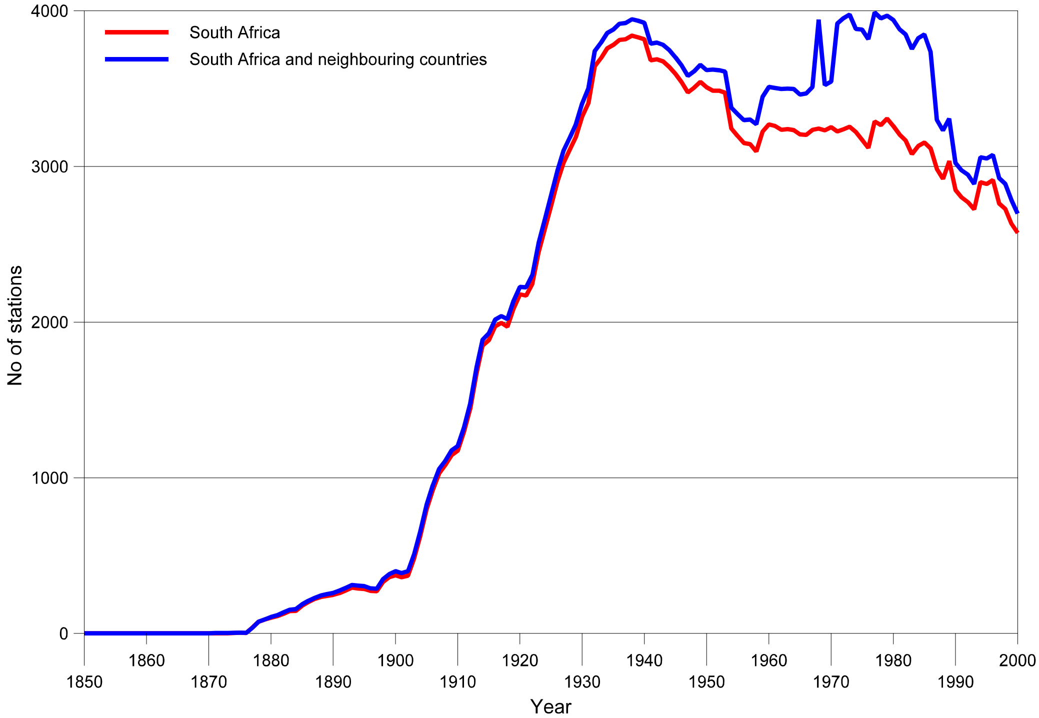


Figure ES3 Number of active rainfall stations over time in southern Africa

In conclusion, the number of active rainfall stations increased from 1 in 1850 to 3 841 in 1938. Thereafter, however, a sharp decrease in the number is noted and this decline tapers out for the period 1959 until 1981 whereafter a sharp decline is noticed once again (Figure ES4). It would appear that the decline in the number of stations, however, does not yet affect the spatial coverage, as stations exist in proximity to those that were closed down.

Missing rainfall records limit the use of these data as daily simulation models, for example, cannot function without a continuous dataset. There are a number of infilling algorithms that have been used in this research project and a complete list is contained in the full report.

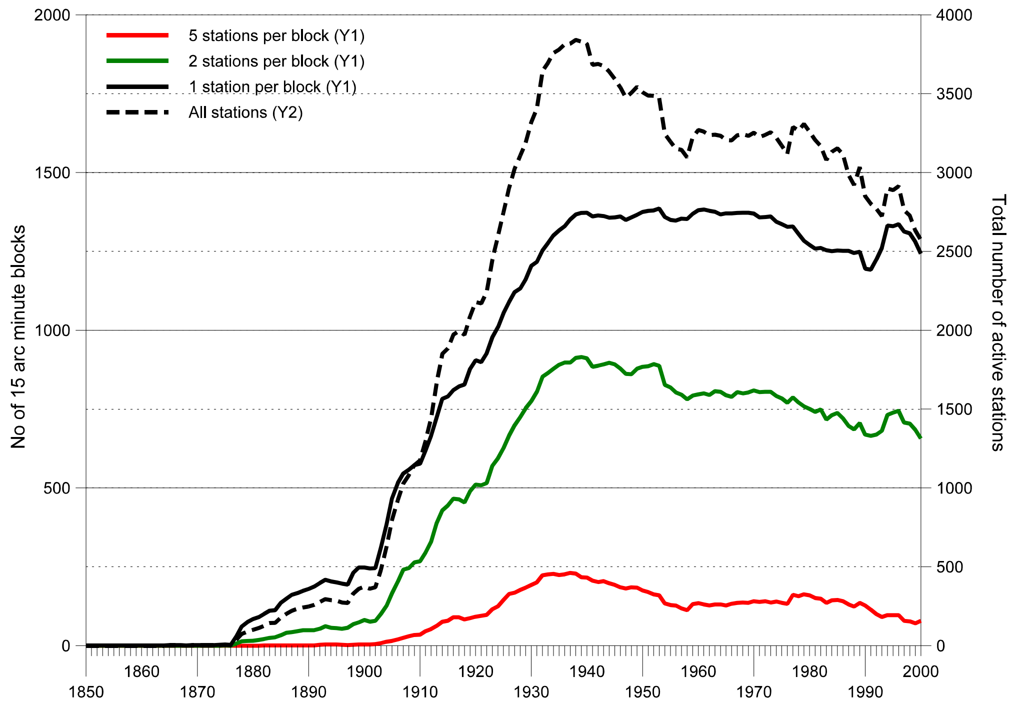


Figure ES4 Analysis of active stations with spatial bias removed

THE INFILLED RAINFALL DATA IN THE DATABASE

The four infilling techniques that have been used have more than doubled the size of the daily rainfall information base (Table ES4). The rainfall information base consists of 105 753 218 daily observed values with 236 154 934 infilled values. The total size of the observed and infilled rainfall database is thus 341 908 152.

Table ES4 Number of infilled daily rainfall values in the rainfall database

Infilling technique	Number of daily values
Expectation Maximisation Algorithm	113 869 517
Median Ratio Method	40 823 148
Inverse Distance Weighting	81 451 381
Monthly Infilling Technique for < 2 mm	10 888
Total	236 154 934

The infilling process has also increased the size of the annual database considerably from an initial 5 118 stations with more than 15 years of complete record to 9 641 stations that have more than 15 years of record.

CHOICE OF INTERPOLATION / REGRESSION TECHNIQUE

It is unwise to use data with any interpolation technique without carefully considering how the results will be affected by the assumptions inherent in the method (Burrough, 1986).

Inverse distance weighting (IDW) techniques are the method of choice when converting point data to a raster format as they are easy to use and are not computer intensive. When the density of the point data is sufficient and the variation in the parameter to be converted to raster is not too complex then IDW would suffice. Rainfall in areas of complex topography, and where the distribution of the stations is lacking in these areas, is better converted to raster using an approach that relies on additional explanatory variables.

The area surrounding the Jonkershoek mountains near Stellenbosch in the Western Cape Province lends itself to a pictorial illustration of the main differences between a regression approach and an IDW approach. The MAP values increase from 1 085 mm to 3 199 mm over a distance of approximately 8.5 km and the altitude increases from approximately 230 m to 1 300 m (Figure ES5).

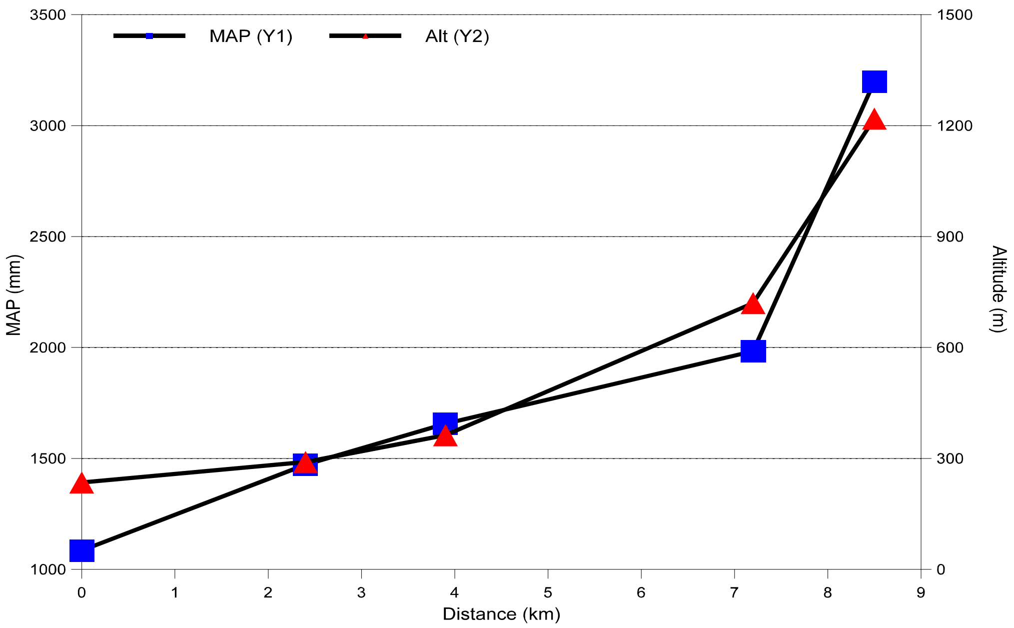


Figure ES5 MAP and altitude relationships along the Jonkershoek mountain range

The area surrounding Jonkershoek has rainfall stations at varying altitudes (Figure ES6) and the areas towards Paarl and Franschoek also have a reasonable rainfall/altitudinal variation. The area further northeast, however, is covered by only two rainfall stations. Two rainfall surfaces, created using IDW and multiple regression have been draped onto a 3-dimension digital elevation model (DEM) to illustrate the differences.

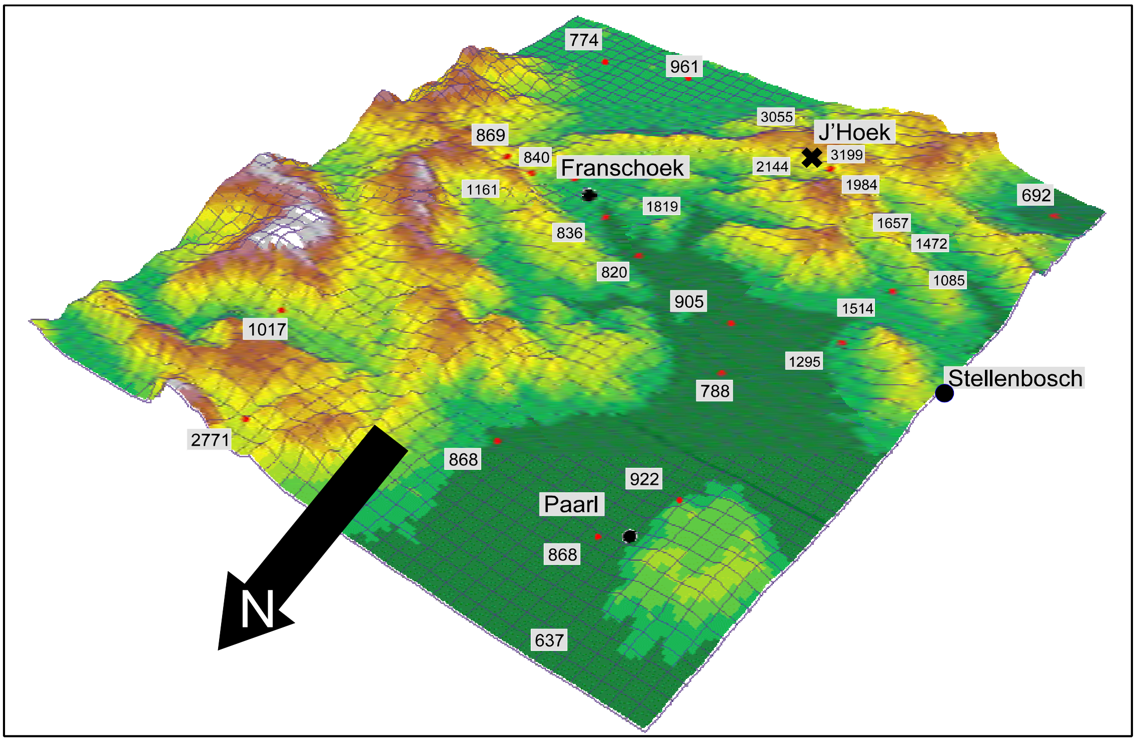


Figure ES6 Digital terrain model (DTM) and rainfall station locations in the area surrounding Jonkershoek

When using an IDW approach the relationship of an increase in MAP with altitude around Jonkershoek is evident, but the mountainous area further northeast has a similar MAP value (Figure ES7). The regression approach, on the other hand, has varying MAP values in the ungauged mountainous area northeast of Jonkershoek. This example highlights the fact that IDW relies on a dense gauge distribution in physiographically complex areas, whereas regression can utilise a relationship that was built up some distance away to estimate the MAP in an area of complex topography that also has a spare rain gauge network.

It is evident, therefore, that techniques used to estimate the MAP raster from the rainfall stations should allow for the inclusion of explanatory variables. This means that of the techniques reviewed, only cokriging and regression remain. The limitations and shortcomings of the cokriging approach therefore suggest the acceptance of the Geographically Weighted Regression (GWR) approach, as suggested by Hughes *et al.* (2001).

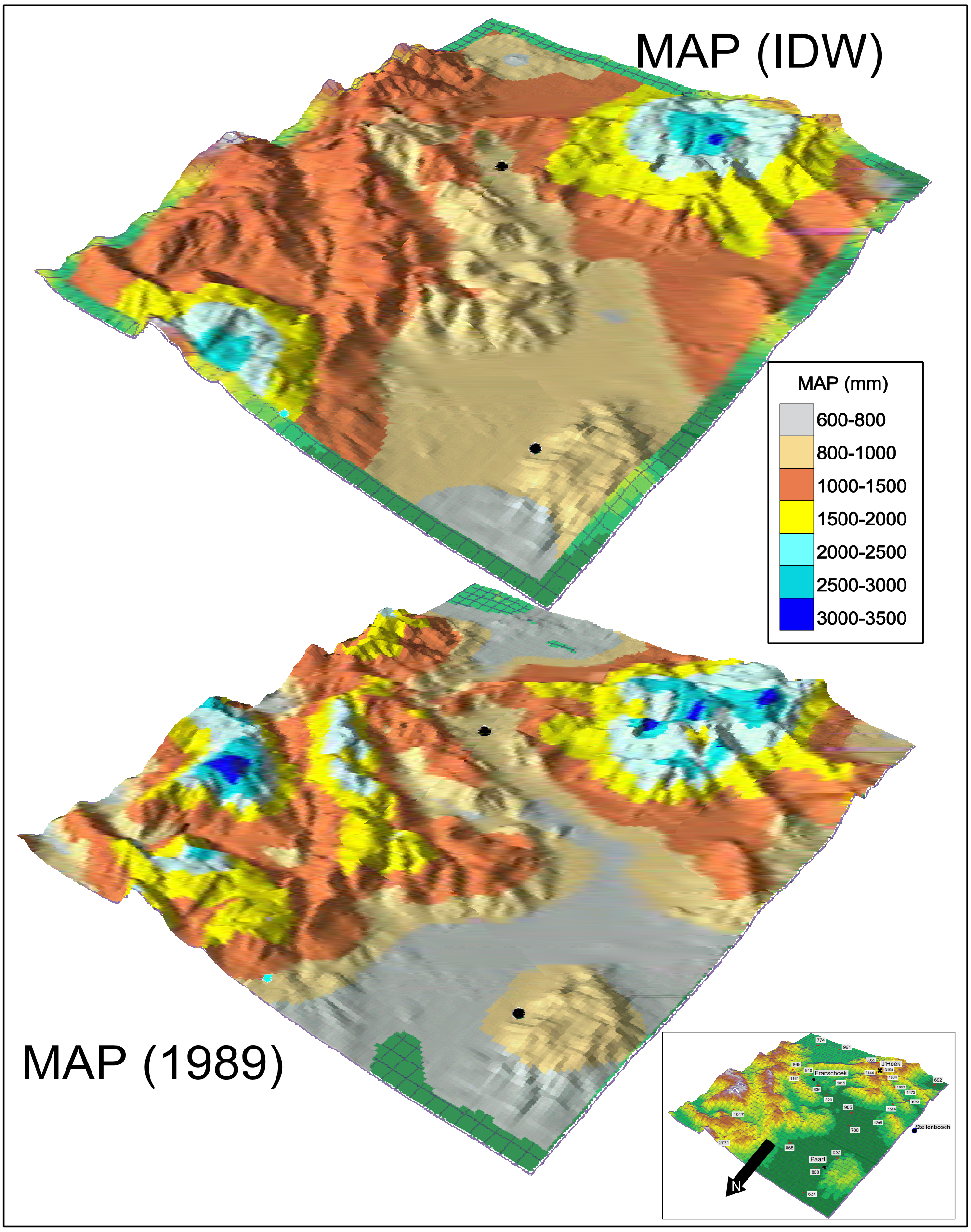


Figure ES7 Surfaces of MAP draped over the DTM

AREA OF INTEREST

At the request of the project Steering Committee, the raster rainfall surfaces were extended to the boundaries of those northern Quaternary catchments from which streamflow enters rivers forming part of South Africa (Figure ES8).

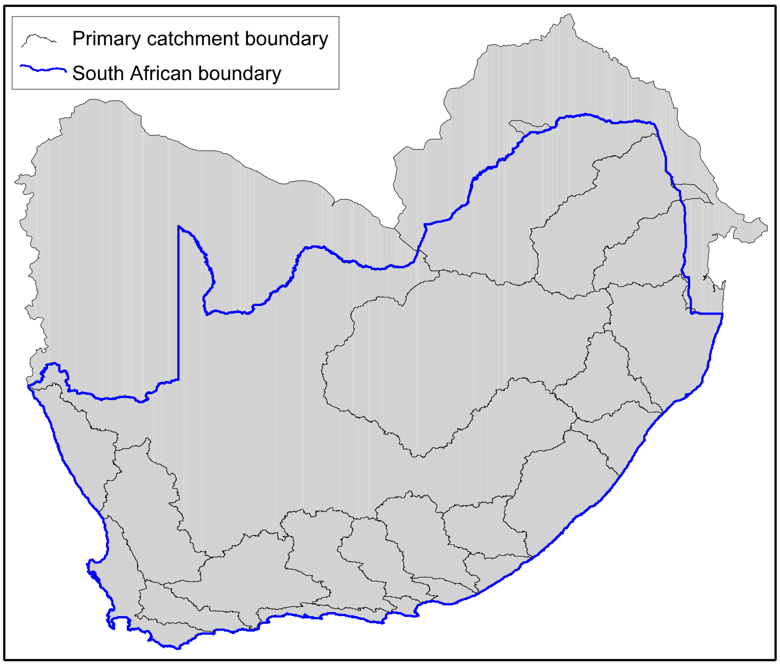


Figure ES8 Area of interest selected on Primary catchment boundaries

ANNUAL PRECIPITATION

Although rain gauges measure rainfall at a point, their record can serve a broader purpose because they reflect locality factors such as continentality, altitude, longitude and aspect, one or more of which often govern the incidence of rainfall (Whitmore, 1967).

MEAN ANNUAL PRECIPITATION

Mean annual precipitation is one of the most widely used variables in hydrological design, water resources planning and agrohydrology. Mean annual precipitation is the total annual precipitation calculated from all yearly total precipitation data for the period of interest divided by the number of years in the period of interest. In this particular study one missing day of rainfall was taken to imply a missing month, which resulted in the whole year being excluded from the calculations.

GEOGRAPHICALLY WEIGHTED REGRESSION APPROACH

The research into finding a suitable technique to represent the spatial variation of MAP concluded that GWR should be used (Hughes *et al*., 2001; Lynch, 2001a; Brunsdon, 2002). Once this had been decided, the decision on which set of explanatory variables was to be used began. A host of different approaches was used to attempt to find which set of explanatory variables would produce the best, or most appropriate, spatial estimate of MAP. Hutchinson (1998) found that the cross validation technique does not always

represent a reliable estimate of model error, especially when short range correlation in data is present.
Another method of deriving the bandwidth and selection of the explanatory variables, which provides a trade-off between goodness-of-fit and degrees of freedom, is to minimise the Akaike Information Criteria (AIC). The AIC has the advantage of being more general in application than cross-validation (Fotheringham *et al*., 2002).

ADJUSTMENT OF THE G.W.R. RASTER

The majority of the regression procedures endeavour to minimise the residuals. The aims of selecting different sets of explanatory variables is to produce a model that will fit the data the best, in other words, the model that produces the smallest residuals. These residuals, however, also mean that when a raster at the station pixel is queried, a different MAP value to that recorded in the point rainfall database may be obtained (e.g. as in Figure ES9). A technique, that was used successfully in the Dent *et al*. (1989) rainfall mapping programme, and which uses these residuals to enhance the GWR raster, was once again used in this research. If the station pixel's MAP is not adjusted, then one would not be able to overlay the station MAP values on the isohyetal map because, as in the example of Figure ES9, the 850 mm MAP point value does not fall between the 700-800 mm isohyets.

The residuals, i.e. the differences between the observed MAPs and the MAPs estimated using GWR, are interpolated onto a rectangular one arc minute raster using IDW. This raster is then added to the one arc minute MAP raster generated using GWR (Figure ES10). This process has adjusted the initial GWR surface locally, to fit where there are observed values, as well as globally at the ungauged pixels, using the residual information. In other words, the surface was not only adjusted to fit the MAP at the rainfall stations, but the surrounding areas were also adjusted according to the interpolated residual surface.

This process can also be explained by thinking of pulling a blanket up with one hand and pressing down on the bed with the other hand. The areas surrounding the hands are also adjusted and not only where the hands touch the blanket, which would happen if only the pixels, where observed data are available, were adjusted.

THE M.A.P. RASTER ESTIMATED USING G.W.R.

The set of explanatory variables that generate the most plausible estimate of MAP (Figure ES11) are ialtCV, latlong, xplusy, xx and slope. These abbreviations represent respectively, the coefficient of variation (CV) of a 5 arc minute mask of altitude values, product of the latitude and longitude co-ordinates (degrees decimal) of the pixel, sum of the latitude and longitude co-ordinates (degrees decimal) of the pixel, square of the longitude co-ordinates (degrees decimal) of the pixel, and slope in degrees of the 8 pixels surrounding the pixel in question. The fact that distance from the sea does not feature as a variable is explained by the fact that the interactions of these selected explanatory variables vary spatially and could act as surrogates for it.

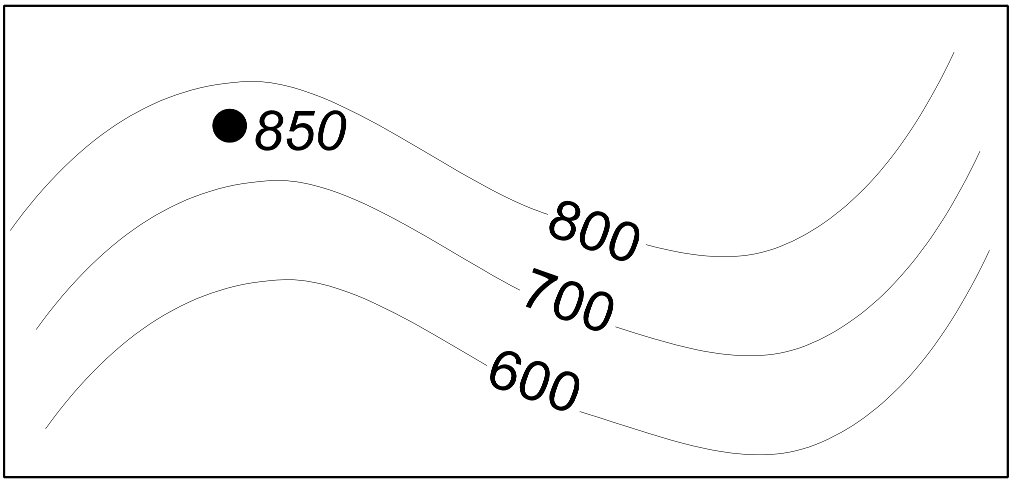


Figure ES9 Reason for adjusting the regressed surface

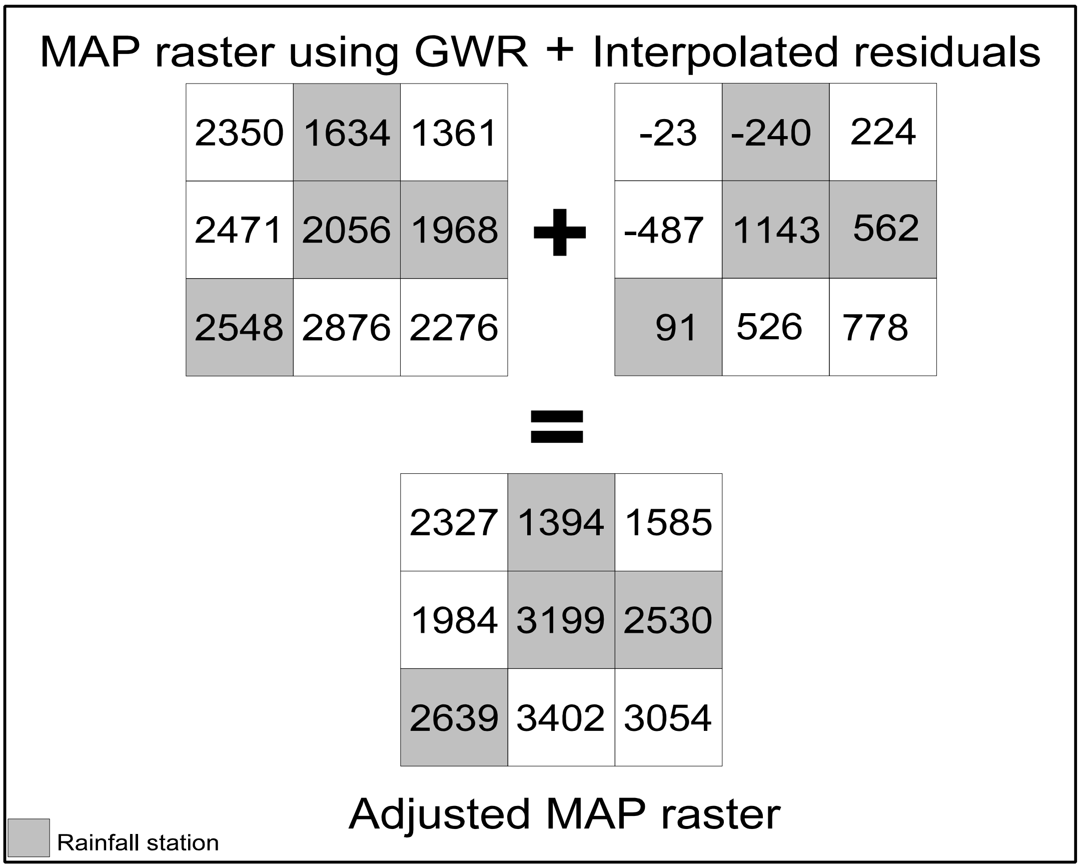


Figure ES10 Technique for adjusting the regressed MAP surface using residuals

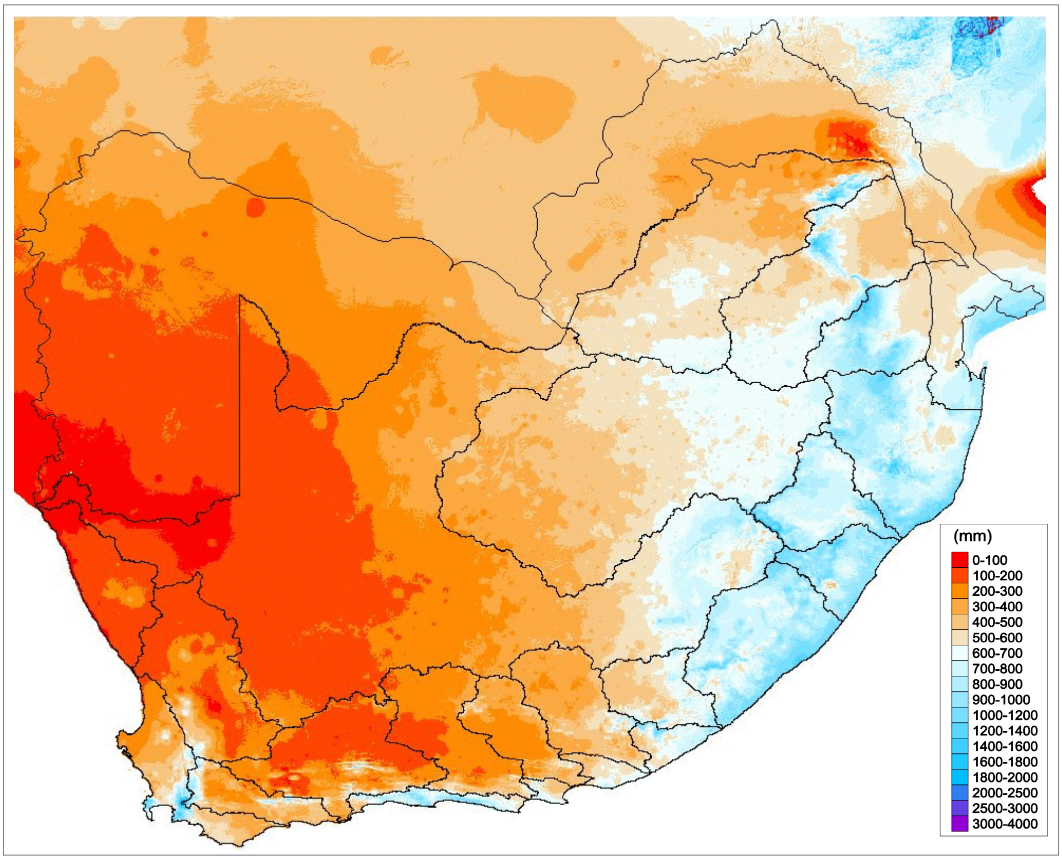


Figure ES11 MAP surface determined using Geographically Weighted Regression (GWR)

The regression coefficients also vary spatially over the area of interest (Figure ES12), which once again emphasises the spatial non-stationarity of the data. The interactions of the coefficients at some pixels are akin to the introduction of a "new" explanatory variable, i.e. in certain areas the product of latitude and longitude (latlong), for example, might have the same effect if distance from the sea, for example, were used at that location.

ANALYSIS OF M.A.P. RASTERS DERIVED USING THE DIFFERENT TECHNIQUES

This section includes some analysis that is performed on the MAP rasters estimated using the different regression and interpolation techniques. The minimum and maximum values (Table ES5) differ, as one would expect. The maximum value of the MAP raster estimated using stochastic values (McNeill *et al*., 1994) is very low because the dataset used to estimate the stochastic parameters did not include some of the key rainfall stations in the Drakensberg and Jonkershoek areas. The CV values are all in the same range, but visual inspection of the MAP surfaces do, however, suggest that some are more smoothed than others.

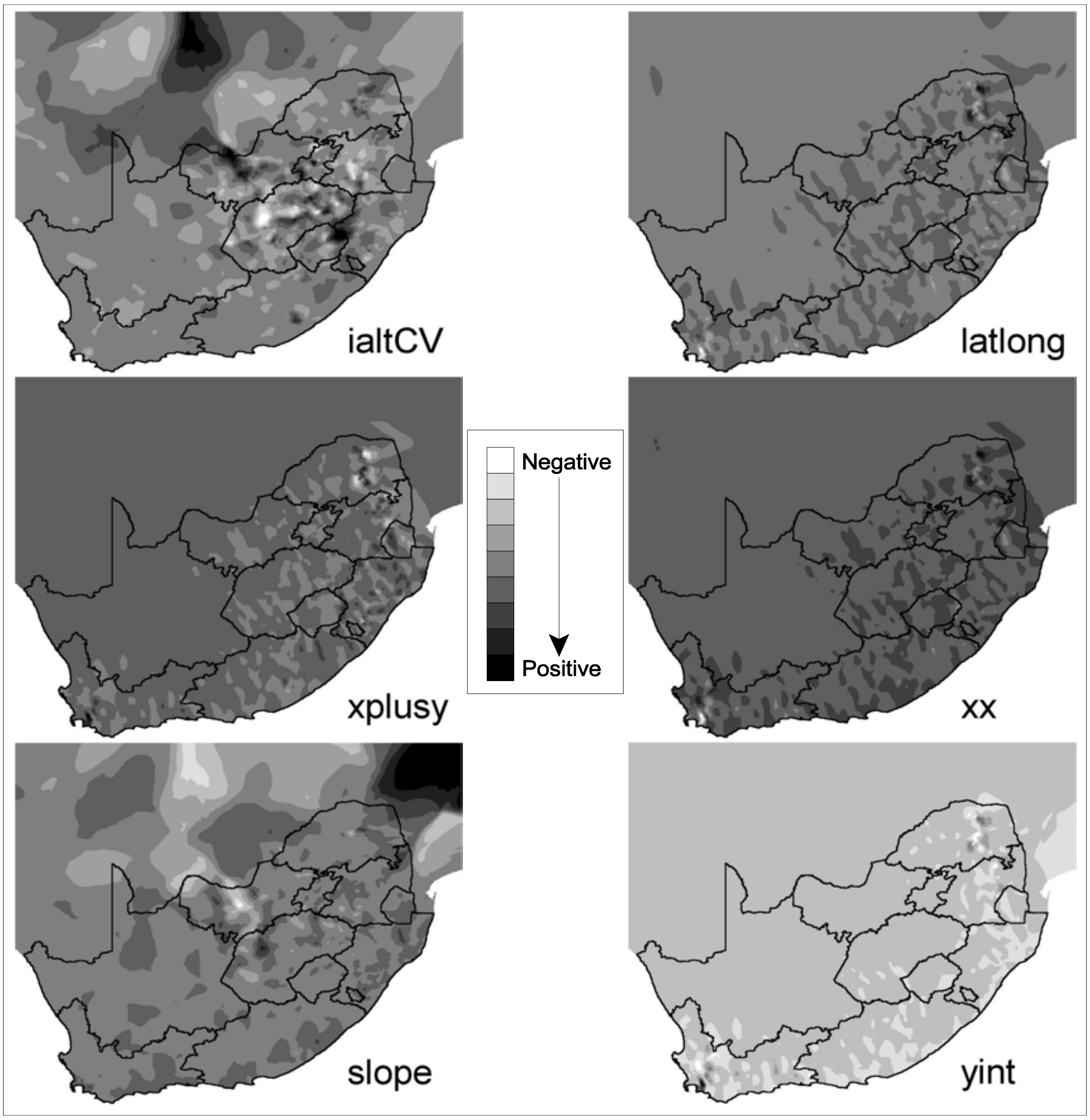


Figure ES12 Spatial variations of the final GWR coefficients used to estimate MAP

Table ES5 Comparative statistics of the MAP for South Africa when using different interpolation and regression techniques (Abbreviations in full text)

Technique / Statistic	MAP'89	GWR	IDW	Cokriging	Stochastic
Minimum (mm)	20	1	43	45	51
Maximum (mm)	3345	3198	3199	2422	1999
Average (mm)	461.62	447.59	452.26	446.15	470.32
CV (%)	56.81	55.26	53.78	54.68	54.54

All the methods reviewed in this section should produce similar estimates of MAP in areas where there is no complex topography. The surface roughness index, ialtCV, is used to exclude topographically complex areas. In others words only pixels where the ialtCV raster values are greater than 5 are selected. These mountainous regions form a mask which is used to determine if there exists any similarity between the MAP rasters generated using a variety of different interpolation and regression techniques.

MONTHLY PRECIPITATION

Raster surfaces of monthly precipitation are calculated using the mean and the median statistics. The Dent *et al*. (1989) rainfall project only considered the surfaces of median monthly precipitation and this was due to a time and computing power constraint. The technique that is used to create these surfaces is similar to that used by Dent *et al*. (1989) which consists, *inter alia,* of expressing the median or mean monthly values as a ratio of the MAP values (done using the existing scatter of rain gauges). These ratios are then interpolated onto a rectangular raster, at a spatial resolution of one arc minute (Figure ES13). This interpolated raster is then multiplied by the raster of estimated MAP values generated using GWR (Figure ES11) and this is repeated for each month. This methodology has been widely accepted as a means of creating monthly surfaces of rainfall without using a regression approach. The use of GWR to estimate monthly rainfall surfaces is not recommended owing to the fact that the monthly data are more variable than the MAP data and the process would be extremely time consuming and computer intensive.

One of the initial aims of this project was to produce rasters of daily rainfall at a spatial resolution of one arc minute. One of these rasters of daily rainfall values takes up roughly 0.5 MB of hard disk space, which means that 365 rasters occupy about 182 MB. This will imply that a CD-ROM will hold approximately 3.5 years of daily rasters. An example of 366 rasters for 1980, with their point shape files, is included on the accompanying CD-ROM for perusal. The rasters were created using IDW. A test was performed at the one arc minute pixel, represented by station 0152482 W, and the 366 pixels produce an annual total of 781.11 mm whereas the observed total for 1980 is 781.6 mm. This slight difference can be attributed to rounding errors in the IDW approach.

DESCRIPTION OF THE DATABASE STRUCTURE

The daily rainfall database consists of more that 300 million rainfall values for approximately 14 000 stations. The data in this database originate from many different organisations and individuals, each having their own structure and set of quality control codes. The author has an extensive library of Fortran based routines to analyse time series information. A flat direct access ASCII data structure is not the most space efficient format, but it allows for quick access to any record in the database. A suite of programs was developed to assist the user in extracting rainfall data and storing it in a number of common data formats.

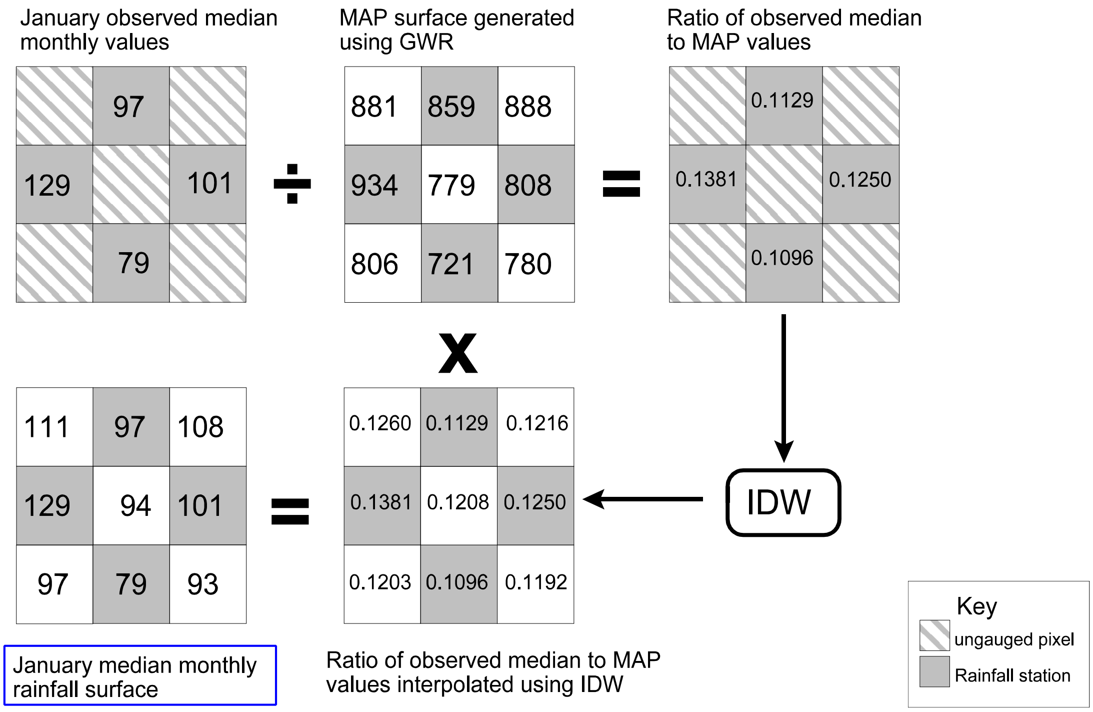


Figure ES13 Procedure used to calculate the median or mean monthly rainfall surfaces

TECHNOLOGY TRANSFER

During the course of this three year WRC funded project the output that had emanated from this component of the project, prior to the publication of the final report, included the following:

- two invited lectures,
- two peer reviewed journal articles,
- two peer reviewed conference proceedings and oral presentations,
- three local conference proceedings and oral presentations,
- one popular article in the local newspaper, and
- five poster presentations.

Copies of the current daily and monthly time-series database have also been requested by a number of research organisations.

FULFILLING CONTRACT OBLIGATIONS

At the beginning of this Executive Summary the objectives of this component of the project were outlined. The Executive Summary, and especially the full report that underpins it, have shown amply that the objectives have been fulfilled. Already the quality controlled rainfall datasets have become a national standard and are being widely used in numerous WRC and other projects.

RECOMMENDATIONS FOR FUTURE RESEARCH

The successful research and use of rainfall data relies on all the data being in a single database. There are many reasons for this. For example, if rainfall data were sought and acquired from different organisations the data would firstly be re-formatted to a single database standard, or format, and the missing data would then need to be infilled, whereas a single database consisting of data that are collated from a host of different organisations and individuals saves an enormous amount of valuable research time. A single database also allows the infilling procedures, for example, to use the best available temporal and spatial distribution of rainfall data. A plea is therefore made that there always be an organisation with a dedicated and interested individual for the collation and creation of a rainfall database for southern Africa.

The research into different interpolation and regression techniques to map annual rainfall has yielded success in the form of the Geographically Weighted Regression (GWR, Brunsdon *et al*., 1996), technique. The spatial interpolation and regression of daily rainfall surfaces, however, has not been entirely successful. The main factors for this are the sheer volume of daily values compared to twelve monthly and a single MAP surface, and the spatial variability of daily values.

Lynch *et al*. (2001a) noted an increase in MAP from 1980 onwards for the rain gauges on the Irwin farm in an area south of Potchefstroom. Initially, this was regarded as an anomaly in the observations, but scrutiny of the original handwritten records, and telephonic contact with Mr Neville Irwin, suggests that the data are correct and that there is definitely an increase in MAP post 1980. A number of rainfall stations in the immediate area were analysed and a similar trend was found. Future research is suggested to determine how widespread this trend is and to determine if similar trends appear elsewhere and at different time periods.

There is a wealth of information pertaining to the occurrences of El Niňo and La Niňa, but little research has been undertaken to determine the extent to which these phenomena really affect South African rainfall amounts. The rainfall database developed in this project will assist research into a possible link between El Niňo and La Niña and the dry and wet spells experienced in South Africa.

ACKNOWLEDGEMENTS

The success of this rainfall component of Project K5/1156 is due largely to the guidance of the late Mr Hugo Maaren, and a special word of appreciation is expressed to him.

The author should like to thank the many people who have contributed to this rainfall component of the project during the past three years. The following organisations and individuals deserve special mention:

The chairman and members of the Project Steering Committee have been particularly helpful in making suggestions and providing constructive advice:

Mr H Maaren (deceased)	Water Research Commission (Chairman)
Dr SS Mkhize	Water Research Commission
Prof PWL Lyne	BEEH, University of Natal
Mr MD Watson	Department of Water Affairs and Forestry
Mr AC Kruger	South African Weather Service
Mr K Monnik	Agricultural Research Council
Mr RA Chapman	CSIR, Environmentek
Prof JC Smithers	BEEH, University of Natal
Dr MC Dent	Computing Centre for Water Research
Mr BC Hohls	Department of Water Affairs and Forestry
Dr CS Everson	CSIR, Environmentek
Dr MS Basson	BKS (Pty) Ltd
Prof GGS Pegram	School of Civil Engineering, Surveying and Construction, University of Natal
Mr P Visser	South African Weather Service
Mr SS Hine	GIMS (Pty) Ltd.

The following organisations and individuals are thanked for their valuable assistance and co-operation during the course of this project:

- Mr Richard Kunz, Institute for Commercial Forestry Research, Pietermaritzburg, for making available monthly rainfall data for 445 stations across South Africa,
- Dr Louis du Pisani, for providing a monthly dataset of rainfall values across Namibia,
- Mr Helio Banze, who assisted in obtaining some daily rainfall data for Moçambique,
- Dr Jerry Ndamba, Institute of Water and Sanitation Development, Harare, Zimbabwe, for the daily rainfall records of 66 rainfall stations across Zimbabwe,
- Prof Jeff Smithers, BEEH, for infilling the missing daily rainfall values using the EMA suite of programs,
- Mr Kevin Meier, LRI, Pietermaritzburg, for infilling missing daily rainfall values using a modified Inverse Distance Weighting technique,
- the erstwhile CCWR for providing the initial daily rainfall database,
- Mr Chris Whyte, erstwhile of Enviromap, for making the digital contour data of Lesotho available,
- Mr Rodney Harding, SA Sugar Association Experiment Station, for providing the daily rainfall data of all their stations,
- Ms Tracey Gill and Ms Elsa de Jager, of the South African Weather Service, for their assistance in providing SAWS daily rainfall data,
- Ms Michelle Warburton, Ms Karen King, Ms Melissa van Rooyen, Mr James Harvey, Mr Gary Morgan, Mr Darryn Knoesen and Mr Kevin Bursey, the Hydrology Honours class of 2002, for assisting in the literature review,
- Mr Mark Summerton, Umgeni Water, for providing their daily rainfall data,
- Ms Joyce Zulu, for her input as a student assistant,
- Mr James Thorpe, of the erstwhile CCWR, for assistance in the preparation of the mean annual precipitation map using stochastic rainfall data,
- Mr and Mrs Irwin, for the donation of the daily rainfall journals and for assisting in interpreting some of the anomalies in their rainfall figures,
- Ms Fay Hughes and Prof Linda Haines, for their work on applying the GWR technique to estimate MAP values,

- Prof Roland Schulze, the project leader, for his valuable assistance, and
- Mr Mark Horan, for all his valuable proof reading efforts and his input on a continuous basis.

NOMENCLATURE

There are often terms and expressions that one uses interchangeably and this section will assist the reader in understanding these terms:

Cell and pixels are square objects of the same values. They can also be thought of as the cells that make up a spreadsheet.

Raster, grid, continuous surface and surface all refer to an object that stores spatial data in a locational data format in which space is partitioned into square cells, and each cell stores a numeric data value.

South Africa, in the context of this study, comprises of the Republic of South Africa and the Kingdoms of Lesotho and Swaziland.

Explanatory and independent variables are normally denoted by X and if Y denotes the dependent variable then a functional relationship of the form Y = f(X) exists. In this document rainfall is the dependent variable and the independent variables include, *inter alia,* altitude, surface roughness and distance from the sea.

Rainfall station, station, rain gauge and gauge refer to the position where the rainfall amount is measured and is represented as a one by one arc minute pixel.

South African Weather Bureau (SAWB) and the South African Weather Service (SAWS) refer to the same organisation, with the latter being the new name.

Precipitation and rainfall will be used synonymously throughout this document to designate rainfall, since fog is not measured at sufficient sites to warrant inclusion and the contribution of snow is not important to total precipitation in southern Africa.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued) **Page**

LIST OF TABLES

CHAPTER 1

THE IMPORTANCE OF RAINFALL DATA

Rainfall measurements were recorded as early as 400 BC and, since its inception, both the principles and the purpose of rainfall measurement have remained unchanged (Biswas, 1967; Ward, 1975). The rainfall of an area helps to structure society in a geographical sense. Water is an essential element for life, thus the more water that exists in an environment, the more potential that environment has for sustenance of life. Early settlers settled in higher rainfall areas, and towns have been built up in and around these areas, shaping settlement as we experience it today (Antevs, 1938).

In order to obtain rainfall data the rainfall needs to be measured. Rainfall measurements can be undertaken by numerous different methods. The most common method is the use of a standard daily non-recording rain gauge, but estimation by radar and satellite is practised as well. While radar and satellite imaging for rainfall estimates are able to provide real time, areal estimates of rainfall values, the primary source of rainfall data is still provided by the rain gauge (Seed and Austin, 1990). This is so mainly because rain gauges are cheap and generally reliable. Rain gauge data are also available for longer time periods, which is advantageous in many respects. One hopes that over time, the rain gauge network density will increase, so as to provide better areal estimates of rainfall from the use of point measurements.

The measurement of rainfall is a simple procedure provided that accuracy is not essential, as an exact measurement of rainfall is impossible to obtain owing to the random and systematic errors which occur in measuring rainfall (Schultz, 1985). As no ‘true rainfall amount’ can be achieved, one can only attempt to improve the estimation of rainfall amounts by minimising the known errors, which are the systematic errors that are associated with the rain gauge used to measure rainfall amounts. Empirical equations have been derived which can be used to account for the systematic errors in point rainfall measurements (Schultz, 1985). Added to these errors, rainfall amounts are extrapolated to give an areal average of rainfall. Boughton (1981) stated that deficiencies of 10-20% could be expected in point measurements of rainfall and that a further 10-20% error is likely in extrapolating data from a point measurement to an areal average. To aid in decreasing this error a sufficiently dense and suitably spaced rain gauge network needs to be used (Schultz, 1985).

Mean annual precipitation (MAP) and the variability thereof also play a large role in determining many aspects of ecosystems. These include distribution of species (Harrison *et al*., 1997), breeding biology in terms of timing, number of offspring produced, degrees of parental care, and fluctuations in populations. Certain rodent species, for example, show marked increase in population sizes following periods of below average rainfall (Mills and Hes, 1997). High rainfall over the past two years (2000-2001) has prompted greater numbers of rare birds that migrate according to local conditions, such as Black Coucal and European Marsh Harrier, to reach South Africa. Indeed it is largely as a result of the variety of rainfall patterns within the country that there is such an incredible diversity of plants and animals.

Seeber (1983), for example, used MAP in an extreme-value expression for the determination of the cumulative distribution of point rainfall rate in a model for use in the telecommunications industry.

The rainfall data that are given to a hydrologist need to be checked thoroughly as 80% of the time should be spent using the data rather than checking the data for inconsistencies (Beven, 2000).

CHAPTER 2

RAINFALL DATASETS

Mean annual precipitation is perhaps the most widely used variable in hydrological design, water resources planning and agrohydrology. The basic sources for MAP in South Africa between the mid 1960s until the end of the 1980s were the 1:250 000 average rainfall map series compiled and drawn by the erstwhile Hydrological Research Division of what was then the Department of Water Affairs. Since then, the rainfall records had lengthened by more than 20 years and techniques of analysis and computerised mapping had increased dramatically and the next series of maps together with the first set of digital raster rainfall values became available towards the end of 1989.

There are many organisations and private individuals that record daily and monthly rainfall data in South Africa and its neighbouring countries. The majority of the rainfall data are recorded at a daily time-step, although there are a number of sites that only record rainfall figures at a monthly time-step. The rainfall data for the Kingdoms of Lesotho and Swaziland have historically been included in the South African dataset and the *status quo* will remain. The current rainfall database, assimilated in this project, consists of data from South Africa, Namibia, Botswana, Zimbabwe and Moçambique.

The task of obtaining rainfall data from the different organisations and individuals is not an easy one. When obtaining the data from neighbouring regions the task becomes more difficult, as there are language and other problems to contend with.

2.1 POINT AND RASTER RAINFALL DATA

The rainfall data described in this research fall into three categories, *viz*.

- point temporal rainfall data that are recorded at a site,
- infilled point temporal rainfall values that are estimated at a site, and
- spatial rainfall values that are stored in a raster.

The point rainfall data are normally recorded at a daily time-step and sometimes at a monthly time-step. These point rainfall values are converted onto a rectangular grid or raster using various regression and interpolation techniques that are discussed in subsequent chapters in more detail.

2.2 SOUTH AFRICA

The initial daily and monthly rainfall datasets were acquired from the erstwhile CCWR early in 2000. The datasets were initially developed for a WRC funded project titled "*Mapping of Mean Annual Precipitation and Other Rainfall Statistics over Southern Africa*" (Dent *et al*., 1989) and were updated annually until the start of this current WRC project.

The current daily rainfall database consists of data from a wide variety of organisations and individuals (Table 1) that include, *inter alia*:
- the South African Weather Service (SAWS),

- the Agricultural Research Council (ARC),
- the South African Sugar Association (SASA), and
- a large number of municipalities, private companies and individuals (PVT).

All the recorders of rainfall data are hereby acknowledged gratefully for their cooperation and diligence without which this project could not have succeeded.

Table 1 Number of rainfall stations per organisation for South African rainfall data

Organisation	No. of stations
SAWS	8 281
ARC	2 661
SASA	161
PVT	1 050
Total	12 153

The Institute for Commercial Forestry Research (ICFR) supplied, to this project, a database of 445 stations, data from which were provided as monthly accumulations.

2.3 NEIGHBOURING COUNTRIES

The datasets for Namibia, Botswana, Zimbabwe and Moçambique (Table 2) that are in this current rainfall database are by no means complete. Some of the countries supplied daily data with the proviso that these data only be used in this project and may not be distributed. The monthly totals, however, that are derived from these daily data may be distributed freely as the daily values cannot be reproduced from the monthly totals.

Table 2 Number of rainfall stations from the neighbouring countries

Country	No. of stations
Namibia	640
Botswana	306
Zimbabwe	69
Moçambique	83
Total	1 098

Monthly rainfall data were obtained for 32 stations in Moçambique. It is not necessary to elaborate on the amount and time spent in obtaining these data, suffice to say that a University of Natal Hydrology Honours student from Moçambique, Helio Banze, was of great assistance. Dr Jerry Ndamba supplied the project with daily rainfall data for 66 rainfall stations scattered across Zimbabwe and Dr Louis du Pisani made rainfall data for Namibia available.

There are many tales of unanswered requests for rainfall data, but the people who responded are acknowledged gratefully. Without their support this project would not have been possible. The final rainfall database consists of daily and monthly values for more than 13 000 stations (Figure 1). These data were collected up to 21 November 2002.

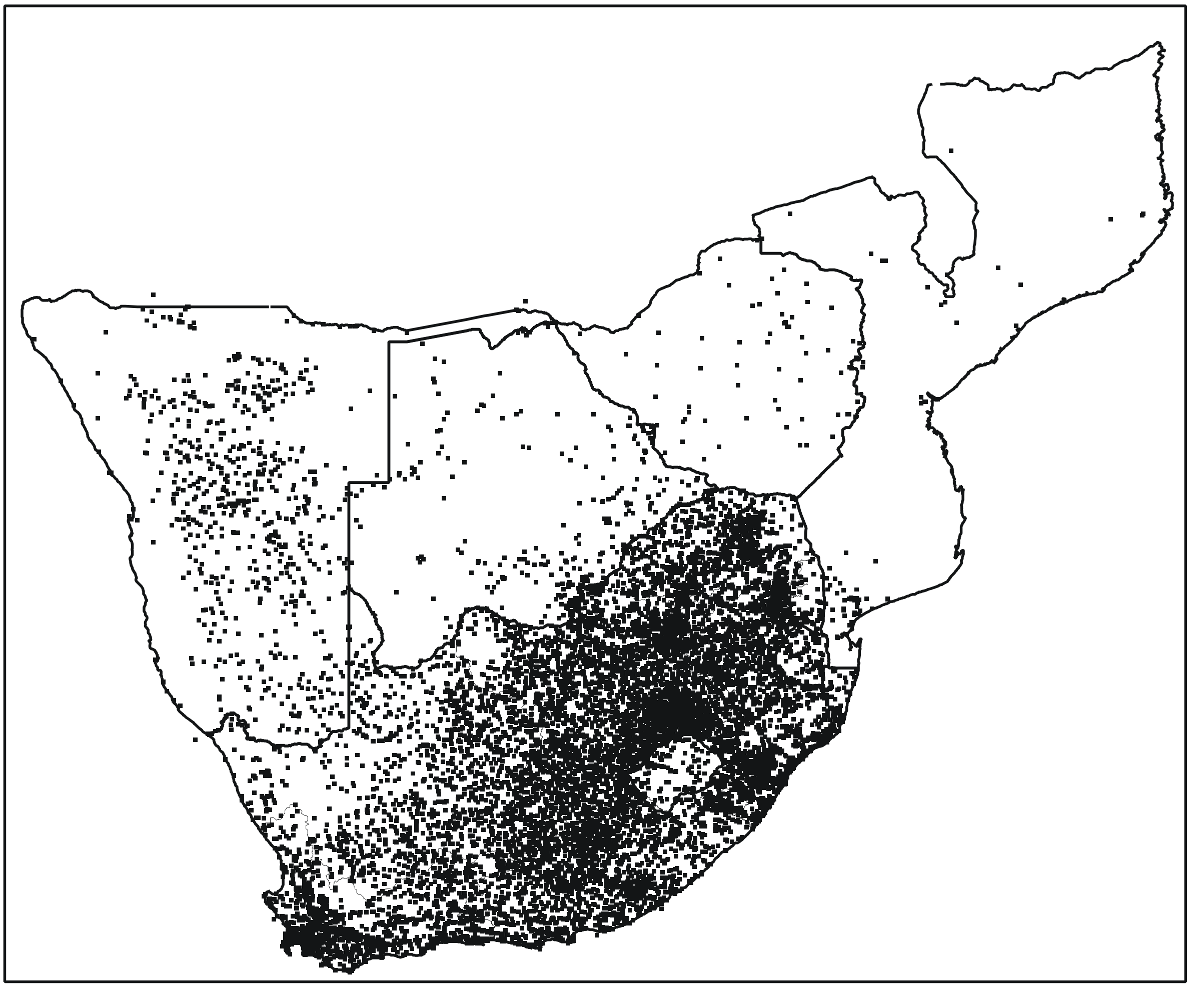


Figure 1 Location of all the rainfall stations available to this project

2.4 QUALITY CONTROL

A rainfall database that has in excess of 300 million values is bound to have some errors. Fotheringham *et al*. (2000) suggest that one needs to find out how *useful* the data are and not if they are *completely free of errors*.

Some of the known errors in the data only become apparent when the rainfall data are used in hydrological modelling exercises and the model results are tested against measured streamflow values. The most common of these errors is the time at which the gauge is read, as the standard time of recording rainfall in South Africa is at eight o'clock each morning. These phasing errors, however, have hardly any effect on the monthly and annual totals.

A rainfall amount of 597 mm was recorded at St Lucia Lake on 31 January 1984 when cyclone Domoina struck the northeastern part of South Africa. It is the largest 1 day total of rainfall recorded thus far in South Africa. ne of the first quality control procedures to be performed on the daily rainfall data was, therefore, to flag all rainfall amounts greater than 597 mm as suspect data.

A paper copy (A1 size) of bar charts depicting total of monthly rainfall amounts was received from the Town Clerk of Tsumeb in Namibia. This archaic paper database (Figure 2) with rainfall records from 1907 until 2000 was viewed with some humour as it has the monthly totals stencilled with amounts in inches and millimetres. Some of these monthly totals were compared with those in the historic rainfall database received from the erstwhile CCWR and many were found to differ by a factor of roughly 2.5 (Figure 2). Further analysis, however, revealed that the CCWR records for this station, prior to 1954, were erroneously stored in imperial units. The lesson learned is to check every piece of information. The complete database was scanned to detect and correct, where possible, more of these linear-plateau time series anomalies where both imperial and metric values are present.

A suite of different computer routines was developed to detect and correct suspect and erroneous rainfall amounts in this large rainfall database, *viz*. in excess of 300 million daily rainfall values and roughly 12 million monthly and annual totals. The initial quality control tests were performed on the monthly data. A computer program, *top10.f90*, was developed to analyse for each station, *inter alia*: the mean/median annual precipitation ratios, stations that have coefficient of variation (CV) of annual totals in excess of 80%, and the maximum, median, standard deviation, skewness and kurtosis values at a monthly and annual time-step. Each set of these statistics is then ranked and the largest 10 are selected and each station is assigned a score of 1 to 10 (10 if it is the highest) depending on whether the station fell into that particular top 10 ranking. After all these statistics are ranked the stations are ranked according to their scores. The station with the highest score is 0022190 W (PARADISE ESTATE-MOUNTAIN) in the Western Cape province (Figure 3). The annual totals between 1945 and 1951 are incorrect. No extreme events, for South Africa, are listed for 1945 and in 1946 there were two severe hailstorms, one in the Free State and the other in the vicinity of Sabie (The Chief Director, 1990). Perusal of the daily records for 0022190 W do not reveal a systematic error for this period and these data should therefore be treated as suspect.

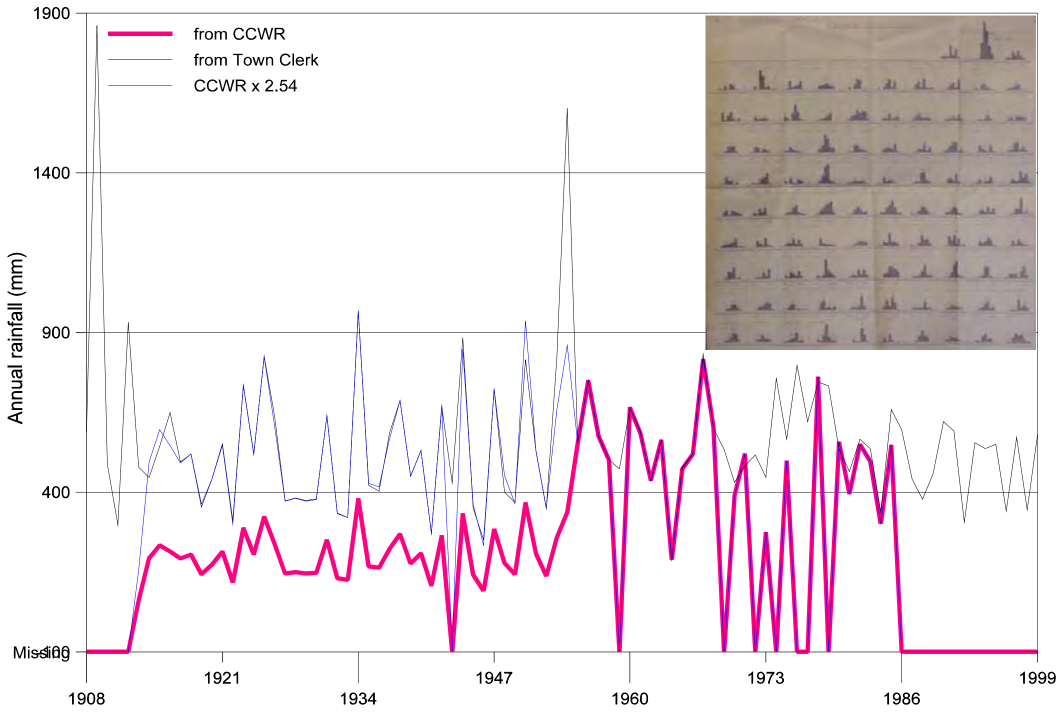


Figure 2 Imperial/metric rainfall amounts for the Tsumeb station

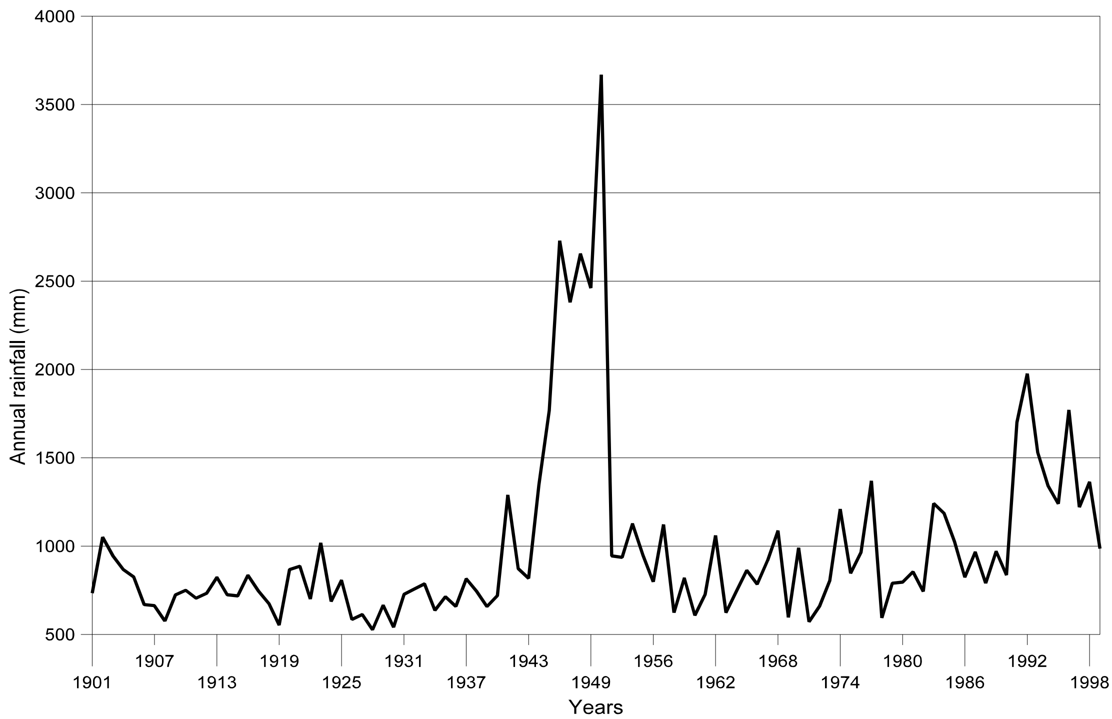


Figure 3 Suspect annual rainfall totals circa 1945 for 0022190 W

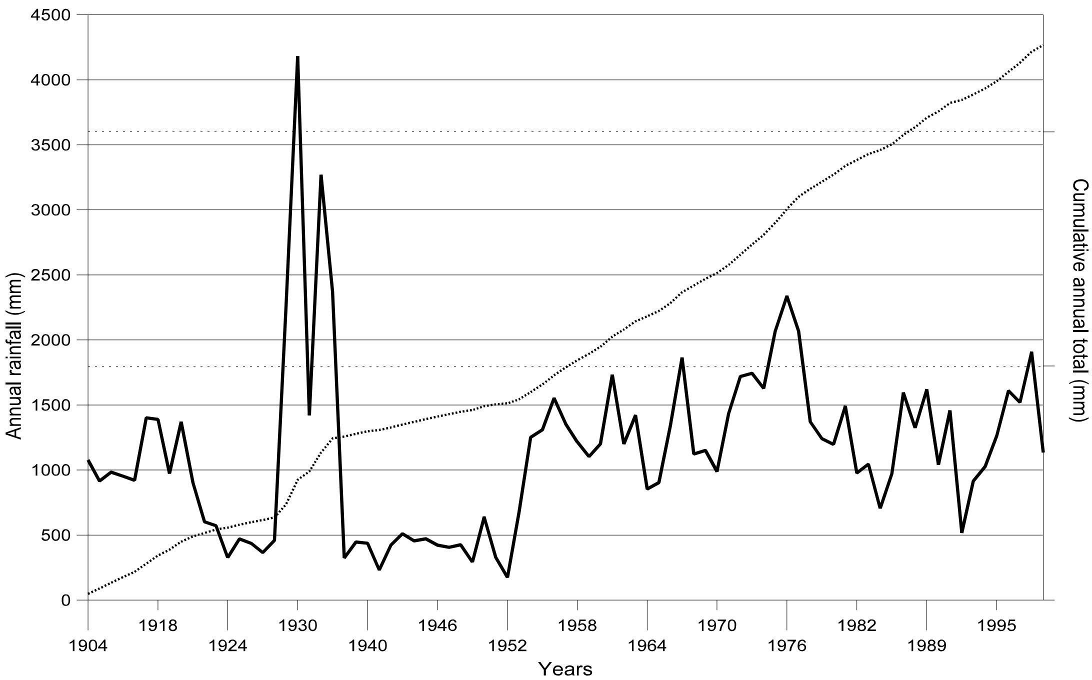


Figure 4 Suspect mean/median annual rainfall totals for 0431548 W

One of the stations identified by ***top10.f90*** as having the highest mean / median annual precipitation ratio is 0431548 W (PLESSISDAM) in the Vryburg area (Figure 4). This particular station has incorrect data around 1930 and most of the data prior to 1952 appears to have been recorded in imperial units.

A further analysis was performed on the annual rainfall totals to determine which rainfall stations have zero annual rainfall totals. There are stations in South Africa, however, that have legitimate zero annual totals and these were excluded from the analysis.

Approximately 80% of the daily rainfall database consists of zero rainfall amounts and amounts less than 5 mm per day account for 10% of the database (Table 3). It appears from the results in Table 3 that the infilling procedures (discussed later in the report) are generating more 0-5 mm amounts but retaining the less than 10 mm cumulative total of approximately 90% of the database.

One of the aims of this project was to collate and populate a daily and monthly rainfall database for southern Africa. Much time was spent detecting and correcting anomalies in the data, but owing to the size of the databases it is impossible to produce an error-free database. It is hoped that these two rainfall databases (daily and monthly) will find favour and that all future observed and infilled rainfall data be included to update it.

Table 3 Frequency analysis of the daily rainfall database

Rainfall (mm)	% of observed values	% of infilled values	% of observed and infilled values
zero	84.9093	77.4160	79.5348
0-5	6.7346	12.4999	10.7158
5-10	3.2739	5.9959	5.2536
10-25	3.5895	3.2651	3.4655
25-50	1.2073	0.6985	0.8559
50-100	0.2569	0.1129	0.1575
100-200	0.0261	0.0108	0.0155
200-300	0.0019	0.0007	0.0011
300-600	0.0005	0.0002	0.0003

2.5 PUBLIC APPEAL FOR RAINFALL DATA

Dent *et al*. (1989) made an appeal to the public to supply rainfall data that were not forwarded to any other rainfall gathering organisation and data for an additional 1 174 stations were received. If the station had less than 10 years of data, or if there were stations in close proximity and the addition of these public stations would not enhance the spatial coverage, then the data were not entered into the database.

At the start of this WRC project, early in 2000, an initial "*by word of mouth*" appeal was launched and this netted 445 recording stations with monthly accumulations of data that were supplied by the Institute of Commercial Forestry Research (ICFR) in Pietermaritzburg. Soon thereafter, Mr Neville Irwin made contact and offered his daily rainfall data of more than 70 years' records for an area south of Potchefstroom. These data include records from five other gauges and these six stations are within a few kilometres of each other. The analysis of these data culminated in a refereed journal article and was presented at a conference (Lynch *et al*., 2001a; Lynch *et al*., 2001b).

Data for an additional 20 stations were received during the course of this project. None of these stations were "gems" in the sense that they improved the density of gauges in topographic complex terrain.

2.6 DECLINE OF THE SOUTH AFRICAN RAIN GAUGE NETWORK

One of the first accounts, from Jan van Riebeeck's journal, of heavy rain in South Africa dates back to 22-23 July 1652 when the garden at the fort was washed away and the packing-shed in the fort was 150 mm under water (The Chief Director, 1990). The Royal Observatory (0020866 W) in Cape Town, however, is the earliest recording station with records dating back to 1850. The second active rainfall station, Ufumba (0375383 W), which is located approximately 20 km north of Hluhluwe in northern KwaZulu-Natal, has records that date back to 1865 and an active rainfall station opened in Clanwilliam, in the Western Cape, around 1869. By 1880 South Africa had more than 100 active daily recording stations and this number increased to a maximum of 3 841 in 1938, with a steady decline in the number of rainfall stations since then (Figure 5). The current daily rainfall database was augmented around 1960 with data from neighbouring countries, but also shows a decline in the number of active rainfall stations after roughly 1985 (Figure 5).

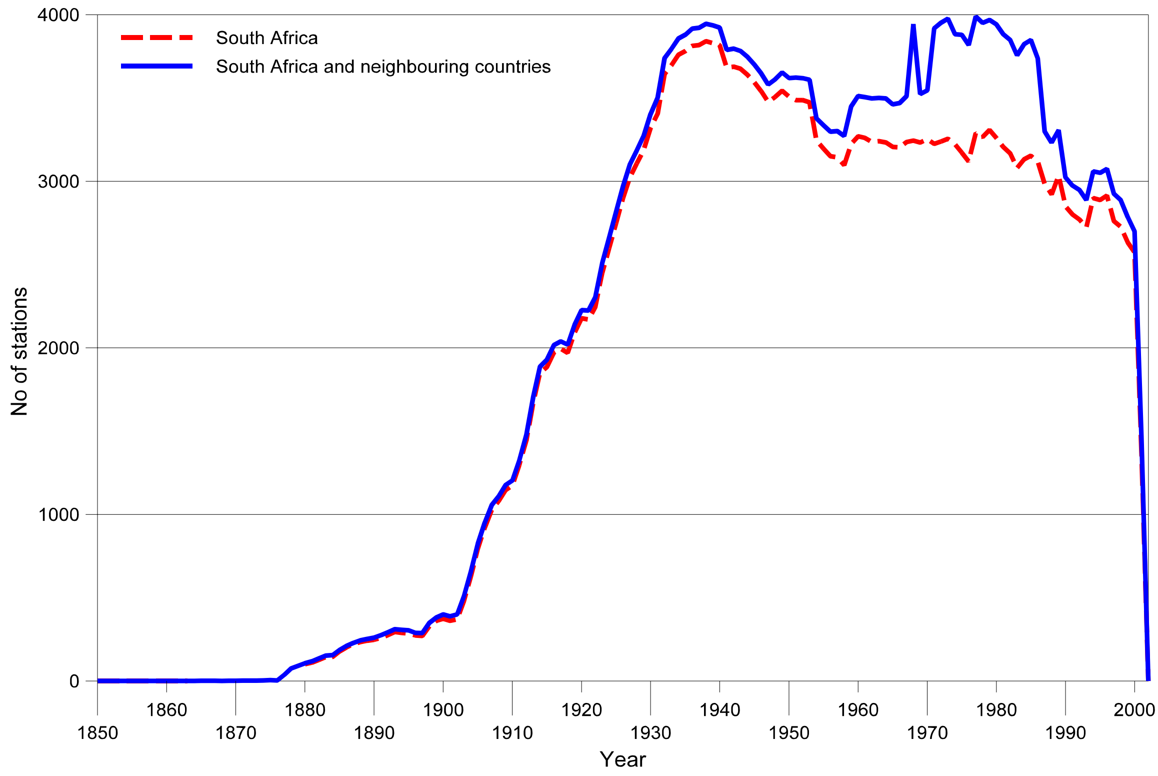


Figure 5 Number of active rainfall stations over time

Spatial distance (SD) is the spatial equivalent of the standard deviation statistic over an area and is defined as the square root of the sum of the standard deviations of all the latitude and the longitude co-ordinates (Burt *et al*., 1996). The SD of active rainfall stations in South Africa moves progressively in a northeasterly direction over time, a path akin to the movement of the early colonists in the last few centuries (Figure 6), with the SD stabilising at a point roughly in the centre of South Africa.

The question as to whether the decreasing number of active rain gauges is going to cause problems in the future, is often asked. The number of active rainfall stations covering South Africa started to decline in 1938, but this decline has become more pronounced after 1980 (Figure 7).

An analysis was performed on the daily rainfall database to determine the number of active rainfall stations per year per 15 arc minute square in an attempt to determine if the closure of the stations affected the spatial coverage of the rainfall monitoring network by removing spatial bias. Initially a test for at least one active station per 15 arc minute square was performed (Figure 7) and the trend is similar to the total number of active stations. The next

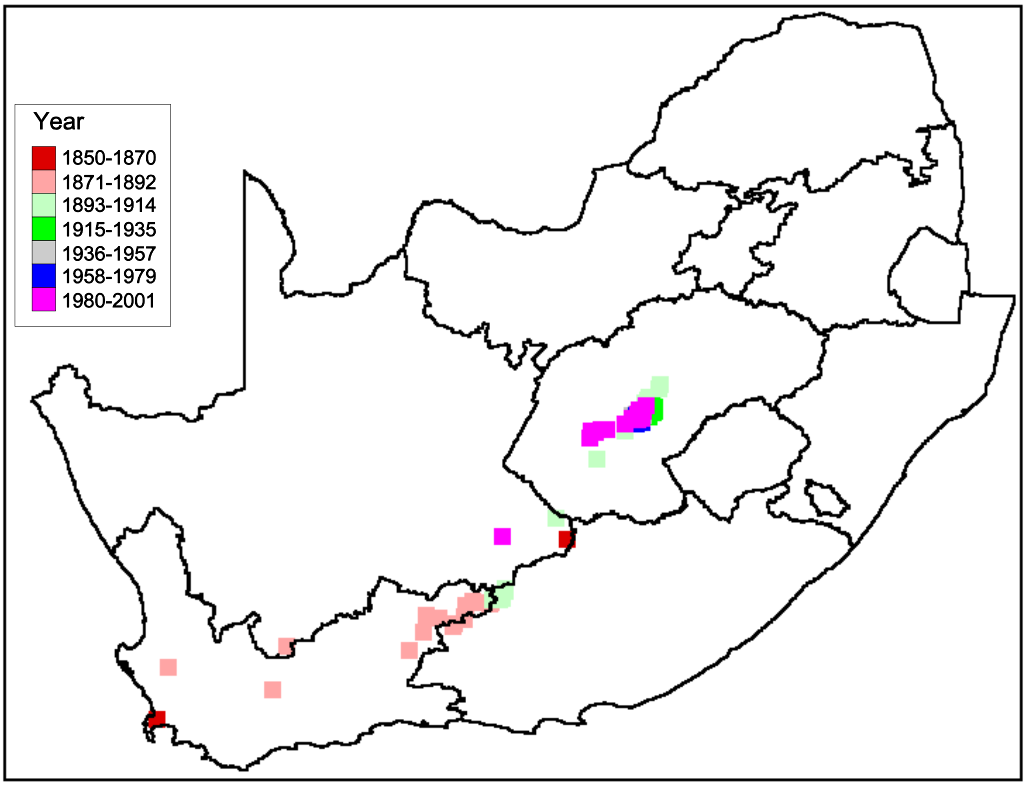


Figure 6 Spatial distance of active number of rainfall stations over time

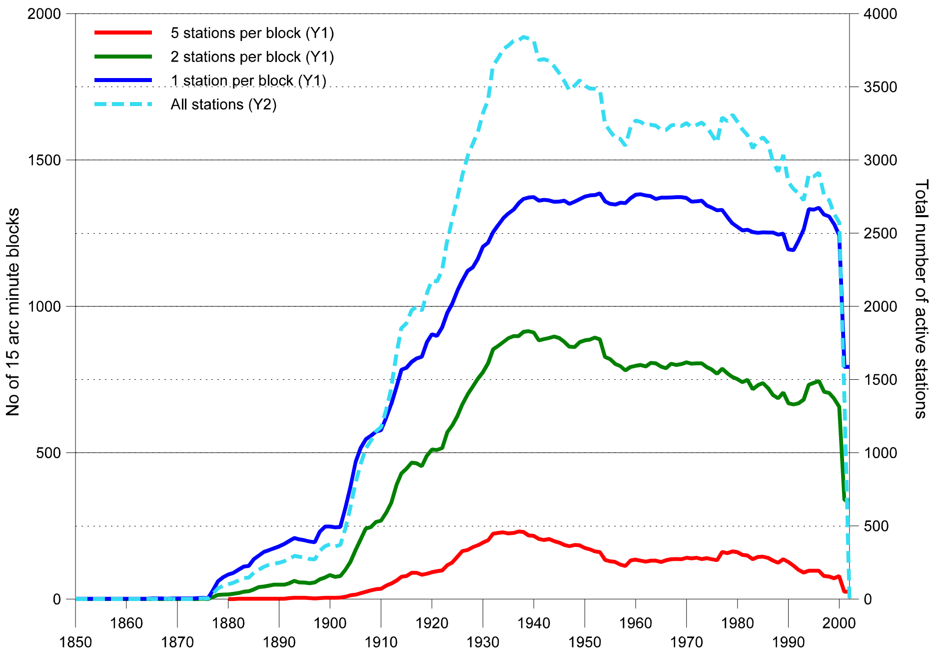


Figure 7 Analysis of active stations with spatial bias removed

test, using at least 2 and then at least 5 stations per 15 arc minute square (Figure 7), all have similar trends, which highlights the declining trend in the spatial coverage of the rainfall monitoring network.

In conclusion, the number of active rainfall stations increased from 1 in 1850 to 3 841 in 1938. Thereafter, however, a sharp decrease in the number is noted and this decline tapers off for the period 1959 to 1981, whereafter a sharp decline is noticed once again (Figure 7). It would appear that the decline in the number of stations, however, does not affect the spatial coverage, as badly as it could have, as stations exist in proximity of those that were closed down.

2.7 TECHNIQUES FOR SELECTING REPRESENTATIVE RAINFALL STATIONS

The inclusion of a key rainfall station with a short record or the exclusion of a station with a long record often relies on the intuition of the researcher. A series of rainfall stations data that have been found, using various techniques, to be suspect are excluded from the analysis. There are approximately 12 stations that must be included as they have been checked manually and have been found suitable, but the checking routines always flag them as suspect. A further analysis ***(extract.f90)*** reveals that an additional 473 rainfall stations need to be excluded from the analysis based on suspect infilled values and an observed record that is too short.

The current monthly rainfall database comprising of annual totals determined using observed and infilled values is used. The infilled data are excluded for a station if these data cause the absolute difference between the mean of observed data and the mean of the infilled data to be more than 20%. The data are also excluded if the station has more than 15 years of record, but the observed record is less than 5 years in length.

The analysis found that approximately 320 stations, of which the observed and infilled means are different, have sufficient observed data and therefore the infilled values are ignored. Unfortunately there are approximately 470 rainfall stations that are excluded as they have insufficient observed data and the observed and infilled means are significantly different from one another.

The monthly rainfall database is initially searched (***extract.f90)*** and then the stations with suspect data are eliminated (***eliminate.f90)***. The spatial and explanatory variable information is attached to each station (***pickup.f90)*** and these data are then combined into one single data file ***(for_jcs.f90).*** The duplicate stations, where more than one station represents a one arc minute pixel, are excluded based on a length of record criteria ***(dup_stns.f90)***. The four output files created ***(dup_stns.f90)*** can now be used as input for the regression and interpolation analysis or can be used to create a point coverage of the selected rainfall stations (Figure 8).

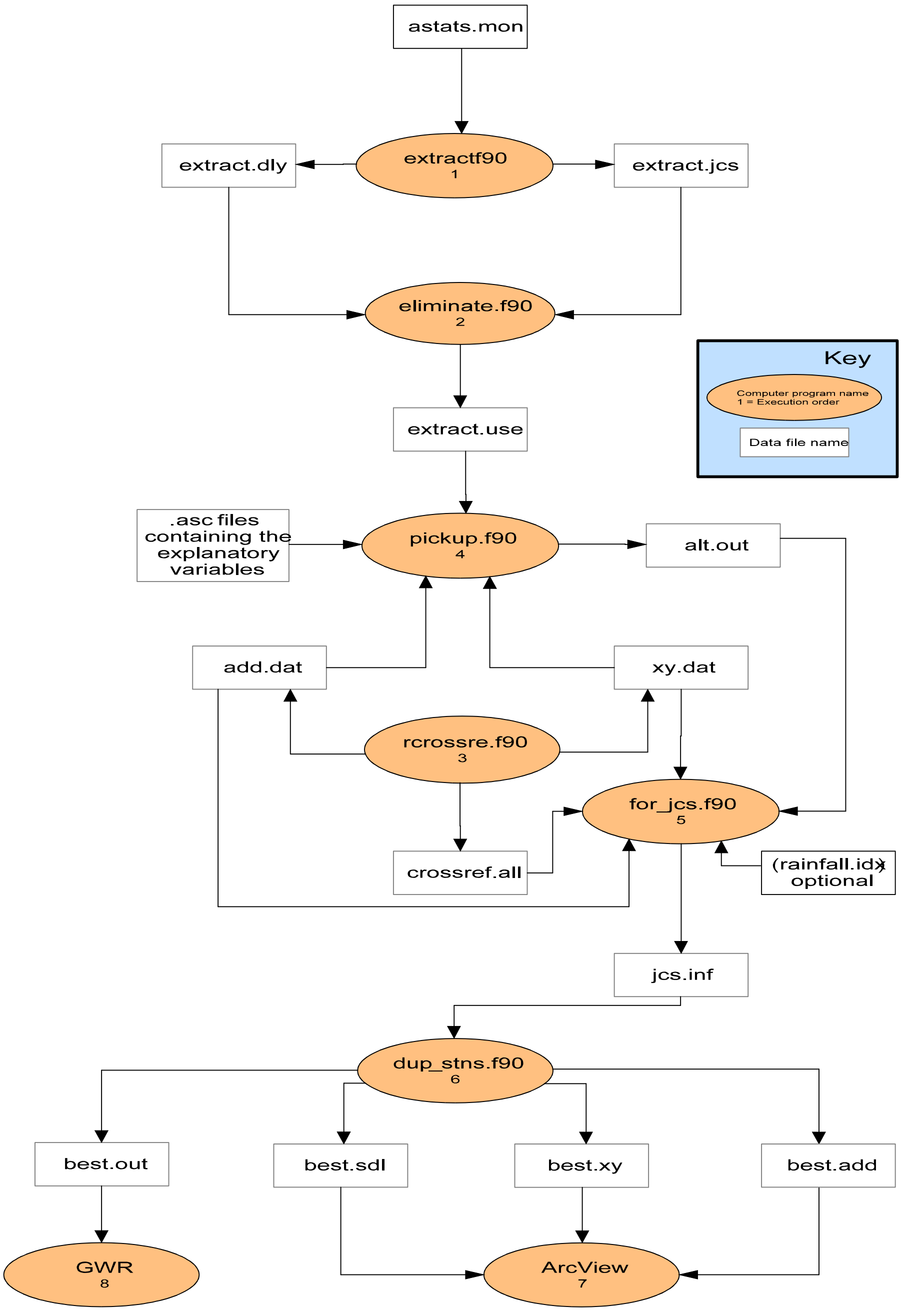


Figure 8 Process outline for selection of rainfall stations to be used in the regression procedure

CHAPTER 3

TECHNIQUES USED FOR THE INFILLING OF RAINFALL DATASETS

Missing records limit the use of rainfall data because daily simulation models, for example, cannot function without a continuous input dataset of rainfall. The length of rainfall record that is required to produce a representative MAP, i.e. so long that MAP is within 10% of the long term mean 90% of the time (Lynch and Dent, 1990), will be achieved by many more stations if rainfall data are infilled. A suite of different infilling techniques was used and the decision as to which one to use was based on a hierarchical process that will be explained later in the document. It must be stressed that all the infilled rainfall values are flagged according to the technique that was used to estimate them and users of the database can ignore values infilled by certain techniques if they choose to do so. *The crux of infilling is, that a missing day implies an incomplete month which implies an incomplete year.* Hence the importance of infilling.

3.1 INVERSE DISTANCE WEIGHTING

The procedure of Inverse Distance Weighting (IDW) revolves around the concept that implements the assumption that things that are in close proximity to another are more alike than those that are further apart. The IDW interpolation technique used by Meier (1997) weights the recorded rainfall from stations surrounding a target station inversely, depending on the distance of those stations from the target station. Because of the spatial variability of rainfall in southern Africa, a quadrant approach was introduced to the selection of the stations to be used. Meier (1997) developed a procedure whereby a certain number of stations were selected from each of the four quadrants surrounding the target station in order to minimise the bias that would be introduced if the closest few stations were all in the same direction from the target station.

3.2 EXPECTATION MAXIMISATION ALGORITHM

The Expectation Maximisation Algorithm (EMA), formalised by Dempster *et al*. (1977), was adopted by Makhuvha *et al*. (1997a; 1997b) to infill missing data in monthly rainfall records. The EMA recursively substitutes missing data and then re-estimates the multiple linear regression relationship between the data at the target station and the data from the nearby control stations. The EMA technique requires that the selection of suitable control stations and these stations are interrogated to determine the suitability of using the selected target and control stations for the simultaneous infilling of missing data (Smithers and Schulze, 2000).

3.3 MEDIAN RATIO METHOD

The median monthly values for the target and nearest control station are computed and the ratio is used to adjust the data from the control station and the result is filled in for the missing day at the target station. If the nearest control station also has missing data on the day, the next closest station is used and so on (Smithers, 2002).

3.4 MONTHLY INFILLING TECHNIQUE

Dent *et al*. (1989) filled in the missing monthly totals using a regression technique (using surrounding stations) described in Zucchini *et al*. (1984). This observed and infilled monthly rainfall database (Dent *et al*., 1989) was interrogated and the monthly infilled values of zero monthly rainfall were extracted and also all the months where the infilled monthly total is less than or equal to 2 mm. The assumption that an infilled monthly total of zero rainfall implies that the data for that month can be assumed to equal zero appears valid. Where the infilled monthly total is less than 2 mm, the first day of the month is assigned this monthly total and the rainfall amounts for the rest of the month are set to zero also seems to be a valid initial assumption.

3.5 THE INFILLED RAINFALL DATA IN THE DATABASE

The EMA technique (Smithers and Schulze, 2000) and the ratio approach are considered to be the premier infilling techniques available to the research project and the missing daily rainfall values are then replaced with these EMA estimates. The rainfall estimates generated by the IDW approach (Meier, 1997) are then used to infill the remaining missing values. Finally, the zero and less than 2 mm rainfall values derived from the Dent *et al*. (1989) database are then used to infill any remaining missing values that have not been infilled.

The infilling techniques used have more than doubled the size of the daily rainfall database (Table 4). The rainfall database consists of 105 753 218 daily observed values with 236 154 934 infilled values giving a total size of the observed and infilled rainfall database of 341 908 152.

Table 4 Number of infilled daily rainfall values in the rainfall database

Infilling technique	Number of daily values
EMA	113 869 517
Ratio	40 823 148
IDW	81 451 381
0-2 mm	10 888
Total	236 154 934

The infilling process has also increased the size of the annual database considerably from an initial 5 118 stations with more than 15 years of complete record to 9 641 stations that have more than 15 years of record.

CHAPTER 4

INTERPOLATION / REGRESSION TECHNIQUES FOR ESTIMATING RAINFALL AT UNGAUGED LOCATIONS

It is a fundamental geographical principle that, generally speaking, things that are closer together tend to be more alike than things that are further apart (Tobler, 1970). One of the primary objectives of this research is to determine rainfall values at ungauged positions using the best suited interpolation or regression procedure or combinations thereof.

4.1 INTERPOLATION

The value at an ungauged site which lies between gauged sites can be interpolated only by fitting some plausible model of variation to the values at the gauged sites and then calculating the value at the desired ungauged site (Burrough, 1986).

Some methods, such as weighted averages, make the surface pass through all the original data points, while others, like splines, place more emphasis on the neighbourhood fit. Kriging on the other hand, produces an "optimal" interpolation by using a statistical model of spatial interdependence (Burrough, 1986). Kriging is optimal if its assumptions of a spatially uniform error structure apply (Chrisman, 1997).

Deterministic interpolation methods, *viz*. inverse distance weighting, are so termed because they are based directly on the surrounding measured values or on a specified mathematical formula that determines the smoothness of the resulting surface. A second family of interpolation methods consists of geostatistical methods, *viz*. kriging, which are based on statistical models that include autocorrelation (statistical relationships among the measured points). Not only do these techniques have the capability of producing a prediction surface, but they can also provide some measure of the certainty or accuracy of the predictions (Johnston *et al*., 2001).

4.1.1 Inverse Distance Weighting

Inverse distance weighting (IDW) is a deterministic interpolation technique. As the name implies, the weight of a value decreases as the distance increases from the prediction location (Johnston *et al*., 2001). The surface resulting from inverse distance weighting methods depend on the function, or on the parameters of the function used, and on the size of the window from which the sample data points are drawn. The size of the window not only affects the average values estimated at a point, but also controls the amount of computer time required for the interpolation.

The power parameter influences the weighting of the measured location's value on the prediction location's value. As the distance increases between the observed locations and the prediction location, the weight that the observed point will have on the prediction will decrease exponentially.

Because moving average methods are, by definition, smoothing techniques, maxima and minima in the interpolated surface can only occur at data points (Burrough, 1986).

4.1.2 Kriging

Kriging is a geostatistical technique for interpolation that uses information about the spatial autocorrelation in the vicinity of each point to provide "optimal" interpolation (Chrisman, 1997). Kriging is a spatial interpolation technique that makes the most convincing claim to be grounded in sound theoretical principles. The basic idea is to "discover" something about the general properties of the surface, as revealed by the measured values, and then apply these properties in estimating the missing parts of the surface. Smoothness is the most important property, and is operationalised in kriging in a statistically meaningful way. It is often argued that the surfaces produced by kriging are too smooth, but Pegram (2002) suggests that it is the "best" possible surface that one could determine from the given set of points and that the less-smooth surfaces produced by other techniques are artifacts of interpolation.

The form of the semivariogram is central to kriging, but is not always easy to ascertain whether a particular estimate of the semivariogram is, in fact, a true estimator of the spatial co-variation in the area. An added advantage of kriging is that the method yields estimates of the errors associated with each interpolation (Burrough, 1986).

Cokriging is a statistical interpolation method that uses data from multiple data types (multiple attributes) to predict (interpolate) values of the primary data type. When interpolating rainfall values (primary data) using, for example, altitude values as an explanatory variable and the predicted surface is created, then the altitude values are also interpolated. Cokriging therefore interpolates all the explanatory variables at the sites where the rainfall values are to be estimated. Rain gauges are often not positioned on mountain peaks and when using altitude as an explanatory variable the cokriging method must then interpolate the altitude value whereas regression methods makes use of the DEM.

Kriging is a powerful and effective tool, but it does not address all the rainfall interpolation problems. Observed rainfall generally shows strong anisotropy in space and in particular relative to the elevation above sea level, where the behaviour is non-linear (Lanza *et al*., 2001).

4.1.3 Thiessen Polygons

The simplest method of interpolation is to use external features to delineate common areas. The concept of Thiessen polygons revolves around the idea that for a given set of two dimensional sampling points, the best information about an unvisited point can be gleaned from the data point nearest to it. Thiessen polygons are used to divide an area up in a way that is totally determined by the configuration of the data points. If the data lie on a rectangular grid, then all the Thiessen polygons are equal, regular squares and if the data are irregularly spaced, then an irregular lattice (Figure 9) of regions in the form of regular polygons results (Burrough, 1986).

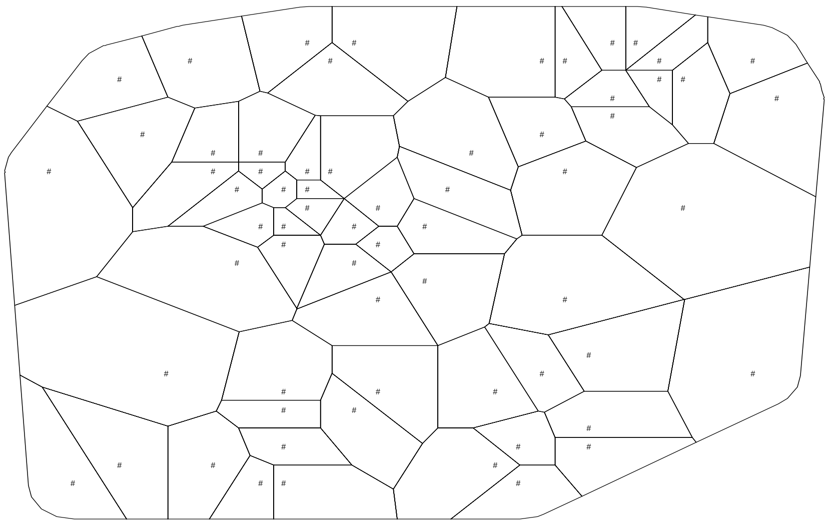


Figure 9 An example of Thiessen polygons

For each rain gauge there is a Thiessen polygon composed of the adjacent rain gauges lying closer to that gauge than to any other.

4.2 REGRESSION

Seldom in the social or physical sciences is it possible to satisfactorily explain the variations in a dependent variable by using a single independent variable. Multiple regression analysis allows for several independent variables to be used to account for the variability of a single dependent variable (Burt *et al*., 1996). In a formal statistical sense, regression analysis allows one to identify the dependence of one variable upon one or more independent variables (Longley *et al*., 2001).

Spatial autocorrelation measures refer to the inter-relatedness of phenomena across space, one attribute at a time. Another important aspect of the nature of geographic data is the tendency for relationships to exist between different phenomena at the same place (Longley *et al*., 2001). This implies that there might exist, for example, more correlation in the product of two explanatory variables than when using them individually.

The statistical methodology that allows one to make inferences assumes that there is no autocorrelation between errors across space or time. This assumption clearly does not conform with Tobler's Law, where everything is related to everything else, but things close to one another are more related than those further apart is manifest in positive autocorrelation (Longley *et al*., 2001).

4.2.1 Multiple Linear Regression

It is possible to extend regression analysis to situations where several independent variables are used to account for the variability of a single response variable and this is termed multiple regression (Burt *et al*., 1996). In other words, multiple linear regression fits a response variable as a linear combination of multiple independent variables by the method of least squares.

Dent *et al*. (1989) fitted a multiple regression model to each of 34 rectangular areas, with a common overlap region, to estimate values of MAP across southern Africa. Following their approach to multiple regression the following transformations of the independent variables were considered: squares, square roots, reciprocals and logarithms. Furthermore, the cross-product of every combination of pairs of variables were calculated and considered, giving a total of 175 variables. For each data set a subset of possible models was first chosen by using forward, backward and stepwise regression, based on R^2 values and then a final model was chosen based on the PRESS (predicted error sum of squares) statistic. The latter was used as the most appropriate model for mapping MAP. It may not be the model that fits the data best, but rather the model that gives the best predictions (Dent *et al*.,1989).

4.2.2 Variable Window Size Multiple Regression

It is a well known fact that the distribution of MAP is dependent on factors which include altitude, distance from the sea and others. One can therefore not simply use an interpolation technique to determine the MAP value at an ungauged position. The regression procedure outlines an initial investigation of a technique to produce a raster of MAP values using a variable window size for the selection of variables to be used in a multiple regression approach.

There are more than 13 000 stations (Figure 1) at which rainfall has been measured over southern Africa which includes South Africa, Lesotho, Swaziland, Namibia, Botswana, Zimbabwe and a small part of Moçambique. Approximately 9 500 of these stations have more than 15 years of observed and infilled data. A procedure was developed to determine what size a square should be such that each station and its 49 closest neighbours are within that square. The value of 50 stations was decided upon after some preliminary investigations. This process is repeated for each of the roughly 9 500 rainfall stations. A stepwise multiple regression is then performed on each of these squares using MAP as the dependent variable and latitude, longitude, altitude, distance from the sea, surface roughness and a combination of these as independent variables. A multiple regression equation is generated for each independent variable combination and the combination with the best adjusted R^2 (R_A^2) is then selected as the model to represent that particular square. This procedure is repeated for each square surrounding each of the roughly 9 500 points. As each square overlaps each other square a test is performed to determine for each pixel what the best R_A^2 value should be by considering the R_A^2 values of each square that overlaps the pixel in question.

A suite of computer programs was developed to determine, for each pixel, what regression equation should be used to estimate the MAP at the pixel. Each square has its own regression equation and an overlay process is used to determine which regression equation is best suited to a particular pixel.

This technique was not pursued due to time constraints.

4.2.3 Geographically Weighted Regression

The mean annual precipitation (MAP) raster of Dent *et al*. (1989) was derived at a spatial resolution of one arc minute using multiple linear regression. South Africa was divided into 34 rectangular regions and a regression equation was manually selected for each region. Hughes *et al*. (2001) and Lynch (2001a) investigated more recent interpolation and regression techniques, *viz.* kriging and geographically weighted regression (GWR, Brunsdon *et al*., 1996), to establish if they would be more suitable for the mapping of MAP than multiple linear regression as used in the 1989 project. Multiple regression gives very detailed maps, but require a large amount of time to implement and may result in physically meaningless values on maps. Kriging is very good at achieving an overall representation, in a short time, but one which lacks fine detail. GWR is far quicker than multiple regression to generate a raster of MAP for the whole region and GWR also requires fewer explanatory variable than multiple regression.

Spatial non-stationarity occurs when the relationships between variables and their underlying processes change over space. GWR is an attempt to account for this spatial non-stationarity, within the framework of traditional regression, by allowing the coefficients to vary locally. Thus the usual model:

$y_i = \sum_{j=0}^{n} x_{ij} \beta_j + \varepsilon_i$ where y_i is the observation measured with explanatory variables x_{ij} and ϵ_i are independent random variables is extended to:

$y_i = \sum_{j=0}^{n} x_{ij} \beta_j (u_i, v_i) + \varepsilon_i$ where (u_i, v_i) is the location in space of the i-th observation.

The GWR procedure is developed to run in the R environment, which is very similar to the statistical programming environment S and S+. The major difference is the cost, because R is free. The GWR technique will be a great asset in future when new surfaces of MAP are to be generated. It uses far fewer independent variables than the classic multiple regression procedures and there is no cost involved in acquiring the software. Furthermore, the software is virtually platform independent. It can be found at www.ncl.ac.uk/geography/GWR.

GWR therefore allows actual parameters for each location in space to be estimated and mapped as opposed to having a trend surface fitted to them (Brunsdon *et al*., 1996). This means that maps of the parameter estimates, $\beta_j (u_i,v_i)$ can be plotted, showing their variation over space and thereby improving the understanding of the process being modelled (Brunsdon *et al*., 1999).

The parameter estimates at any regression point are dependent not only on the data supplied but also on the kernel chosen and the bandwidth selected for that kernel.

The bandwidth in GWR determines the rate at which the regression weights decay around a given point (u_i,v_i). If the bandwidth is small, weights decay quickly with distance and the values of the regression coefficients change rapidly over space. The present GWR approach allows for the kernel type to be either fixed (Gaussian) or adaptive (bi-square).

The kernel bandwidth (Figure 10) is determined either by cross-validation or Akaike Information Criteria (AIC) minimisation (Fotheringham *et al*., 2002). It is possible to think of the bandwidth as a 'smoothing' parameter, with larger bandwidths causing greater smoothing. An over-smoothed model will produce parameters that are similar in value across the study area and an under-smoothed model will produce parameters with so much local variation that it would be difficult to determine whether there are patterns at all (Brunsdon *et al*., 1999).

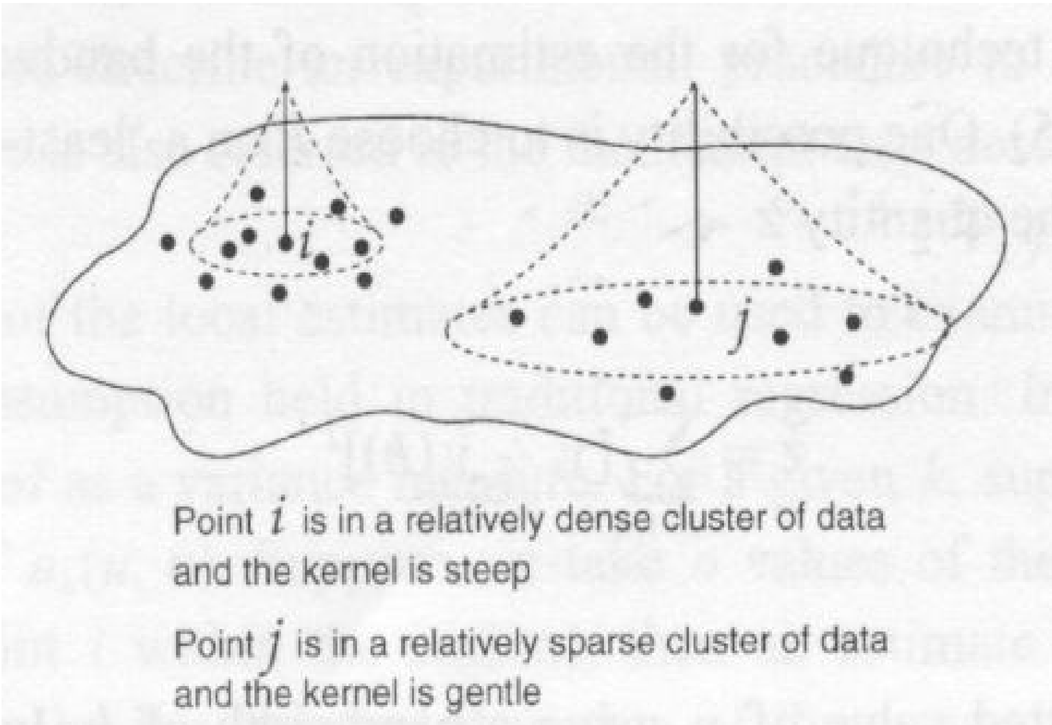


Figure 10 Example of an adaptive spatial kernel in GWR (after Fotheringham *et al*., 2000)

As in all statistical models one of the most important questions is how to select a "best", or most appropriate, model. In GWR this amounts to choosing the bandwidth as well as the explanatory variables. For the choice of explanatory variables there is no automated procedure, such as the forward or backward selection procedures for multiple regression, and the choice of such variables must be guided by a good understanding of the underlying processes. For choosing a bandwidth for the model there are two possible methods, *viz*. cross validation and Akaike's Information Criteria (AIC). Brunsdon (2002) favours AIC because the bandwidth selection via the cross validation technique tends to under-smooth the data. Hutchinson (1998) found that the cross validation technique does not always represent a reliable estimate of model error, especially when short range correlation in data are present.

GWR recognizes that spatial variations in relationships might exist and provides a way in which they can be measured (Fotheringham *et al*., 2000). If the bandwidth tends to infinity, the estimated parameters become uniform and GWR becomes equivalent to ordinary least squares. Conversely, as the bandwidth decreases, the parameter estimates will increasingly depend on observations in close proximity to the observation point and hence will have increased variance (Fotheringham *et al*., 2000).

4.3 EXPLANATORY VARIABLES

The analysis of data can be thought of as part art and part science (Burt *et al*., 1996). This is the case when deciding on which variables can and should be used as independent variables to describe rainfall patterns. The single explanatory variable, such as distance from the sea, might be replaced by a combination of other explanatory variables. The influence of altitude and other terrain indices on the spatial description of rainfall, and in particular MAP, are well documented in the literature and a comprehensive discussion can be found in Dent *et al*. (1989).

A regression model with numerous explanatory variables may be difficult to maintain and a regression model with fewer explanatory variables is easier to work with and to understand. The presence of many highly intercorrelated explanatory variables may substantially increase the sampling variation of the regression coefficients, detract from the model's descriptive abilities, increase the problem of roundoff errors, and not improve, or even worsen, the model's predictive ability. Computerised approaches can be very helpful in identifying appropriate sets of explanatory variables, but in order to develop a useful regression model one needs to be pragmatic and needs to utilise large doses of subjective judgment (Neter *et al*., 1996).

4.3.1 Digital Elevation Model

The one arc minute DEM (also referred to as the minute by minute altitude grid, DEM16 or 1600 m DEM) was captured and checked for the previous rainfall mapping exercise (Dent *et al*., 1989) and was corrected a few years later when more up-to-date 1:50 000 topographical maps became available (School of Bioresources Engineering and Environmental Hydrology). The 200/400 m DEM (Directorate Surveys and Mapping, 1990) and the 800 m DEM (USGS, 1996) are used to correct DEM16 once again. Some of the procedures that were used include:

- A DEM for Lesotho was generated from scanned contour information supplied free of charge by the erstwhile Enviromap (2000). This DEM was checked for sinks and edge matched to the existing 200 m DEM.
- The Lesotho patch was inserted into the DEM, and areas surrounding the patch were error checked and filled/adjusted.
- The DEM was checked for edge effects, coastal errors and errors with neighbouring states. Holes and errors in the DEM were filled using a manual nearest neighbour approach, with special attention being given to the creation of erroneous sinks. Where contour data were available, these were used in conjunction with the 800 m DEM to assist this filling process.
- Computer programs were deployed to identify areas where the 1600 m DEM differed from the 200 m and 800 m DEMs. The points of difference were identified, and then all three DEMs investigated at those points. Based on values from the other DEMs,

surrounding values, available contours and inspection, the values in the 1600 m DEM were adjusted or left depending on the findings. If the 200 m and 800 m DEMs (aggregated to 1600m) were found to be erroneous then it was adjusted to avoid revisiting the same errors. The original 200 m DEM, however, was NOT changed except for the inclusion of the Lesotho data.

A DEM (Figure 11) now exists at 1 600 m, approximately one arc minute, resolution which is different from the aggregated 200 m DEM by 100 m or less. In turn the DEM16 is less that 150 m different from the aggregated 800 m DEM supplied by the USGS. Many of the approximately 11 390 errors found, Figure 12, were in the Lesotho region, the Drakensberg, the Cape fold mountains, and all areas of variable terrain (Figure 11). Surprising though was that in comparison to the available contours, the 1 600 m DEM often offered a more realistic representation of these areas than either of the aggregated rasters.

A more detailed account of the procedures involved in identifying suspect pixels can be found in Lynch (2002).

The area of interest decided upon for this research project includes part of Namibia, Botswana, Zimbabwe and Moçambique and the 800 m, roughly 30 arc seconds in South Africa, DEM (USGS, 1996) was re-sampled to a one arc minute pixel size and combined with the one arc minute South African DEM.

4.3.2 Surface Roughness Indices

There are a host of DEM derived indices (Figure 13) that one can calculate to assist in the spatial description of rainfall data. A surface roughness index, ialtCV, is described as the coefficient of variation (CV) of a 5 arc minute mask of altitude values (Figure 13). Areas that have similar altitude values have a low CV and areas that are topographically complex have high CV values and these areas also often have highly variable spatial rainfall patterns. A slope raster (Figure 13) and an aspect raster (Figure 13) is generated using the one arc minute DEM. A curvature raster (Figure 13), which is defined as the curvature of the DEM at each pixel centre, is also derived from the DEM. A positive curvature indicates that the surface is upwardly convex at that pixel and a negative curvature indicates that the surface is upwardly concave at that pixel.

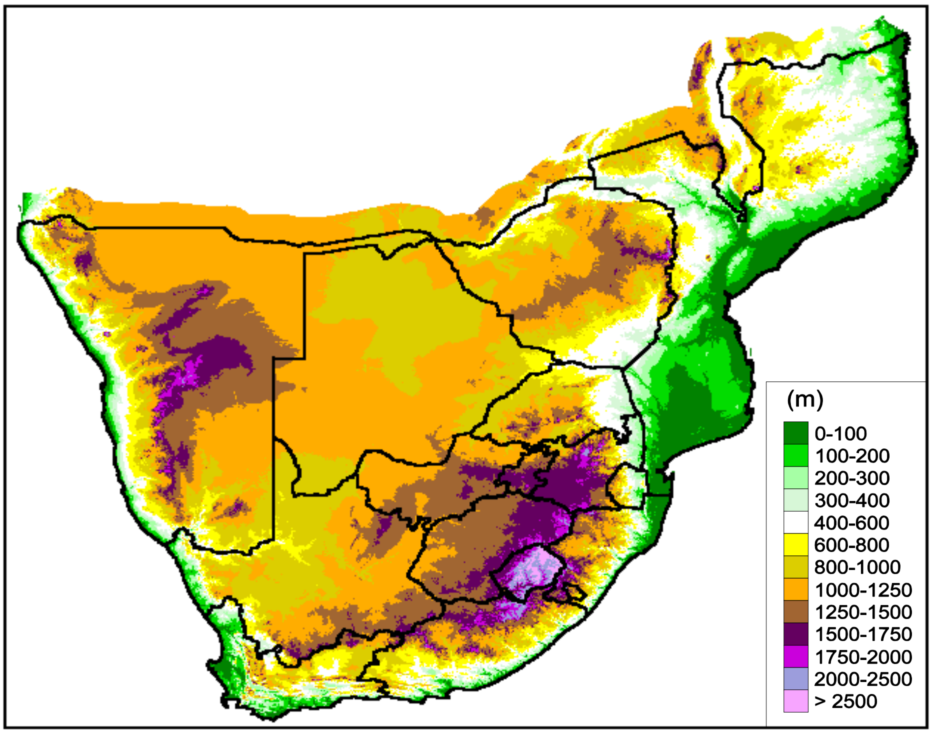


Figure 11 One arc minute DEM of South Africa and its neighbours

Figure 12 Locations of the 11 390 corrected altitude pixels

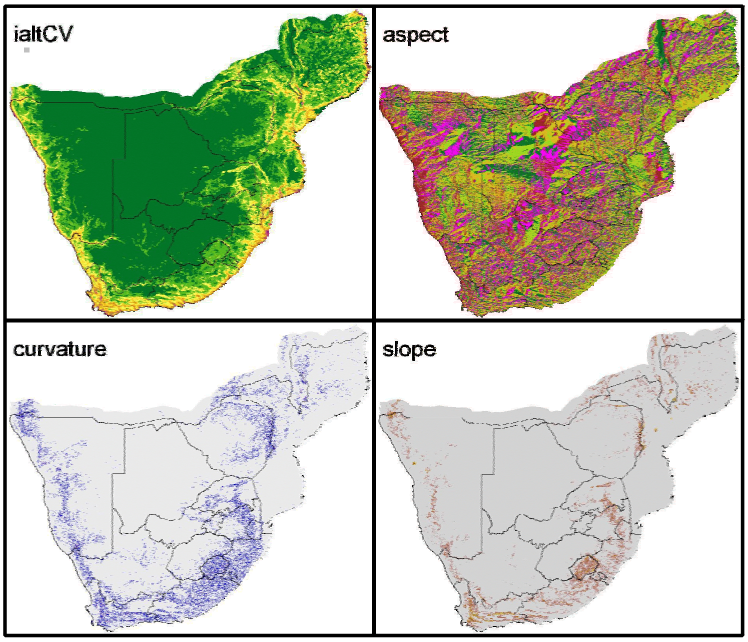


Figure 13 Pictorial representation of the surface roughness indices

The question may be posed as to whether these surface roughness rasters are correlated in any way, as they are all derived from the one arc minute DEM. A technique was developed whereby the rasters are compared to one another on a corresponding pixel basis. All the rasters have been normalised, pixel value minus the raster minimum all divided by the range, to have values between 0 and 1. Each pixel in the one raster is compared to the corresponding pixel in the other raster and if the values are within 20% of one another then those two pixels are deemed to be the same, and this process continues for all the roughly 1.4 million pixels in each raster. Only two pairs of surface roughness raster appear to have some commonality, *viz*. the DEM and the curvature raster have 19.28% of their pixels within 20% of one another spatially, and the aspect raster and the curvature raster have 20.77% of their pixels within 20% of one another spatially (Table 5). These similarities reported in Table 5 suggest that the surface roughness rasters can be used as they are not correlated to one another.

Table 5 Percentage similarity between the normalised surface roughness rasters

	DEM	ialtCV	slope	aspect	curvature
DEM		2.13	3.53	10.95	19.28
ialtCV			6.36	1.75	0.11
slope				2.73	0.56
aspect					20.77

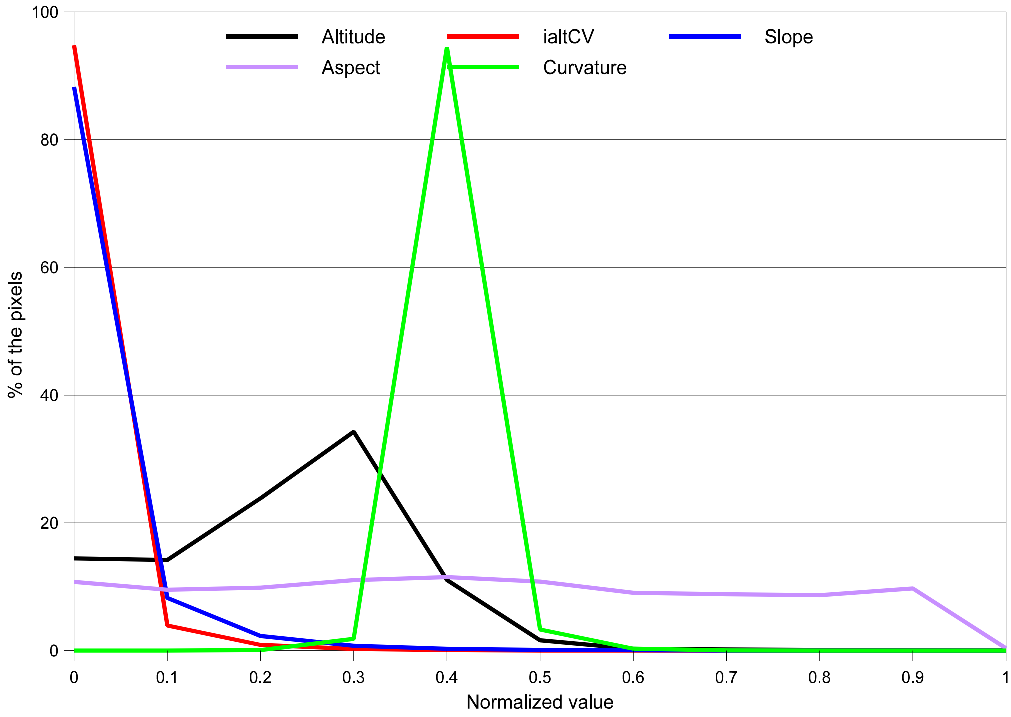


Figure 14 Frequency distribution of the surface roughness indices

It is also interesting to note that the frequency distributions of the majority of these surface roughness rasters have different shapes (Figure 14). The ialtCV and the slope rasters, however, have a similar frequency distribution (Figure 14) but only 6.36% of their pixels have values that are within 20% of each other (Table 5).

It is well documented that there exists a strong statistical relationship between MAP and altitude. This, however, does not mean that the MAP is similar for the same altitude value. The one altitude value might be for a coastal mountain peak and the other altitude value may be far inland on a plateau. The correlation coefficient between altitude and MAP is only -0.033 (Table 6) whereas the correlation coefficient between MAP and ialtCV is 0.264 when analysing the complete database of rainfall stations. The correlation coefficients (Table 6) are spatially non-stationary over southern Africa. GWR is a technique that attempts to handle this problem of non-stationarity based on the idea of a local statistic.

4.4 CHOICE OF INTERPOLATION / REGRESSION TECHNIQUE

The conclusion to be drawn from these studies is that it is unwise to use ones data with the first available interpolation technique without carefully considering how the results will be affected by the assumptions inherent in the method (Burrough, 1986).

Inverse distance weighting techniques are the method of choice when converting point data to a raster format as they are easy to use and are not computer intensive. When the density of the point data is sufficient and the variation in the parameter to be converted

Table 6 Selected explanatory variables vs MAP correlation coefficients

Explanatory Variable	Correlation with MAP
$Longitude^2$	0.645
Latitude*Longitude	-0.547
Latitude+Longitude	0.470
Distance from the $sea^{1/2}$	-0.321
Slope	0.299
ialtCV	0.264

to raster is not too complex, then IDW would suffice. Rainfall, in areas of complex topography and where the distribution of the stations is sparse in these areas, is better converted to raster using an approach that relies on additional explanatory variables.

The area surrounding the Jonkershoek mountains near Stellenbosch in the Western Cape Province lends itself to a pictorial illustration of the main differences between a regression approach and an IDW approach. The MAP values increase from 1 085 mm to 3 199 mm over a distance of roughly 8.5 km and the altitude increases from approximately 230 m to 1 300 m (Figure 15).

The area surrounding Jonkershoek has rainfall stations at varying altitudes (Figure 16) and the areas towards Paarl and Franschoek also have a reasonable rainfall / altitudinal variation. The area further northeast, however, is covered by only two rainfall stations. Two rainfall surfaces, created using IDW and multiple regression have been draped onto a 3-dimension DEM to illustrate the differences**.**

When using an IDW approach the MAP increase with altitude around Jonkershoek is evident, but the mountainous area further northeast has a similar MAP value as the surrounding area (Figure 17). The regression approach, on the other hand, has varying MAP values in the ungauged mountainous area northeast of Jonkershoek. This example highlights the fact that IDW relies on a dense gauge distribution in complex areas, whereas regression can utilise a relationship that was built up some distance away to estimate the MAP in an area of complex topography that also has a spare rain gauge network.

It is, therefore, evident that techniques used to estimate the MAP raster from the rainfall stations should allow for the inclusion of explanatory variables. This implies that of the techniques discussed only cokriging and regression remain. The limitations and shortcomings of the cokriging approach therefore suggest the acceptance of the GWR approach as suggested by Hughes *et al*. (2001).

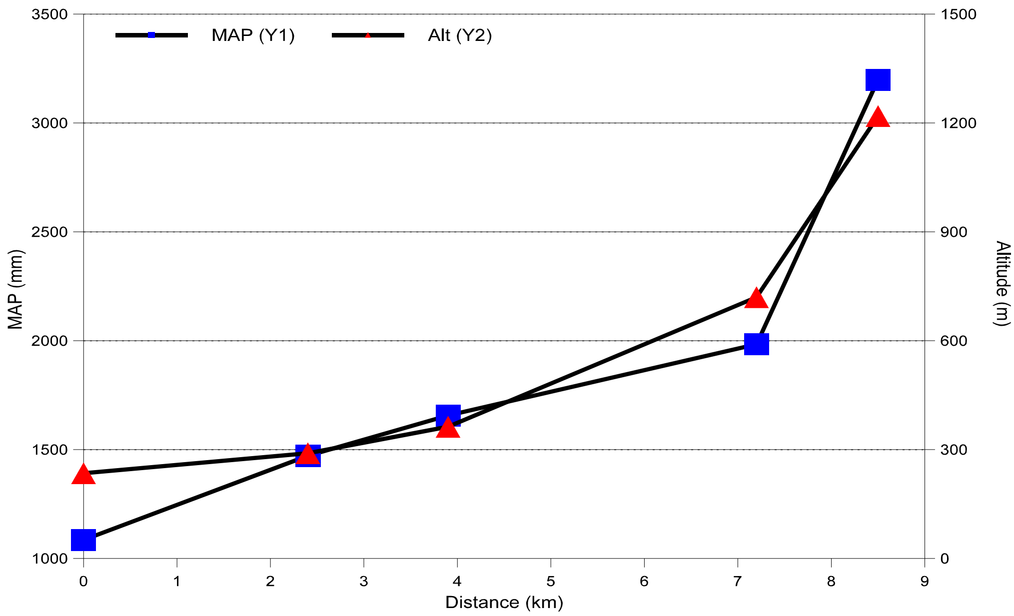


Figure 15 MAP and altitude relationship along the Jonkershoek mountain range

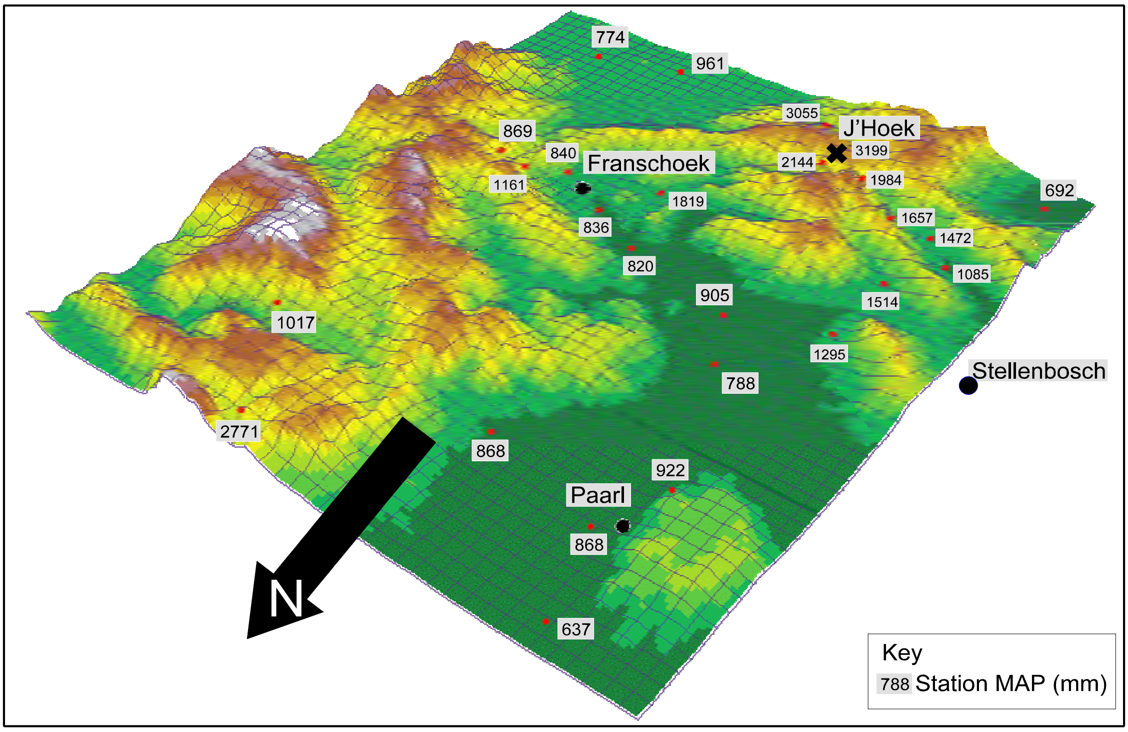


Figure 16 DTM and rainfall station locations in the area surrounding Jonkershoek

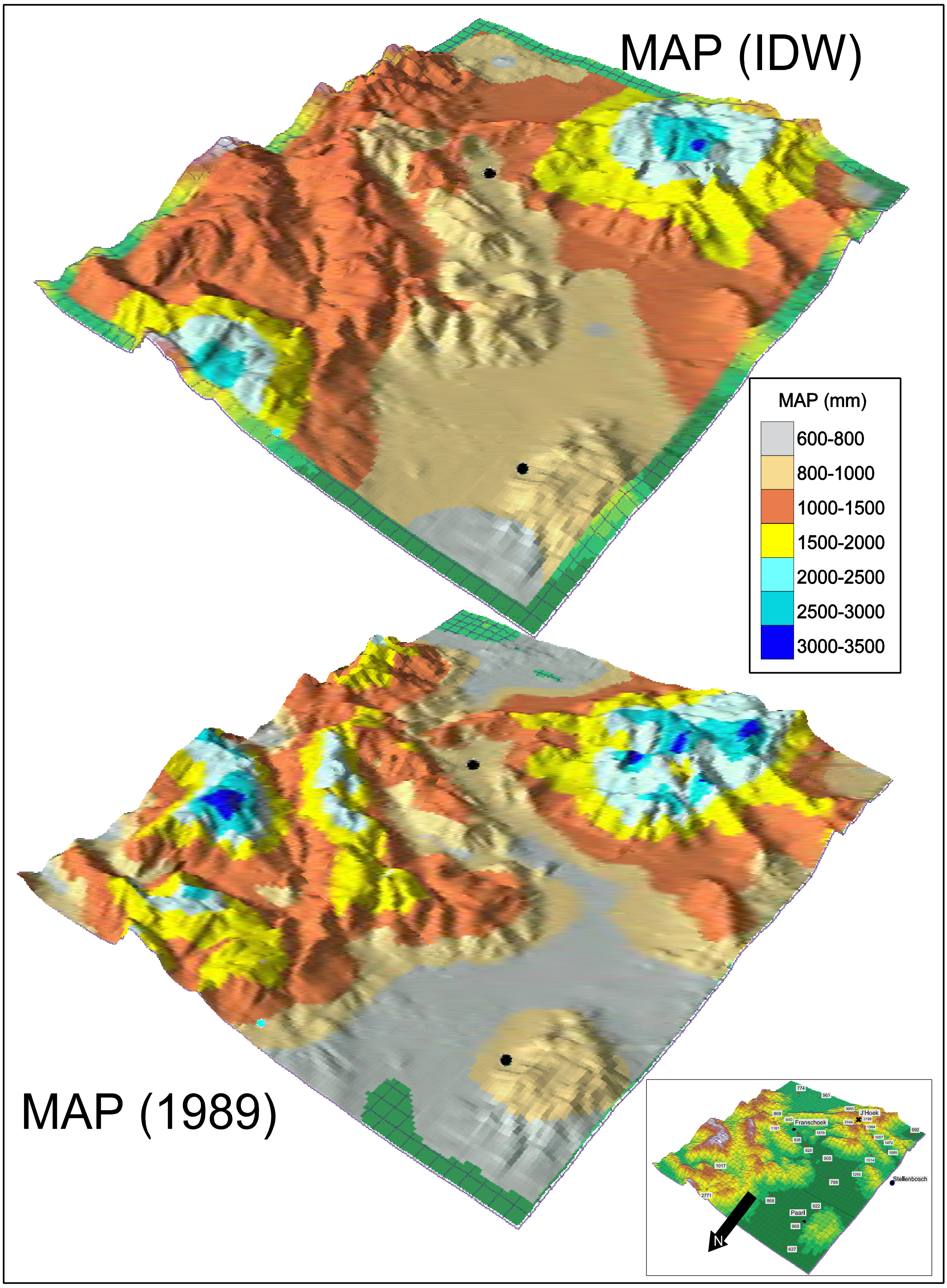


Figure 17 MAP surfaces draped over the DTM

CHAPTER 5

DERIVED RAINFALL RASTER DATASETS

This chapter covers all the various raster surfaces that are created from, *inter alia,* the point rainfall data. These include the annual, monthly and daily raster rainfall datasets or surfaces that form one of the main objectives of this project.

5.1 AREA OF INTEREST

Based on the total areas of the Primary catchments of South African rivers the raster rainfall surfaces extend to the northern Quaternary catchments boundaries (Figure 18) at the request of the project Steering Committee.

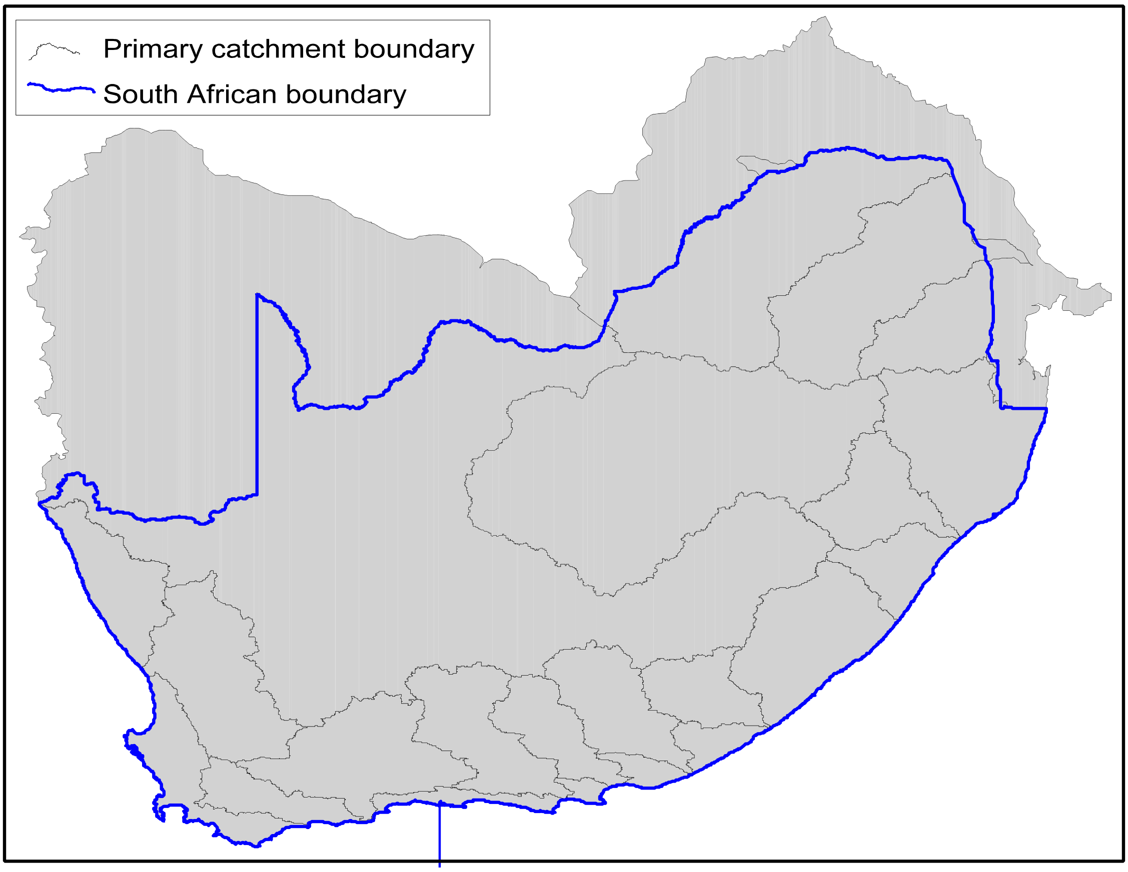


Figure 18 Area of interest selected on Primary catchment boundaries

5.2 ANNUAL PRECIPITATION

Although rain gauges measure rainfall at a point, their record can serve a broader purpose because it reflects locality factors such as continentality, altitude, longitude and aspect which govern the incidence of rainfall (Whitmore, 1967).

5.2.1 Mean Annual Precipitation

Mean annual precipitation is one of the most widely used variables in hydrological design, water resources planning and agrohydrology. Mean annual precipitation is the total annual precipitation calculated from all yearly total precipitation data for the period of interest divided by the number of years in the period of interest. One missing day of rainfall, implies a missing month which results in the whole year being excluded from the calculations, although this is only applicable to this study and cannot be generalised.

5.2.1.1 Geographically Weighted Regression Approach

The research into finding a suitable technique to represent the spatial variation in MAP concluded that GWR should be used (Hughes *et al*., 2001; Lynch, 2001a; Brunsdon, 2002). Once this conclusion had been made the decision into which set of explanatory variables should be used began. A host of different approaches were used in an attempt to find which set of explanatory variables would produce the best, or most appropriate, spatial estimate of MAP. Hutchinson (1998) found that the cross validation technique does not always represent a reliable estimate of model error, especially when short range correlation in data is present.

Another method of deriving the bandwidth and selection of the explanatory variables, which provides a trade-off between goodness-of-fit and degrees of freedom, is to minimise the AIC (Akaike Information Criteria). The AIC has the advantage of being more general in application than cross-validation (Fotheringham *et al*., 2002).

A spatial database consisting of all the different MAP rasters, created using GWR, using more than 20 different combinations of explanatory variables was created to assist in the selection of the appropriate model. In other words, this database contains a raster of MAP, estimated using a particular set of explanatory variables. The idea behind this is to determine which areas have the same estimate of MAP irrespective of the GWR model. The analysis yielded a one arc minute raster for which each pixel contains a number representing the number of times that the more than 20 rasters at that pixel have estimates that are within ±10% of the average of the values at that pixel (Figure 19). The larger part of South Africa comprises of areas that are not topographically complex (Figure 11) and this is highlighted once again in Figure 19. The areas where less that half of the GWR models do not produce similar MAP estimates correspond well with areas that are topographically complex (Figure 11) and where there is a large void in rainfall stations (Figure 1) and these are the areas where more detailed analysis should be performed to determine which GWR is the most appropriate.

5.2.1.1.1 Adjustment of the GWR Raster

The majority of the regression procedures endeavour to minimise the residuals. The aims of selecting different sets of explanatory variables is to produce a model that will fit the data the best, in other words, the model that produces the smallest residuals. These residuals, however, also mean that when the raster at the station pixel are queried, a different MAP

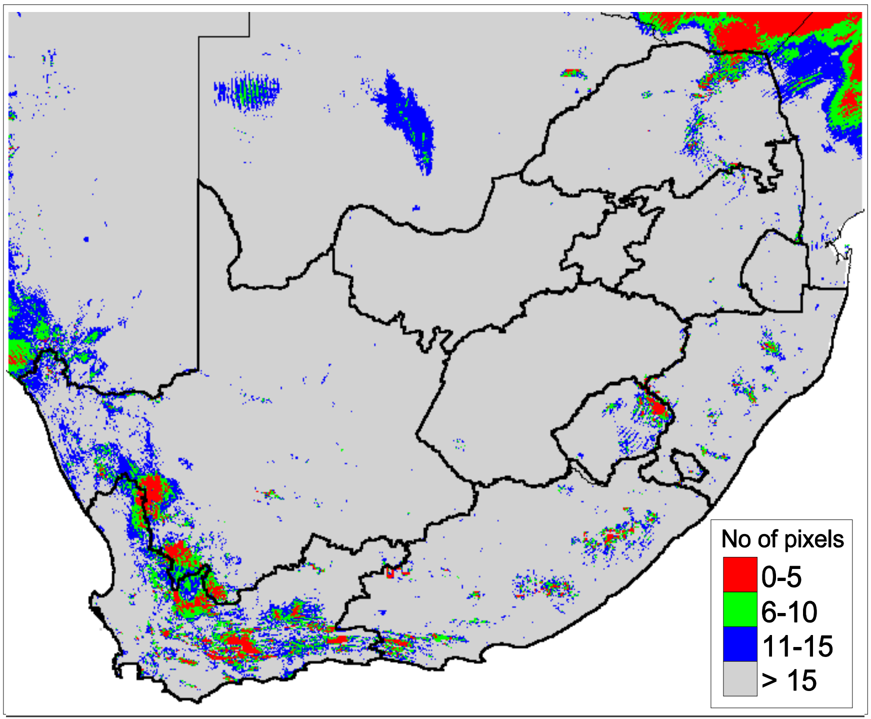


Figure 19 The number of pixels out of 20, calculated using all the different GWR models, that are within 10% of the average value at that pixel

value to that recorded in the point rainfall database may be obtained. A technique that was used successfully by Dent *et al*. (1989), *viz*. using these residuals to enhance the GWR raster, was once again used. If the MAP raster is not adjusted then one would not be able to overlay the station MAP values on the isohyetal map (Figure 20) as the 850 mm MAP point value should not fall between the 700-800 mm isohyet.

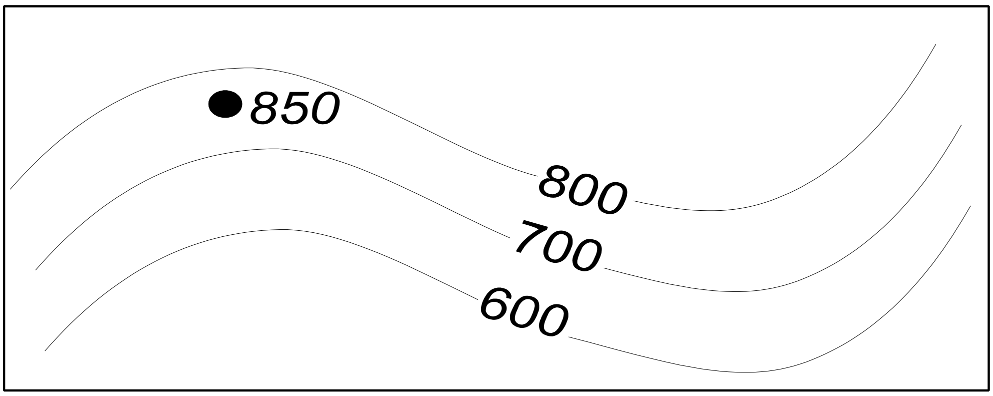


Figure 20 Reason for adjusting the regressed surface

The residuals, i.e. the difference between the observed MAP and the MAP estimated using GWR, are interpolated onto a rectangular one arc minute raster using IDW. This raster is then added to the one arc minute MAP raster generated using GWR (Figure 21). This process has adjusted the initial GWR surface locally, to fit where there are observed values, and globally, at the ungauged pixels, using the residual information.

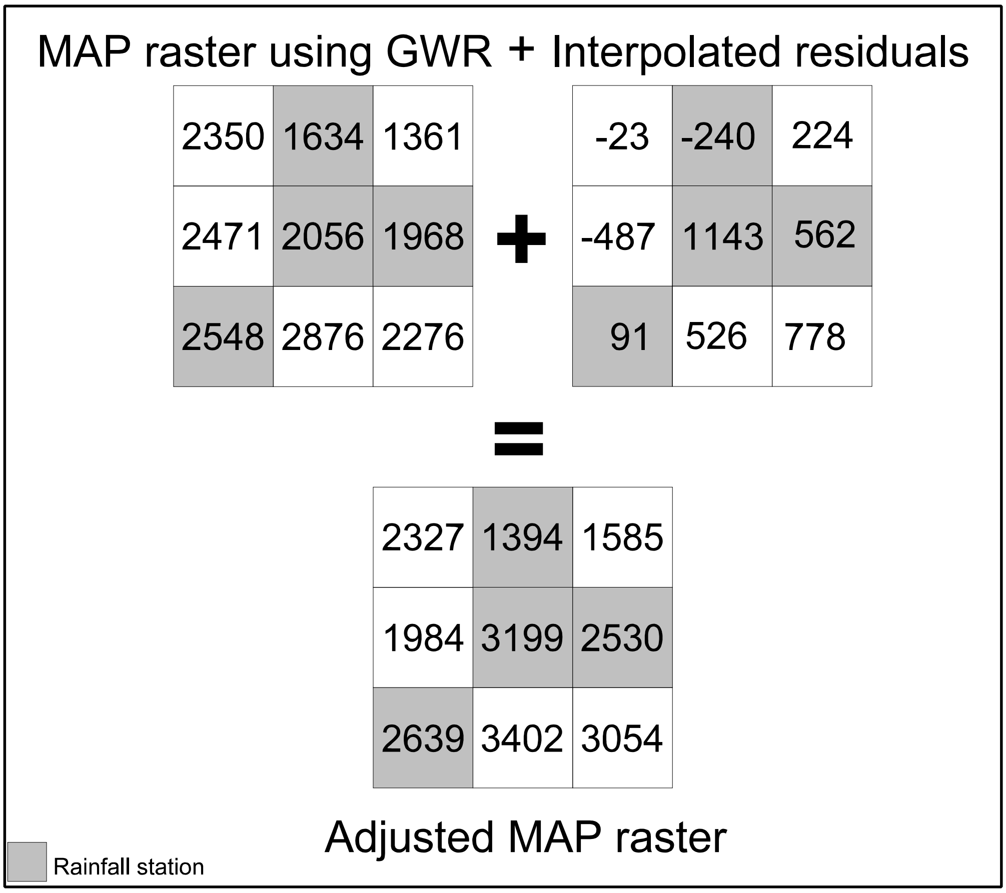


Figure 21   Technique for adjusting the regressed MAP surface

In other words, the surface was not only adjusted to fit the MAP at the rainfall stations but the surrounding areas were also adjusted according to the interpolated residual surface.

This process can also be explained by thinking of pulling a blanket up with one hand and pressing down on the bed with the other hand. The areas surrounding the hands are also adjusted and not only where the hands touch the blanket, which would happen if only the pixels, where observed data are available, were adjusted.

5.2.1.1.2   The MAP Raster Estimated using GWR

The set of explanatory variables that generate the most plausible estimate of MAP are:

- ialtCV, i.e. the coefficient of variation (CV) of a 5 arc minute mask of altitude values,

- latlong, i.e. the product of the latitude and longitude co-ordinates (degrees decimal) of the pixel
- xplusy, i.e. the sum of the latitude and longitude co-ordinates (degrees decimal) of the pixel,
- xx, the square of the longitude co-ordinates (degrees decimal) of the pixel, and
- slope, i.e. the slope in degrees of the 8 pixels surrounding the pixel in question.

The fact that distance from the sea does not feature could be explained by the fact that the interactions of these selected explanatory variables vary spatially and could mimic it.

The MAP surface was estimated using an adaptive bandwidth of 0.006782, which implies that 0.6782% of the data are used at each regression point. The MAP raster (Figure 22) was then adjusted using the residual interpolation technique.

The regression coefficients that are used to produce the final adjusted MAP raster (Figure 22) vary from negative to positive values (Table 7).

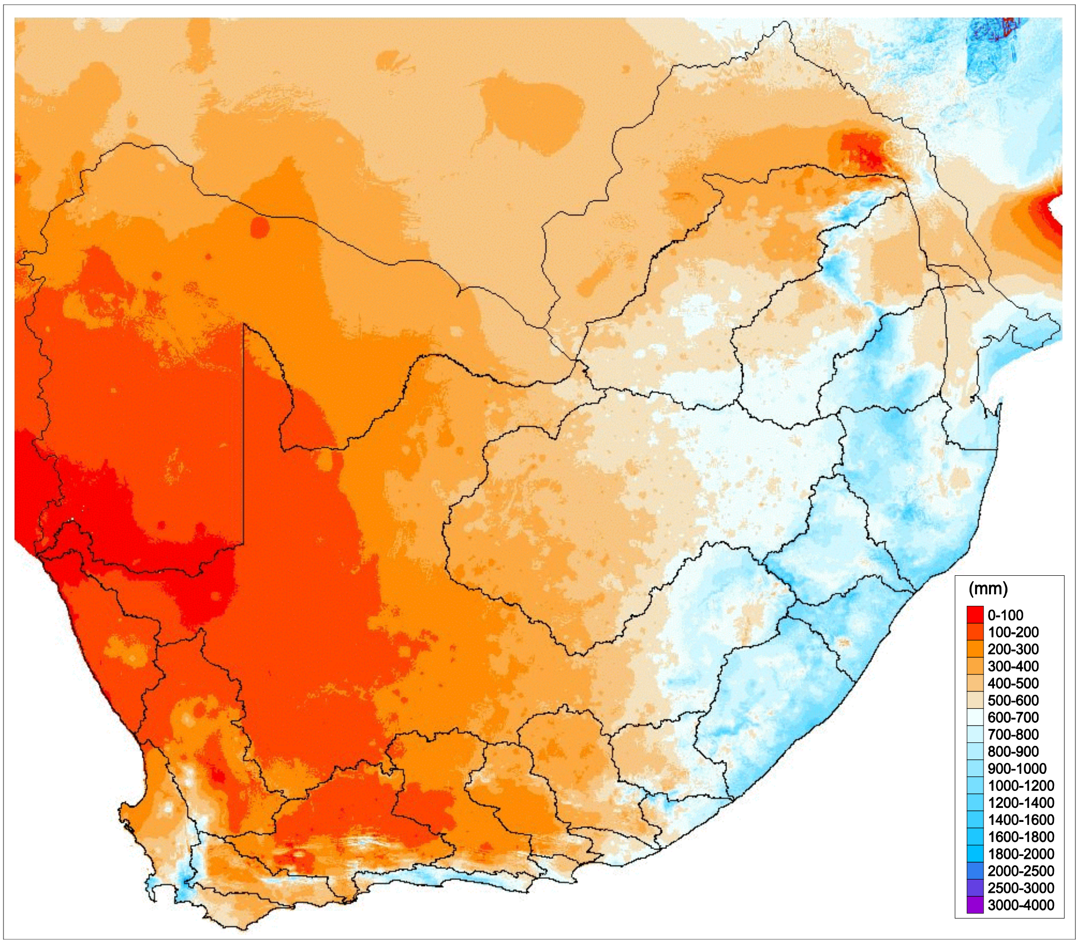


Figure 22 Final adjusted MAP surface determined using GWR

Table 7 Extremes of the regression and explanatory variables, with abbreviations explained in the text

Explanatory Variables	Regression coefficients		Explanatory Raster	
	Minimum	Maximum	Minimum	Maximum
ialtCV	-68	76	0	254
latlong	-381 587	312 473	-1 186	-318
xplusy	-95 983	72 208	-18.8	14.3
xx	-532 833	280 461	259	1 156
slope	-309	281	0	20.2
yint	-176 865	569 947		

The regression coefficients also vary spatially over the area of interest (Figure 23), which once again emphasises the spatial non-stationarity of the data. The interactions of the coefficients at some pixels are akin to the introduction of a "new" explanatory variable, i.e. in certain areas the product of latitude and longitude (latlong), for example, might have the same effect if distance from the sea, for example, were used at that location.

The regression equation (Eq. 1) used to generate the final MAP raster (Figure 22) is simple in form, but the interactions of the coefficients over space allow for the best estimation of MAP.

$$MAP_{Estimate} = ialtCV_C*ialtCV_R + latlong_C*latlong_R/100 + splusy_C*splusy_R + xx_C*xx_R/100 + slope_C*slope_R + yint \quad \text{...... Equation 1}$$

where the subscripts C and R refer to the regression coefficients and to the explanatory variables respectively and

ialtCV = the percentage coefficient of variation (CV) of a 5 arc minute mask of altitude values
latlong = product of the latitude and longitude co-ordinates (degrees decimal) of the pixel
xplusy = sum of the latitude and longitude co-ordinates (degrees decimal) of the pixel
xx = square of the longitude co-ordinates (degrees decimal) of the pixel
slope = slope in degrees of the 8 pixels surrounding the pixel in question
yint = y intercept term.

An example is used to illustrate how Eq. 1 is used to generate a MAP estimate for a single one arc minute pixel in the Western Cape province. The values of the explanatory variables for the pixel in question (lat = 34°S; long = 19.0167°E) are listed in Table 8, as are the regression coefficients derived using GWR.

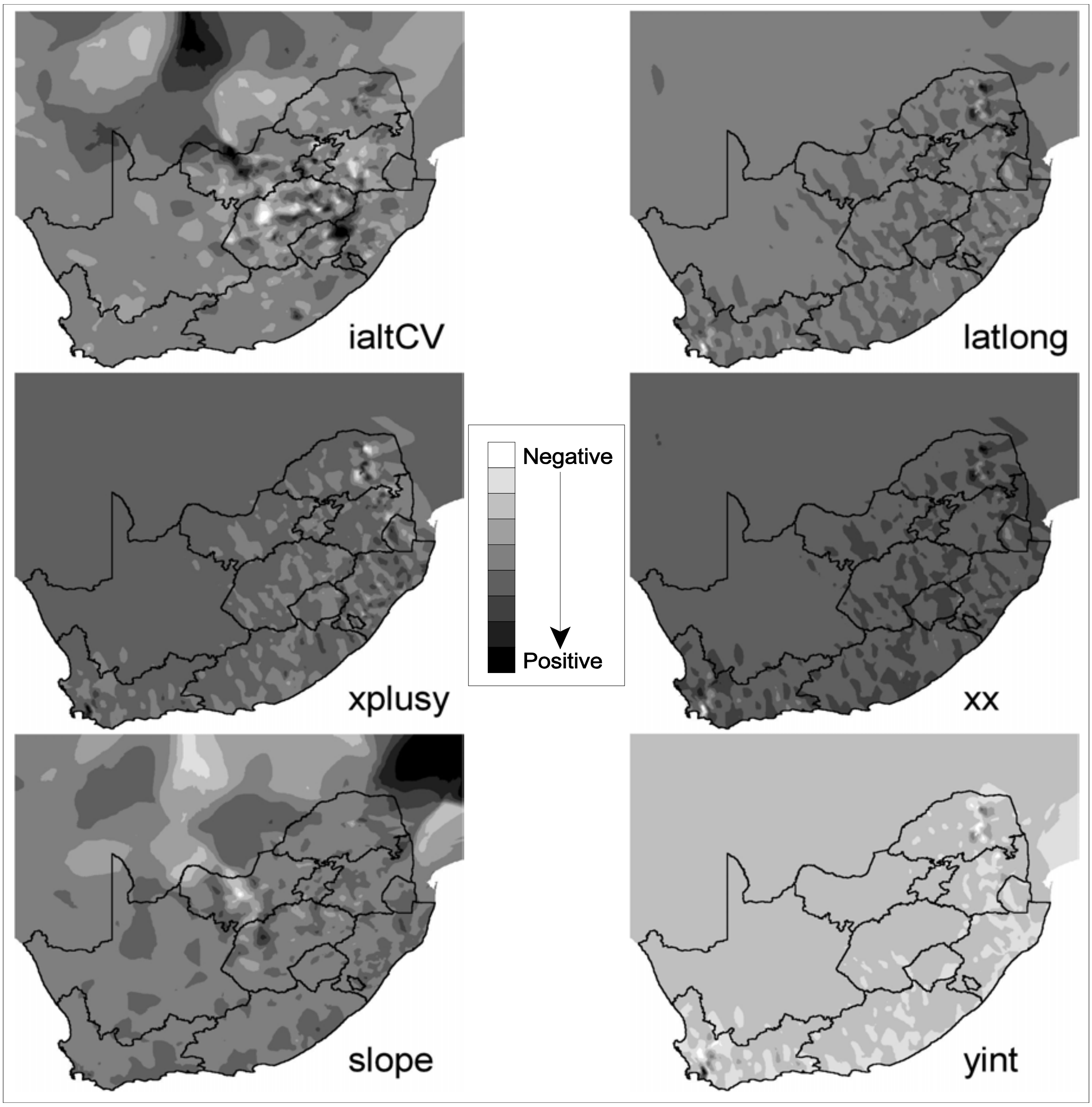


Figure 23 Spatial variations of the final GWR coefficients used to estimate MAP

This procedure is repeated for each pixel (there are roughly 980000 one arc minute pixels comprising the rectangle surrounding the area of interest) and thereafter the residual adjustment technique is applied to create the final MAP raster (Figure 22).

The data used to produce the final MAP raster are listed in detail in the Appendix ***resources required.pdf***.

Table 8 Worked example of a regression at a selected pixel

Explanatory variable	Coefficient at pixel	Raster value at pixel	Result
ialtCV	-29.8552134947446	21.0000000000000	-626.9594833896366
latlong / 100	-336456.60457950100	-646.56780000000	2175420.06618437900
xplusy	65743.3460279914000	-14.9833000000000	-985052.27654120350
xx / 100	-477754.02300087800	361.63487889000	-1727725.1824713280
slope	16.8819428786078	7.3310670000000	123.7626543332466
yint	539790.145578250100		539790.145578250100
$MAP_{Estimate}$			1929.55592104

5.2.1.2 Inverse Distance Weighting Technique

An Inverse Distance Weighting interpolation was performed on the current database of MAP values. An optimal value of 2.5227 for the power term was determined, in other words, the weighting term of the IDW is raised to the power of 2.5227 and the octant search method, with at least 10 samples per sector, was used (Johnston *et al*., 2001). The resultant MAP surface (Figure 24) is “spotted”. This is a characteristic trend when using IDW (Burrough, 1986). The IDW method is often used as it is a “quick and easy” and is freely available. The IDW method can be used in areas where the MAP values do not vary much spatially, but it should under no circumstances be used in areas where the MAP varies spatially.

5.2.1.3 Multiple Linear Regression Procedure

Dent *et al*. (1989) used a multiple linear regression approach to produce the MAP raster. South Africa was divided in 34 rectangular regions (Figure 25), with a 15 arc minute overlap or common areas. These regions were manually and subjectively selected and in hindsight one would rather prefer to have applied a more scientific selection method, *viz.* the adaptive bandwidth employed in the GWR process.

For each region a subset of possible models was first chosen by using forward, backward and stepwise regression, based on R^2 values and then a final model was chosen based on the PRESS (predicted error sum of squares) statistic. The latter was used as the most appropriate model for mapping MAP. It may not be the model that fits the data best, but rather the model that gives the best predictions (Dent *et al*.,1989).

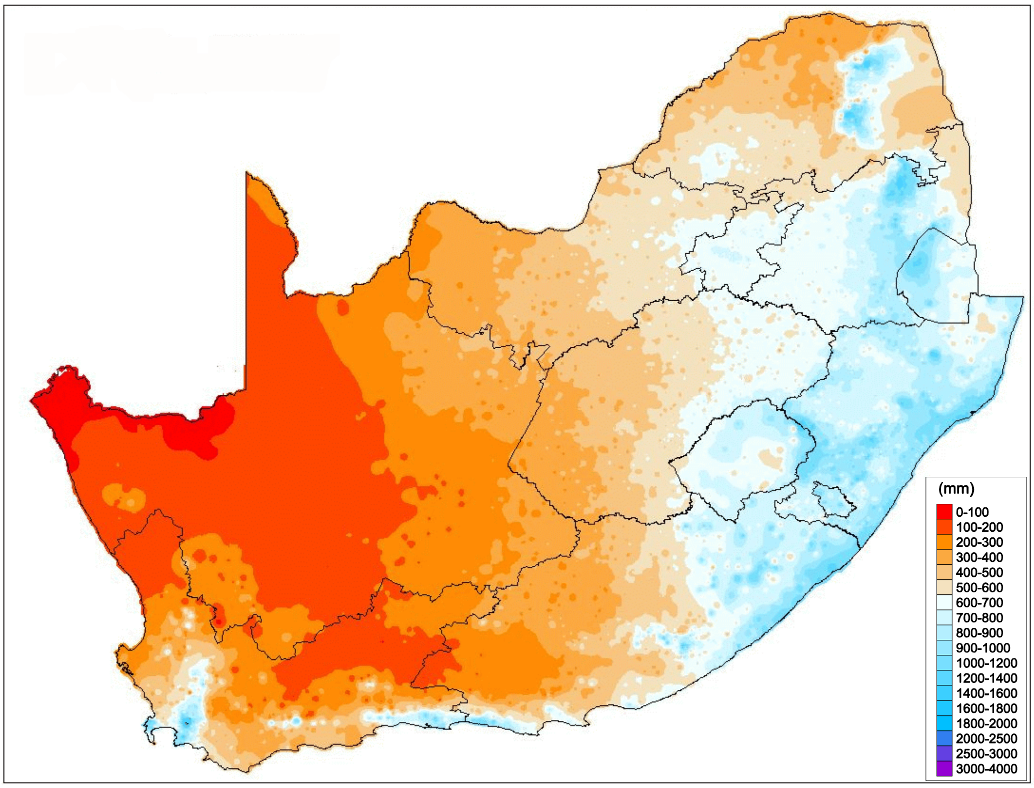


Figure 24 MAP estimated using IDW with an optimal power term of 2.5227

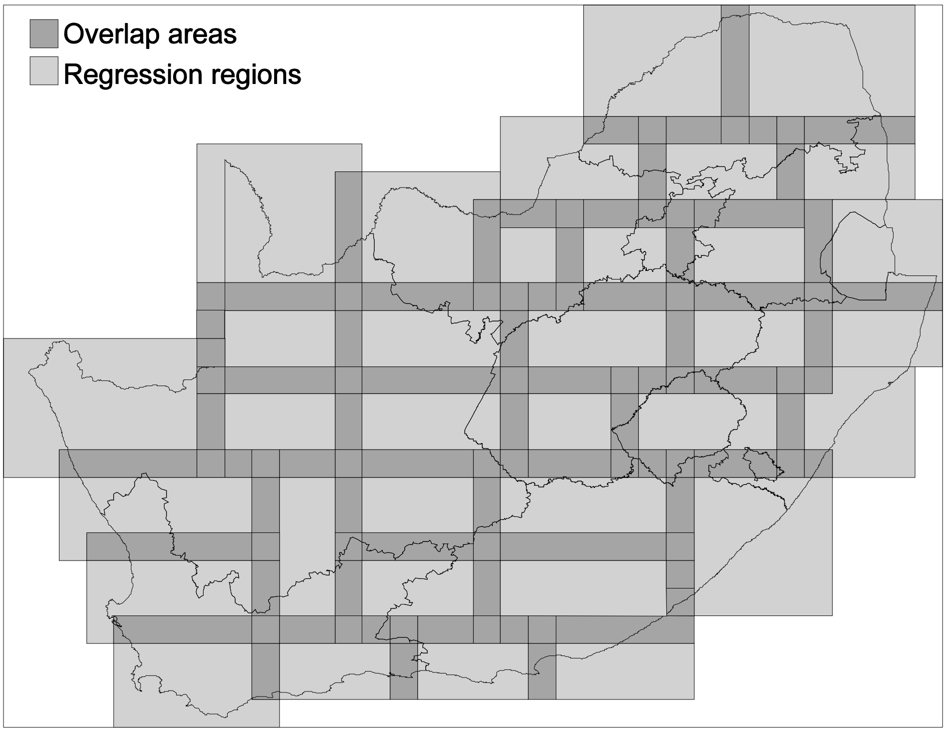


Figure 25 Regression regions used by Dent *et al*. (1989) to estimate MAP

This MAP surface (Dent *et al*., 1989; Figure 26) has been used for the past decade as the *deo facto* MAP raster for South Africa. Any new MAP raster that is produced should, at least, bear major similarities to this one. There are, however, areas where more accurate altitude data are now available (Lynch, 2002) and these regions will definitely have different MAP values, especially in Lesotho and the Northern Cape province.

There are areas where the differences will also be apparent as a result of more detailed altitude data (Lynch, 2002) and the fact that the lengthening of the rainfall database by more than 15 years (since the Dent *et al*., 1989 maps) will yield a denser network of gauges.

5.2.1.4 Stochastic Daily Rainfall Method

McNeill *et al*. (1994) produced, for each one arc minute pixel covering South Africa, a series of parameters for the estimation of stochastic daily rainfall. These stochastic rainfall time-series datasets are not chronologically linked; rather, they are "packets" of 365 values. The long term statistics and the occurrence of rainfall events are similar to the observed data. Fifty years of daily stochastic data for each one arc minute pixel were generated and the corresponding MAP values were determined (Figure 27). Some of the key stations in South Africa, *viz*. Jonkershoek, the highest recording station which only has monthly records, were not included in the McNeill *et al*. (1994) dataset as only daily rainfall was used. This MAP raster does not exhibit the same variability in the mountainous regions as do the rasters estimated using additional explanatory variables. This is due to the omission of key stations that were used by McNeill *et al*. (1994).

5.2.1.5 Cokriging Approach

Cokriging is a statistical interpolation method that uses explanatory variables to assist in interpolating independent variables. Hughes (2001) and Hughes *et al*. (2001) found that kriging gives a very good overall impression for estimating MAP values. In particular as a result of the way it is defined, it will never give values that are far outside the original data values. However, it is unable to give fine detail, even though universal kriging does increase the amount of detail achieved. Also, kriging only indirectly takes into account the other explanatory variables, such as altitude, distance from the sea and roughness and thus does not fully utilise the information which is available. However kriging, both ordinary and universal, is quick and straightforward to implement, and is thus a method which can be used rapidly to give an initial overall map of the data.

Cokriging does not use the DEM, but only uses the altitude values at the station locations and then estimates an altitude value at the ungauged site. When considering an example where there are two rain gauges on hills that are separated by a valley it is clear that the valley feature would not be reflected.

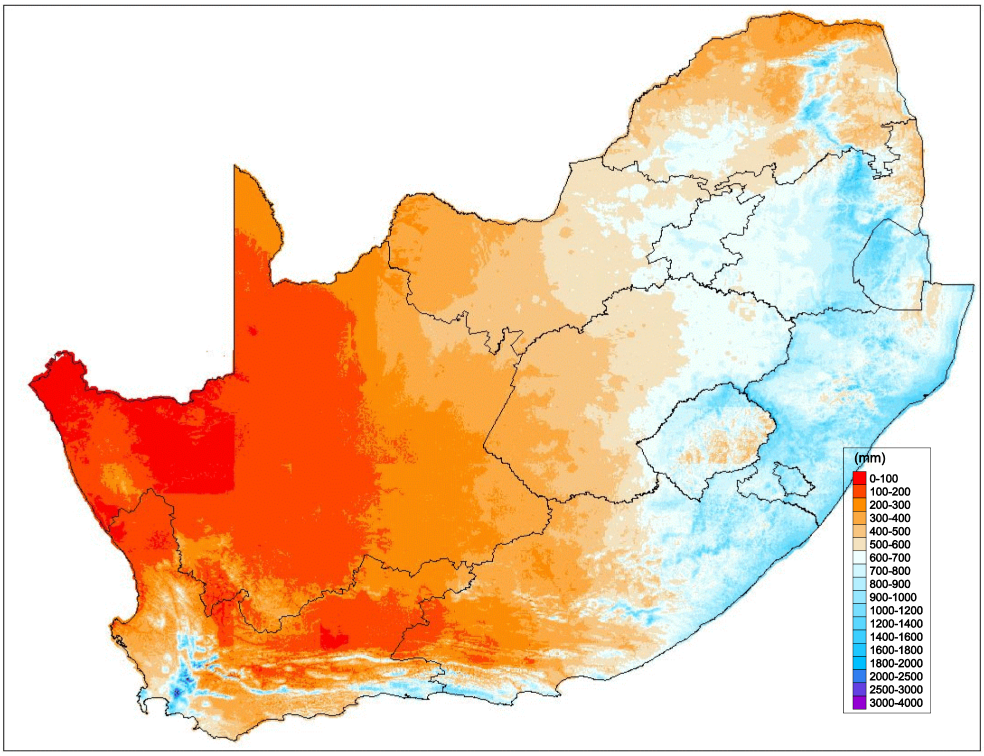


Figure 26 MAP surface estimated by Dent *et al*. (1989)

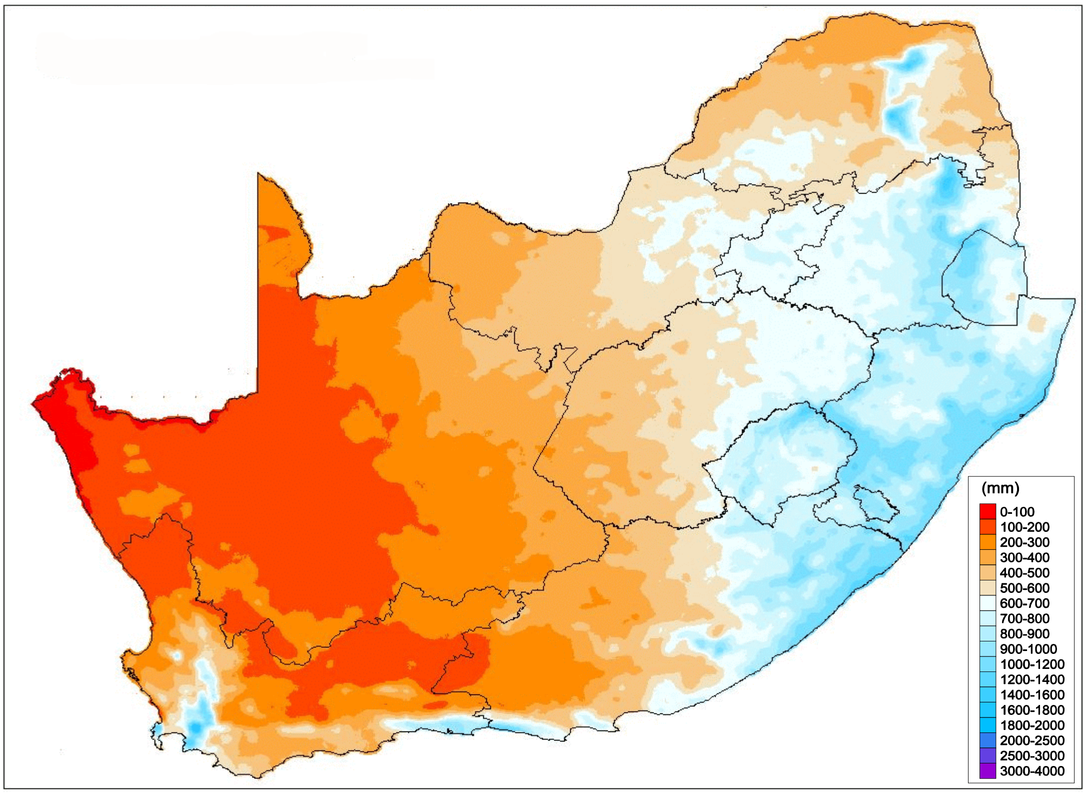


Figure 27 MAP surface estimated using stochastic rainfall values

The raster of MAP (Figure 28) estimated using ordinary cokriging (Johnston *et al*., 2001) was generated with ialtCV, latitude*longitude and slope as explanatory variables using the exponential semivariogram method (Figure 29) suggested by Clarke (1999).

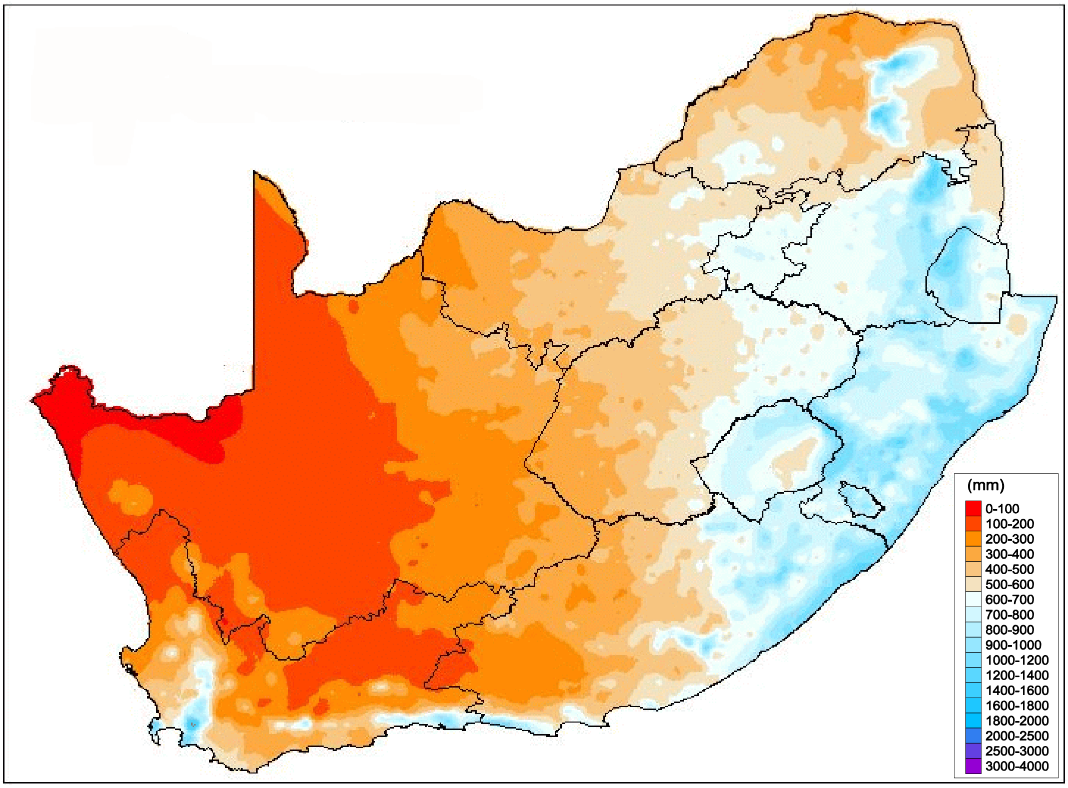


Figure 28 MAP surface estimated using cokriging

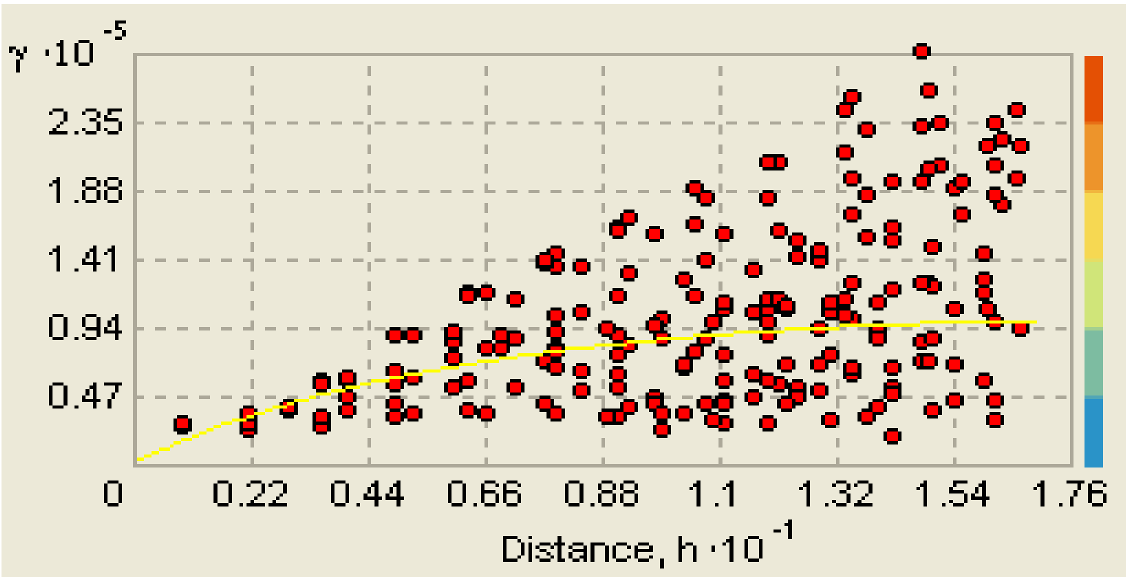


Figure 29 Semi-variogram for the cokriged estimate of MAP

The cokriged MAP surface (Figure 28) displays the same west-east trend of increasing rainfall shown in the Dent *et al*. (1989) MAP surface (Figure 26). The criticism that kriging over-smooths is, however, evident.

5.2.1.6 Analysis of MAP Rasters Derived Using the Different Techniques

This section includes an analysis of the MAP rasters estimated using the different regression and interpolation techniques. The minimum and maximum values (Table 9) differ, as one would expect. The maximum value of the MAP raster estimated using stochastic values (McNeill *et al*., 1994) is very low because the dataset used to estimate the stochastic parameters did not include some of the key rainfall stations in the Drakensberg and Jonkershoek areas. The CV values are all in the same range, but visual inspection of the MAP surfaces does, however, suggest that some are more smoothed than others, e.g. Figure 28, where the MAP is estimated using cokriging.

Table 9 Comparative statistics of the MAP rasters for South Africa when using different interpolation and regression techniques

Technique / Statistic	MAP'89	GWR	IDW	Cokriging	Stochastic
Minimum (mm)	20	1	43	45	51
Maximum (mm)	3 345	3 198	3 199	2 422	1 999
Average (mm)	462	448	452	446	470
CV (%)	56.81	55.26	53.78	54.68	54.54

All the methods reviewed in this section should produce very similar estimates of MAP in areas where there is no complex topography. The surface roughness index, ialtCV, is used to exclude topographic complex areas. In others words, only pixels where the ialtCV raster values are greater than 5% are selected (Figure 30). The mask of iAltCV>5 is used to select areas that are topographically similar. The pixels defined by this mask are then used to determine if there any similarity between these MAP rasters that were generated using a variety of different interpolation and regression techniques.

A large part of South Africa consists of areas where the rainfall and the topography are spatially not complex. The different regression and interpolation techniques produce similar MAP estimates for these areas (Figure 31). In order to estimate MAP in these areas one could, therefore, use the "easiest and fastest" technique at ones disposal.

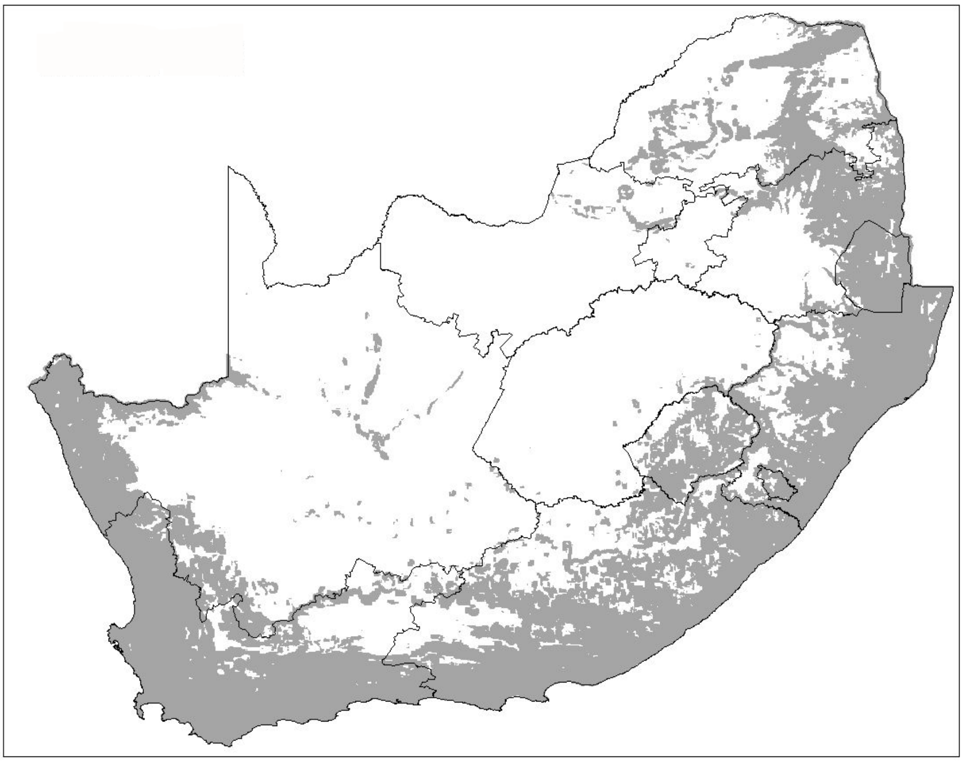


Figure 30 Areas where the explanatory variable ialtCV exceeds 5%

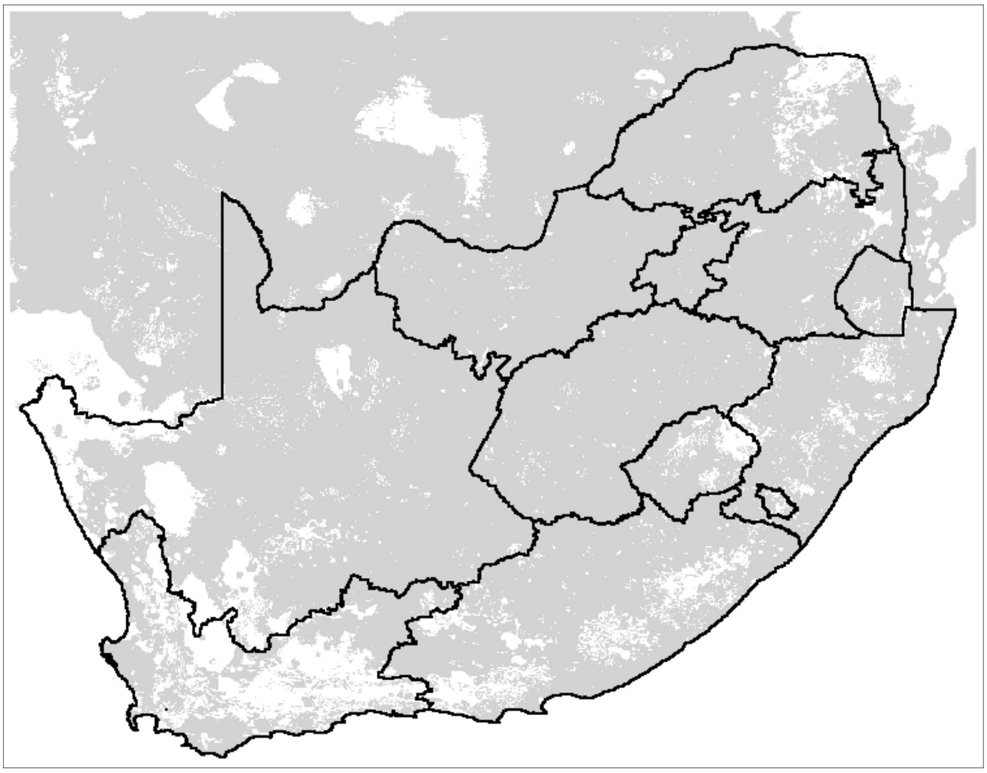


Figure 31 Areas that have the same MAP value irrespective of interpolation or regression technique used

The rasters generated using inverse distance weighting and cokriging have the highest percentage of similarity, *viz*. 94.36% (Table 10). In other words, 94.36% of the pixels defined by the mask, i.e. iAltCV.5% (Figure 31), are within 20% of one another. This is what would be expected, as they are both interpolation techniques. The results presented in Table 10 fail to suggest which interpolation/regression technique best mimics the MAP surface developed by Dent *et al*. (1989).

5.2.1.7 Differences Between the Previous and the New MAP Raster

There will be differences between the MAP raster of Dent *et al*. (1989) and any of the newer rasters for many reasons. The Dent *et al*. (1989) raster has been used extensively over the past 15 years and this section will try and point out where there are major differences and the reasons therefore. The GWR models that are used in this section are not necessarily the final one, as this research has been done throughout the course of the project and time does not allow for the results to be re-calculated using the final GWR model.

Table 10 Percentage of the masked pixels that are within 20% of each other raster

Technique	GWR	IDW	Cokriging	Stochastic
MAP'89	73.46	73.09	74.16	74.40
GWR		87.80	85.49	85.12
IDW2.5227			94.36	88.77
Cokriging				90.33

The first area displaying large differences is in the vicinity of Wellington and Worcester in the Western Cape province (Figure 32). There are 4 pixels (Figure 32) showing a difference of more than 2 000 mm in the MAP estimates. The 380 km^2 area has 6 rainfall stations in the current rainfall database. The database for the Dent *et al*. (1989) MAP raster, however, only had one rainfall station (Table 11) available for this area. The daily rainfall infilling techniques and the additional 15 years of data therefore enhance the spatial density of gauges in this area that has a complex topography.

The DEM that was used in 1989 has also been enhanced substantially (Lynch, 2002) and the differences are listed in Table 12. The CV of the raster of this study area has changed from a high value to a lower value and the fact that altitude plays an important role in the estimation of MAP in mountainous regions could account for some of the other differences.

The 1989 MAP surface was incorrectly estimated as a result of the altitude values in the initial one arc minute DEM. The differences between the DEMs and the MAP surfaces are illustrated in Figure 33.

Table 11 Comparison of data available in the Wellington/Worcester region in the Western Cape province during 1989 and 2001

Station-id	Available for Dent *et al*. (1989) MAP raster		Data in Lynch (2003) database	
	MAP (mm)	No of years	MAP (mm)	No of years
0022399AP	n/a	n/a	770	99
0022104 W	1 145	4	950	99
0022253 W	n/a	3	900	99
0022284 A	1 030	11	960	142
0022140 P	n/a	n/a	840	20
0022471 W	810	31	705	143

Table 12 Differences between the 1989 and 2002 DEM in the Wellington/Worcester region in the Western Cape province

Statistic	DEM in 1989	DEM in 2002
Minimum (m)	131	212
Maximum (m)	1 798	1 722
Mean (m)	854	892
CV (%)	48.9	43.7

5.2.2 Errors Associated with the MAP Raster

The question that is often asked when one uses raster data is, what errors are associated with the estimation of the raster? The interpolation and regression techniques cannot produce exact estimates of rainfall values and one has to produce some or other map or table to indicate the accuracy of the estimates. The fact that kriging produces error maps is often the reason that it is used to determine estimates at ungauged positions.

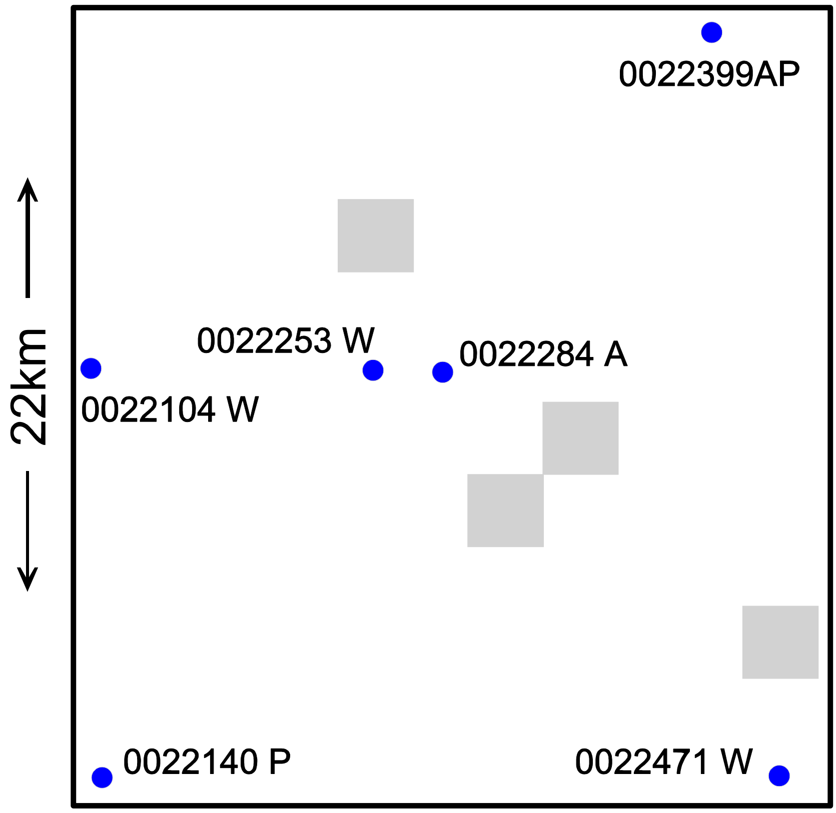
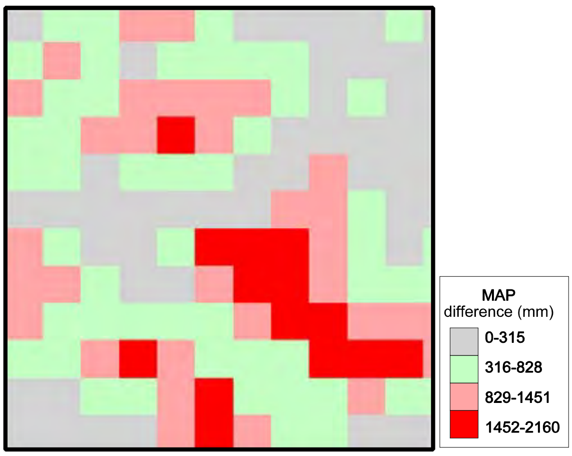


Figure 32 High residual values in the Wellington/Worcester region in the Western Cape province



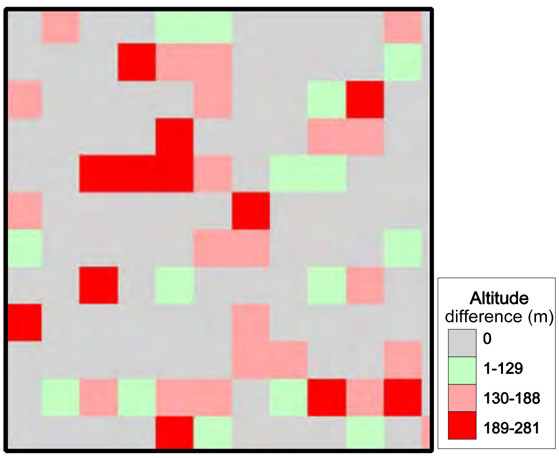


Figure 33 Incorrect estimates of MAP (Dent *et al*., 1989) as a result of incorrect altitudes values in the Wellington/Worcester region in the Western Cape province

The GWR technique does not produce an estimate of the error *per se*, but the residuals do, however, allow for the estimation of the accuracy associated with the particular raster. The residual values are interpolated, using IDW, onto a one arc minute raster, as was done in the adjustment of the GWR surface earlier in this chapter. This residual raster can be used to associate some form of error with the GWR technique that was used to estimate the MAP. The fact that these residuals have been incorporated into the final MAP surface therefore implies that one can only assume that the MAP estimates are more accurate that those generated using only GWR.

There are 7 pixels, out of a total of nearly one million, that have residual values of more than 500 mm and 99.2% of the area has residuals between ±100 mm. Rainfall station 0678883 W northwest of Tzaneen in the Limpopo province (Figure 34) has the largest negative residual of 538 mm.

The selected area (Figure 34) is roughly 32 km^2 and there exists a sharp increase in MAP from the northwest to the southeast. The highest residual is also located in an area void of other rainfall stations and the altitude varies slightly with a CV of 24%. The adjustment technique whereby the residuals are incorporated into the GWR surface allows for a seamless transition which would otherwise have had a significant rainfall gradient surrounding the pixel with the large residual.

The highest positive residuals are located in the vicinity of Jonkershoek and an example of their distribution can be found in Figure 21 in the adjustment of the GWR raster section.

5.2.3 Annual Rainfall Totals

An hydrological year in South Africa commences on 1 October and runs through to the next calendar year and ends on 30 September. Approximately 8 000 rainfall stations with more than 15 years of rainfall data were analysed to determine which month is the dominant month to start the hydrological year in South Africa. The smallest and the largest percentages where the one month has more monthly rainfall than the preceding month appears around September and October (Table 13). Schulze (2002), on the other hand, suggests that the hydrological year starts in October because most of the world have defined rainfall and thus runoff seasons which would be split if hydrological year started in January. A mid-season (equinoctal) date thus makes sense worldwide.

5.2.3.1 Calendar Year Approach

One of the criticisms levelled at using a calendar year to determine annual rainfall is that in the summer rainfall region December and January are both high rainfall months, but would be split into two calendar years. The layperson, however, invariably uses the calendar year approach. Many scientific computer programs have also been developed using this yardstick and a possible explanation might be the few additional lines of computer code that are required to split a year into three and nine month periods.

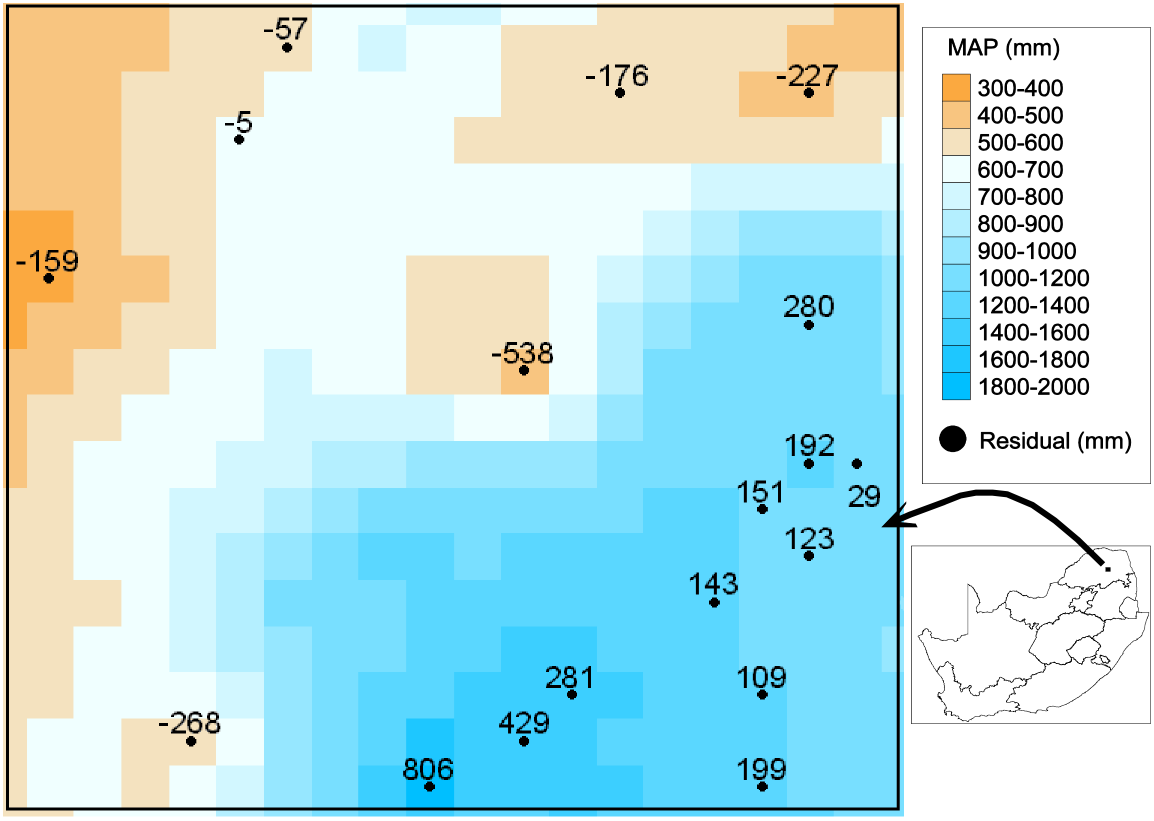


Figure 34 Area near Tzaneen highlighting some high residual values

Table 13 Percentage of rainfall stations that have more rainfall than in the preceding calendar month and *vice versa*

Boolean expression	% of the rainfall stations where the boolean expression is	
	True	False
Jan > Feb	57.05	42.95
Feb > Mar	57.44	42.56
Mar > Apr	91.46	8.54
Apr > May	90.68	9.32
May > Jun	91.26	8.74
Jun > Jul	55.16	44.84
Jul > Aug	20.02	79.98
Aug > Sep	13.46	86.54
Sep > Oct	8.15	91.85
Oct > Nov	12.33	87.67
Nov > Dec	30.13	69.87
Dec > Jan	22.83	77.17

The raster for the 1903 calendar year (see Appendix) has the most red pixels (391 257 of 429 700 one arc minute pixels) and is, therefore, the driest year spatially (Table 14). The raster for the 1976 calendar year (see Appendix ***calendar_yrs.pdf***) has 412 040 blue of 429 700 one arc minute pixels, which makes it spatially the year in which the most of South Africa experienced above normal rainfall (Table 14).

5.2.3.2 Hydrological Year Approach

The hydrological summation process starts in October of a year and ends in September of the next year. The rationale behind this is that a complete summer or winter rainfall period should be accounted for in each total.

The raster for the 1932/33 hydrological year (see Appendix ***hydrological_yrs.pdf***) has the most red pixels (233 370 of 429 700 one arc minute pixels) and is, therefore, the driest year spatially (Table 15). The raster for the 1975/76 hydrological year (see Appendix ***hydrological_yrs.pdf***) has 268 476 blue of 429 700 one arc minute pixels which makes it spatially the year in which the most of South Africa experienced above normal rainfall (Table 15).

Table 14 Ten driest and wettest years using calendar years

Driest years spatially	Wettest years spatially
1903	1976
1992	1996
1966	1917
1919	1921
1926	1909
1927	1974
1945	1925
1908	1967
1916	1939
1984	1955

5.2.3.3 Similarities Between the Calendar and Hydrological Year Approach

Approximately one third of South Africa's surface area is covered by only 11.73% of the rainfall stations and 5.74% of the surface area is covered by 11.20% of the rainfall stations (Table 16).

The metropolitan areas are generally well gauged whereas the Northern Cape province, the largest of the provinces, has the lowest density of rainfall stations. However, it also does not receive much rainfall (Figure 35).

Table 15 Ten driest and wettest years using hydrological years

Driest years spatially	Wettest years spatially
1932/33	1975/76
1925/26	1924/25
1902/03	1973/74
1948/49	1938/39
1965/66	1960/61
1991/92	1995/96
1944/45	1906/07
1926/27	1949/50
1982/83	1996/97
1918/19	1956/57

Table 16 Frequency analysis of surface area represented by rainfall stations

MAP (mm)	% RSA surface area	% Rain gauges
0-300	32.17	11.73
300-600	38.13	39.27
600-900	23.96	37.80
>900	5.74	11.20

Raster surfaces of annual rainfall, using an IDW technique, are developed for the hydrological years between 1899/1900 and 1999/2000 and for the calendar period between 1900 and 1999 (see Appendix ***hydrological_yrs.pdf*** and ***calendar_yrs.pdf***).

The annual rainfall totals derived from these raster surfaces are generally less than the rainfall totals, calculated using only the rainfall station data, by a factor of 0.8 (Figure 36 and Figure 37 respectively). What this means is that the northwestern part of South Africa has a few number of stations, but occupies a large portion of the surface area, i.e. for this area there are many pixels with low MAP values and not many station MAP values.

The sequence of raster annual totals derived using hydrological years exceeds the 700 mm mark on eight occasions, whereas the calendar years raster totals only exceed it six times (Figure 36 and Figure 37 respectively).

A five term moving average filter is applied to both the hydrological and calendar year totals derived using rainfall station point data and the overlap is significant (Figure 38).

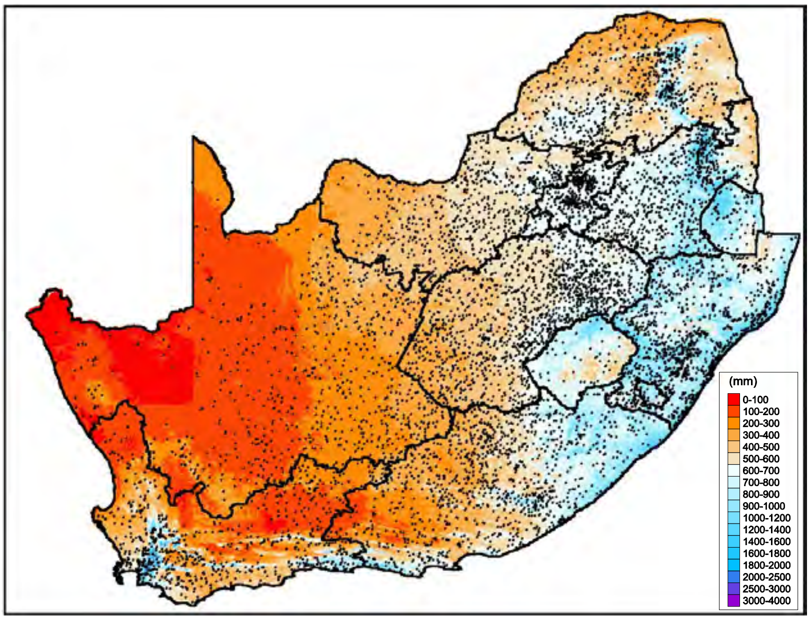


Figure 35 Rainfall station positions on the MAP surface estimated by Dent *et al*. (1989)

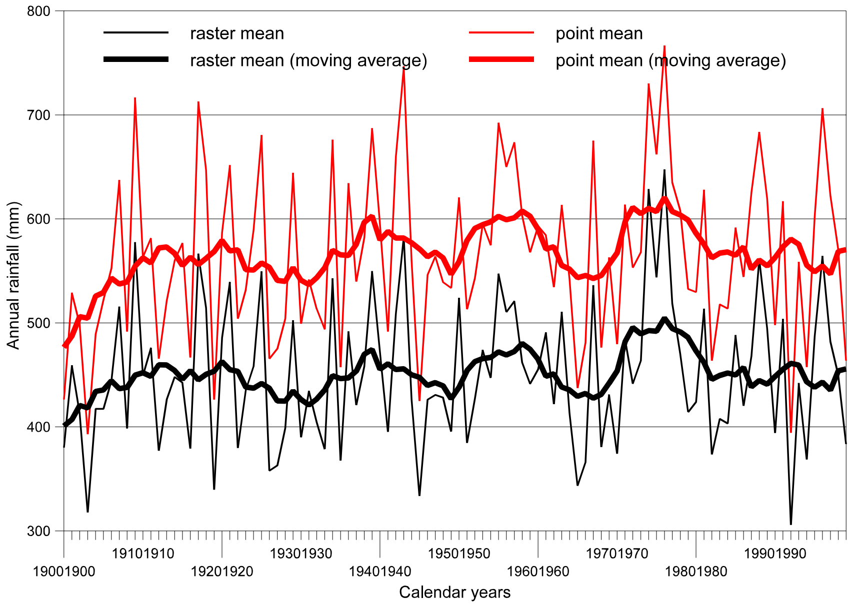


Figure 36 Raw and smoothed annual rainfall totals of raster and point values calculated using calendar years

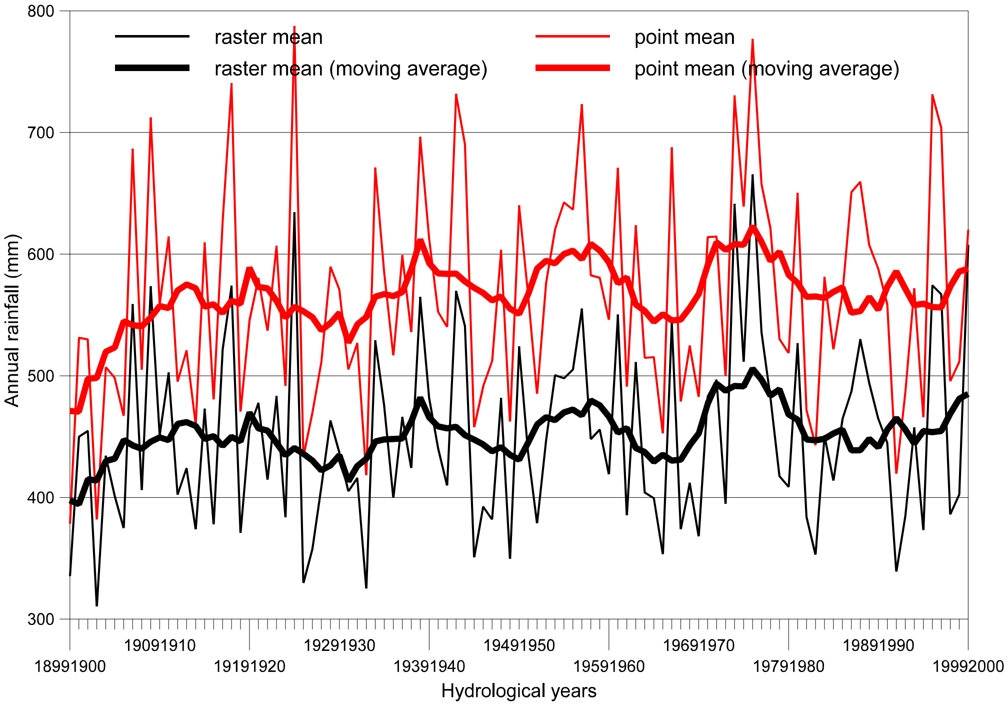


Figure 37 Raw and smoothed annual rainfall totals of raster and point values calculated using hydrological years

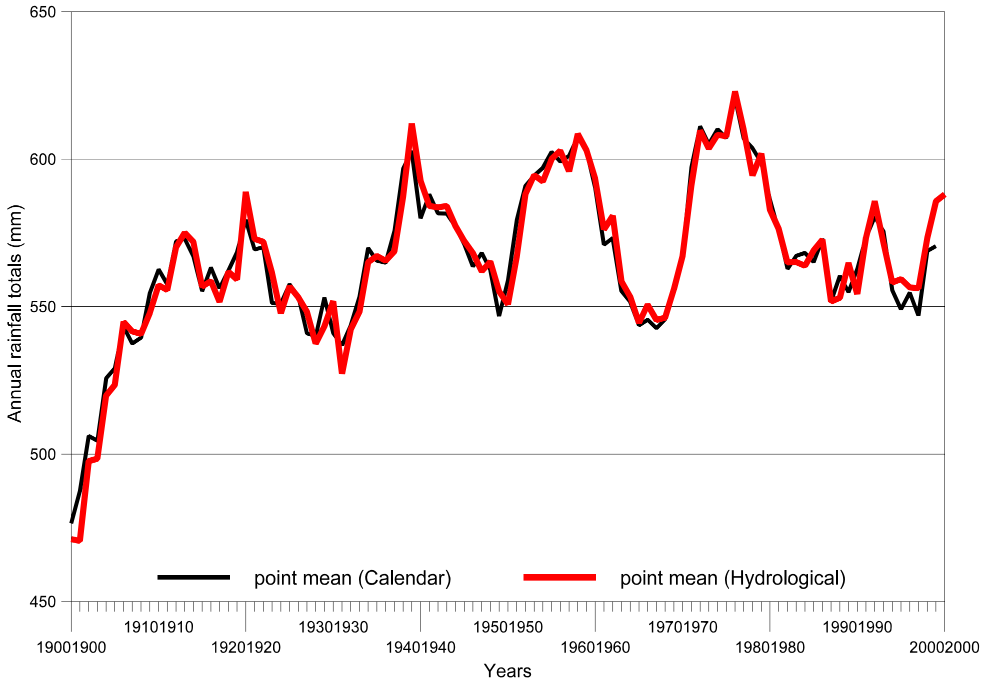


Figure 38 Smoothed annual rainfall totals calculated from point values using calendar and hydrological years

When comparing maps of rainfall for hydrological years and calendar years it is possible to see that even for the same timeframe there are differences within the areas that show up as low rainfall and high rainfall areas. This is due mainly to the hydrological year spanning from October to September. It is only in the southern and southeastern Cape province where there is an all year rainfall region. For the remainder of the country the rainfall is divided into areas of early summer rainfall which peaks over December, mid summer which peaks in January, late summer which is in February and very late summer rainfall which peaks from March to May.

The winter rainfall periods are from June to August, and only affect the Cape west coast. With most of the rainfall for the country occurring during the summer months, i.e. from December to February, the calendar year divides the season, with different parts contributing to different years' records. With the hydrological year being divided in the September/October months implies that the only area that will be divided is the region of all year rainfall. However, because this area has an even spread of rain throughout the year, the calendar year would also split the season. The hydrological year split also takes into account areas that have winter rainfall and it begins in spring after the winter rainfalls have ended and before the summer rainfalls begin. This allows for the wet season to be "contained" to a single study period. It is because of this split that the maps for the different time spans look different; when comparing rainfall data for a hydrological year and a calendar year, the second year in the hydrological year correlates better with the calendar year than the first part of the hydrological year.

The roughly 100 rasters of annual rainfall determined using calendar and hydrological years (see Appendix ***hydrological_yrs.pdf*** and ***calendar_yrs.pdf***) were examined using a raster overlay process to determine if any significant patterns exist over time and space. Each one arc minute pixel was analysed, over time, to determine the number of times that it was classified dry or wet (Figure 39).

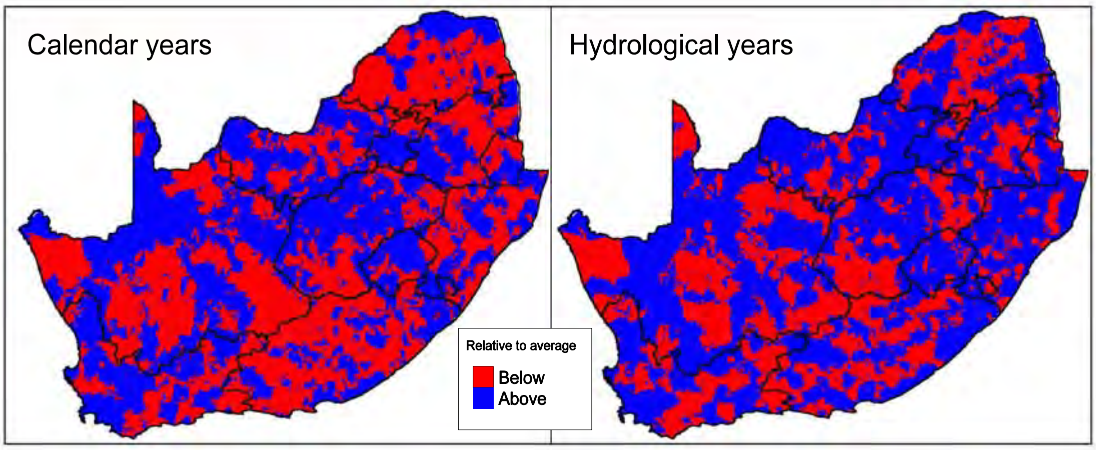


Figure 39 Dominant areas that have below (red) or above (blue) average rainfall over the past 100 years using calendar and hydrological year approaches

The calendar years (Figure 39) have 49% of the pixels occupying the dry state over time and the wet state being occupied 51% of the time. The scatter of red and blue pixels appear to be rather random and it is, therefore, difficult to decide if any pattern exists over time. The hydrological years, on the other hand, have 39% of the pixels occupying the dry state with 61% of the pixels called wet over time. The two maps depicted in Figure 39 have similar trends in certain regions, the Free State province being an example. KwaZulu-Natal, on the other hand, has many more dry classifications using the calendar year approach than when the annual totals are calculated using hydrological years.

Spatially, more than 65% of the time the pixel occupies the same state when the two maps (Figure 39) are subjected to a similarity test (Figure 40).

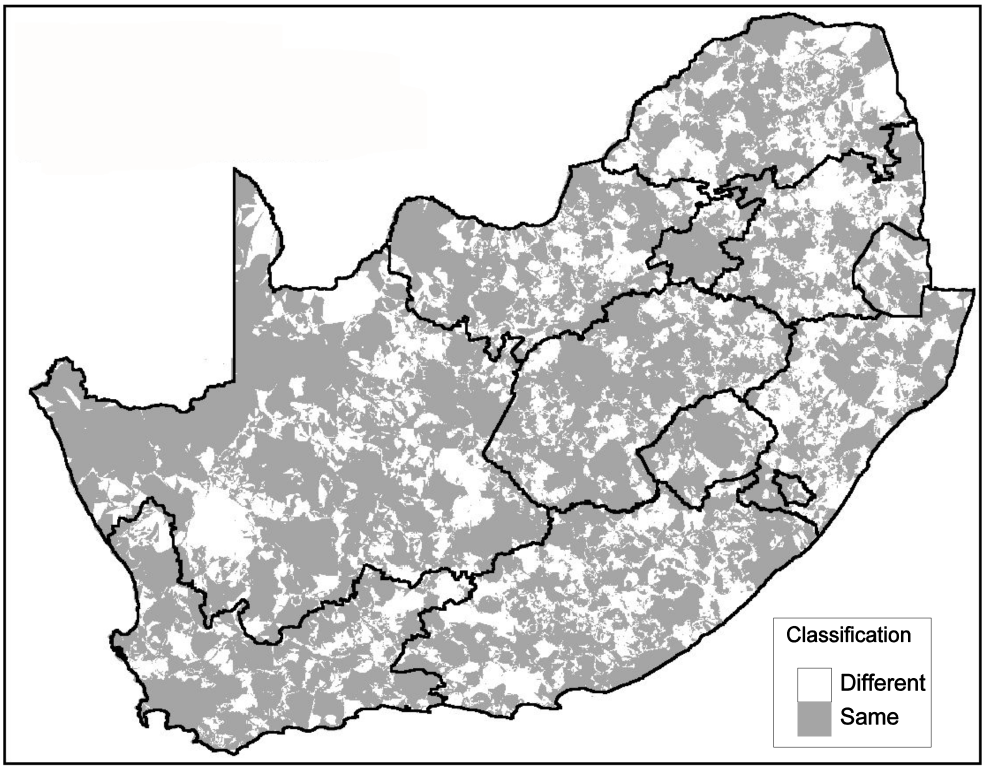


Figure 40 Areas where the calendar and hydrological year approach yield similar values averaged over the past 100 years

5.2.3.4 El Niňo and La Niňa Patterns for South Africa

The scope of this research project does not allow for a detailed investigation into the occurrences of El Niňo and La Niňa, but does warrant the inclusion of some preliminary thoughts as to whether there is a link with extreme rainfall in South Africa.

The Pacific ocean warm (El Niňo) and cold (La Niňa) episodes vary in their magnitude of Sea Surface Temperature (SST) anomalies. Since both the warm and cold episodes tend to reach their peak during the Northern Hemisphere winter, they tend to span two consecutive calendar years in South Africa. In some cases years are listed as both cold and warm episode years as there was a transition between the two extreme states during the course of the calendar year (Kousky, 2002). The periods when strong intensities of SST anomalies were observed are investigated to determine if there is a link between the anomalies and the annual totals of South African rainfall. Perusal of Table 17 suggests that there is no significant overlap between spatial rainfall extremes in South Africa and the occurrence of either El Niňo or La Niňa.

5.3 MONTHLY PRECIPITATION

Raster surfaces of monthly precipitation are calculated using the mean and the median statistics. The Dent *et al*. (1989) report only considered the surfaces of median monthly precipitation and this was due to a time and computing power constraint. The technique that is used to create these monthly precipitation surfaces in this project is similar to the technique used by Dent *et al*. (1989) which consists, *inter alia,* of expressing the median (or mean) monthly values as a ratio of the MAP values (done using the existing scatter of rain gauges). These ratios (Figure 41) are then interpolated onto a rectangular raster, at a spatial resolution of one arc minute. This interpolated raster is then multiplied by the raster of estimated MAP values generated using GWR (Figure 22) and this is repeated for each month. This methodology has been widely accepted as a means of creating monthly surfaces of rainfall without using a regression approach. The use of GWR to estimate monthly rainfall surfaces is not recommended because the monthly data are more variable than the MAP data and the process could be extremely time consuming and computer intensive.

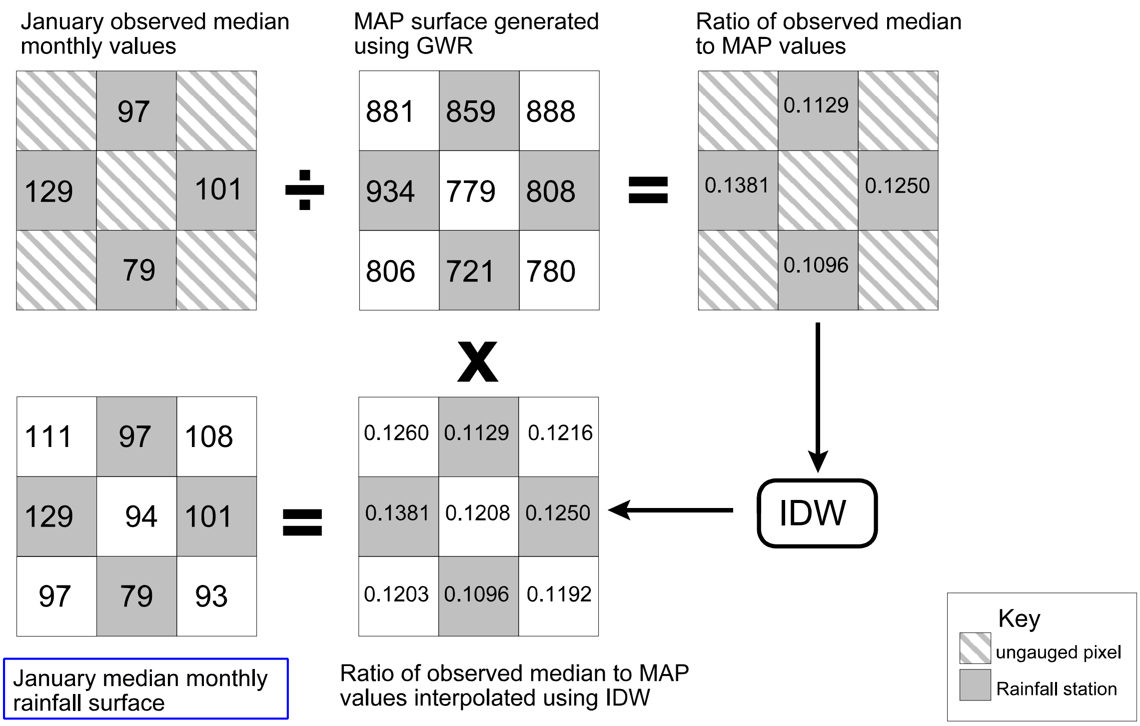


Figure 41 Procedure used to calculate the median (or mean) monthly rainfall surfaces

Table 17 Frequency of extreme annual rainfall totals calculated according to calendar and hydrological years and the occurrence of definitive El Niňo and La Niňa events (after Kousky, 2002)

Hydrological Year			El Niňo	La Niňa	Calendar Year		
Year	Driest Ranking	Wettest Ranking	Warm	Cold	Year	Driest Ranking	Wettest Ranking
1899/1900	13	70	weak		1900	56	40
1900/1901	52	42			1901	70	33
1901/1902	64	57			1902	39	58
1902/1903	3	93	weak		1903	1	100
1903/1904	47	65		strong	1904	35	59
1904/1905	48	79			1905	26	77
1905/1906	19	89	strong		1906	68	38
1906/1907	82	7		strong	1907	67	31
1907/1908	27	76		strong	1908	8	93
1908/1909	88	13			1909	93	5
1909/1910	60	52		strong	1910	34	74
1910/1911	73	31			1911	65	37
1911/1912	43	83	strong		1912	14	88
1912/1913	66	63			1913	54	57
1913/1914	16	58	moderate		1914	59	51
1914/1915	44	35			1915	38	48
1915/1916	12	74			1916	9	92
1916/1917	101	21		strong	1917	99	3
1917/1918	89	18		strong	1918	77	19
1918/1919	10	91	strong		1919	4	98
1919/1920	70	41			1920	73	29
1920/1921	72	27			1921	98	4
1921/1922	17	69			1922	17	84
1922/1923	80	44	moderate		1923	51	53
1923/1924	26	94			1924	53	44
1924/1925	95	2		moderate	1925	92	7
1925/1926	2	100	strong		1926	5	95
1926/1927	8	97			1927	6	97
1927/1928	36	67			1928	33	68
1928/1929	77	50		weak	1929	82	20
1929/1930	39	61			1930	16	86
1930/1931	45	80			1931	44	52
1931/1932	51	60	moderate		1932	18	83
1932/1933	1	101			1933	12	94
1933/1934	74	17			1934	84	14
1934/1935	56	26			1935	29	69
1935/1936	34	71			1936	78	27
1936/1937	69	46			1937	28	76

Hydrological Year			El Niňo	La Niňa	Calendar Year		
Year	Driest Ranking	Wettest Ranking	Warm	Cold	Year	Driest Ranking	Wettest Ranking
1937/1938	54	62			1938	60	42
1938/1939	85	4		strong	1939	87	9
1939/1940	63	37	strong		1940	72	26
1940/1941	57	40	strong		1941	27	63
1941/1942	35	85			1942	81	23
1942/1943	99	12			1943	91	12
1943/1944	86	15			1944	31	62
1944/1945	7	82			1945	7	91
1945/1946	24	77			1946	43	60
1946/1947	18	96	moderate		1947	40	72
1947/1948	68	36			1948	47	65
1948/1949	4	92			1949	15	79
1949/1950	78	8		strong	1950	64	32
1950/1951	55	55		strong	1951	20	82
1951/1952	22	90			1952	52	50
1952/1953	62	45			1953	71	34
1953/1954	92	24			1954	55	43
1954/1955	79	32		strong	1955	90	10
1955/1956	90	28		strong	1956	95	11
1956/1957	65	10			1957	83	21
1957/1958	49	51	strong		1958	58	41
1958/1959	61	47	strong		1959	49	54
1959/1960	28	66			1960	50	49
1960/1961	98	5			1961	79	28
1961/1962	15	64			1962	32	71
1962/1963	76	16			1963	69	25
1963/1964	25	59			1964	24	75
1964/1965	38	87		moderate	1965	13	89
1965/1966	5	88	moderate		1966	3	99
1966/1967	84	11			1967	94	8
1967/1968	14	78			1968	21	81
1968/1969	40	72	moderate		1969	41	64
1969/1970	11	98	moderate		1970	11	90
1970/1971	71	33		moderate	1971	74	30
1971/1972	58	29			1972	48	56
1972/1973	23	84	strong		1973	66	36
1973/1974	96	3		strong	1974	96	6
1974/1975	87	23		strong	1975	89	13
1975/1976	97	1		strong	1976	100	1
1976/1977	94	14	weak		1977	88	15
1977/1978	42	38	weak		1978	45	47
1978/1979	33	56			1979	30	73
Hydrological Year			El Niňo	La Niňa	Calendar Year		

Year	Driest Ranking	Wettest Ranking	Warm	Cold	Year	Driest Ranking	Wettest Ranking
1979/1980	37	68	weak		1980	42	66
1980/1981	93	20			1981	80	24
1981/1982	31	81			1982	22	85
1982/1983	9	75	strong		1983	36	70
1983/1984	32	43		weak	1984	10	87
1984/1985	30	48		weak	1985	62	35
1985/1986	50	30			1986	37	61
1986/1987	53	34	moderate		1987	46	55
1987/1988	75	19	moderate		1988	75	17
1988/1989	67	22		strong	1989	85	16
1989/1990	83	54			1990	23	78
1990/1991	41	49	strong		1991	76	22
1991/1992	6	86	strong		1992	2	96
1992/1993	21	53	strong		1993	61	45
1993/1994	59	39			1994	19	80
1994/1995	20	99	moderate		1995	86	18
1995/1996	91	6		weak	1996	97	2
1996/1997	100	9			1997	63	39
1997/1998	29	95	strong		1998	57	46
1998/1999	46	73		strong	1999	25	67
1999/2000	81	25			2000		

5.3.1 Median Monthly Precipitation

The CV (%) of the median monthly rainfall values calculated per province (Table 18) exhibits some very large values for the Northwest province for the months June, July and August. The reason for this is that the Northwest province consists of 37 788 one arc minute pixels and 37 730 of these pixels have a zero median rainfall value for July and the remaining 58 pixels have median rainfall values varying between 1 and 4 mm. The average July median monthly rainfall for the Northwest province is 0.00172 mm and the standard deviation is 0.048 which yields a CV of roughly 2 788% (Table 18). The median values analysed per province are skewed in the non-rainy months and the users of this information should realise this. The complete set of the 12 median monthly raster images are presented pictorially in the Appendix.

5.3.2 Mean Monthly Precipitation

The average of the mean monthly precipitation values analysed per province (Table 18) do not exhibit such extremes as in the average of the median monthly precipitation analysis. This can be attributed to the fact that the median values, i.e. what happens 50% of the time, has less smoothing than the mean values. The complete set of the 12 mean monthly raster images are presented pictorially in Appendix ***mean and median monthly rfl.pdf***.

Table 18 Frequency analysis of the median and mean monthly raster rainfall

Month and Province/ Country	Median Monthly Rainfall					Mean Monthly Rainfall				
	mean (mm)	CV (%)	max (mm)	min (mm)	median (mm)	mean (mm)	CV (%)	max (mm)	min (mm)	median (mm)
JANUARY										
Limpopo	87	29	308	23	84	100	29	351	28	95
Mpumalanga	111	21	268	54	107	122	22	309	63	115
North-West	71	29	180	20	72	82	24	176	28	83
Northern Cape	14	100	73	0	10	22	72	85	0	19
Gauteng	105	8	157	51	106	114	9	166	51	114
Swaziland	119	21	217	65	112	137	20	246	77	131
Free State	72	32	195	21	71	79	28	198	27	79
KwaZulu-Natal	119	23	299	35	117	129	20	317	38	126
Lesotho	106	21	283	46	101	114	19	292	62	110
Eastern Cape	57	63	204	3	45	64	57	207	5	52
Western Cape	10	116	111	0	7	15	82	119	0	13
FEBRUARY										
Limpopo	72	34	307	19	68	86	33	349	22	81
Mpumalanga	89	25	265	46	82	101	27	302	56	93
North-West	64	19	117	20	65	73	18	124	26	73
Northern Cape	22	78	81	0	19	30	57	88	0	28
Gauteng	82	8	130	50	82	91	9	137	55	93
Swaziland	103	23	194	51	98	119	21	208	61	113
Free State	66	20	167	26	66	74	19	187	28	73
KwaZulu-Natal	107	21	282	31	103	118	20	320	35	115
Lesotho	96	19	259	56	94	104	19	297	60	102
Eastern Cape	62	52	199	6	55	69	48	210	10	61
Western Cape	12	96	108	0	9	19	72	122	1	16
MARCH										
Limpopo	54	34	229	12	52	64	33	270	16	60
Mpumalanga	74	24	206	34	69	81	24	232	39	75
North-West	61	18	149	19	62	70	14	152	26	71
Northern Cape	28	59	81	0	26	36	50	91	0	36
Gauteng	71	10	111	42	72	79	10	128	45	80
Swaziland	83	24	156	43	77	93	22	171	49	87
Free State	66	14	134	31	66	73	13	151	40	72
KwaZulu-Natal	91	22	249	30	88	101	22	271	33	98
Lesotho	85	18	234	46	87	93	17	254	48	95
Eastern Cape	66	41	193	14	63	74	38	204	16	70
Western Cape	20	73	127	0	17	25	64	132	1	22

Month and Province/ Country	Median Monthly Rainfall					Mean Monthly Rainfall				
	mean	CV	max	min	median	mean	CV	max	min	median
	(mm)	(%)	(mm)	(mm)	(mm)	(mm)	(%)	(mm)	(mm)	(mm)
APRIL										
Limpopo	23	34	106	4	22	31	28	112	6	30
Mpumalanga	34	25	94	14	32	41	22	108	20	39
North-West	27	17	52	4	27	35	15	59	14	35
Northern Cape	14	51	40	0	13	21	35	48	0	21
Gauteng	32	12	55	18	32	39	9	65	22	39
Swaziland	41	18	66	21	41	49	17	75	25	48
Free State	34	16	65	14	34	41	13	67	18	40
KwaZulu-Natal	42	25	104	14	41	50	25	118	17	47
Lesotho	42	22	84	20	43	48	20	99	25	49
Eastern Cape	34	34	90	8	34	41	34	105	10	41
Western Cape	21	64	211	1	18	28	55	233	1	23
MAY										
Limpopo	5	63	34	0	4	11	34	51	1	10
Mpumalanga	10	33	31	3	9	16	23	40	5	15
North-West	6	38	15	0	6	14	18	26	4	14
Northern Cape	7	74	48	0	6	13	45	61	0	12
Gauteng	8	22	15	2	8	16	13	25	4	16
Swaziland	14	22	26	6	13	22	16	34	11	21
Free State	12	28	28	0	11	18	18	34	3	18
KwaZulu-Natal	18	51	80	0	16	27	43	121	5	24
Lesotho	19	26	32	6	20	25	20	39	11	25
Eastern Cape	18	47	82	4	17	27	40	106	5	24
Western Cape	29	85	314	1	22	35	75	376	1	29
JUNE										
Limpopo	1	261	22	0	0	4	55	32	0	4
Mpumalanga	1	107	11	0	1	7	32	22	0	6
North-West	0	556	3	0	0	5	20	14	0	5
Northern Cape	5	160	62	0	1	9	102	79	0	6
Gauteng	0	221	10	0	0	5	19	16	1	6
Swaziland	4	29	9	1	4	12	16	19	6	12
Free State	2	74	14	0	2	7	26	17	0	7
KwaZulu-Natal	7	94	62	0	5	16	55	76	2	13
Lesotho	5	38	13	1	5	11	22	22	4	11
Eastern Cape	10	84	88	0	8	17	61	97	2	14
Western Cape	33	97	441	0	25	38	90	444	1	30

Month and Province/ Country	Median Monthly Rainfall					Mean Monthly Rainfall				
	mean	CV	max	min	median	mean	CV	max	min	median
	(mm)	(%)	(mm)	(mm)	(mm)	(mm)	(%)	(mm)	(mm)	(mm)
JULY										
Limpopo	1	242	20	0	0	4	68	37	0	3
Mpumalanga	1	144	14	0	0	8	36	25	0	7
North-West	0	2788	4	0	0	4	31	8	0	3
Northern Cape	4	180	62	0	0	8	101	73	0	5
Gauteng	0	1180	2	0	0	5	19	11	0	5
Swaziland	4	32	9	0	4	12	16	20	6	11
Free State	1	112	9	0	1	8	24	24	1	7
KwaZulu-Natal	7	91	53	0	5	16	51	68	2	13
Lesotho	4	32	11	1	4	11	17	19	5	11
Eastern Cape	9	91	87	0	7	17	61	98	3	14
Western Cape	30	98	470	0	23	35	87	432	1	27
AUGUST										
Limpopo	0	296	27	0	0	4	69	38	0	3
Mpumalanga	3	79	22	0	3	9	37	30	1	9
North-West	0	390	5	0	0	5	30	13	1	5
Northern Cape	4	190	60	0	0	8	92	66	0	6
Gauteng	1	97	10	0	1	7	19	19	3	7
Swaziland	7	39	19	0	6	15	21	29	8	15
Free State	4	58	14	0	3	11	29	29	3	10
KwaZulu-Natal	15	53	63	0	13	24	35	75	8	22
Lesotho	10	19	28	5	10	18	15	45	10	18
Eastern Cape	14	80	106	1	11	22	57	120	4	19
Western Cape	30	93	444	0	23	35	81	491	1	28
SEPTEMBER										
Limpopo	5	76	46	0	4	12	41	63	2	11
Mpumalanga	17	39	57	3	16	25	29	63	7	24
North-West	3	83	16	0	3	11	31	24	3	11
Northern Cape	2	151	31	0	0	7	61	39	0	6
Gauteng	11	19	28	4	12	19	17	33	6	19
Swaziland	28	26	54	12	27	35	19	61	21	35
Free State	10	56	40	0	9	19	32	53	7	18
KwaZulu-Natal	33	35	93	11	30	45	30	118	14	41
Lesotho	20	21	63	8	20	31	15	75	16	31
Eastern Cape	23	67	99	2	20	34	54	119	4	30
Western Cape	19	92	310	0	14	25	76	342	1	20

Month and Province/ Country	Median Monthly Rainfall					Mean Monthly Rainfall				
	mean	CV	max	min	median	mean	CV	max	min	median
	(mm)	(%)	(mm)	(mm)	(mm)	(mm)	(%)	(mm)	(mm)	(mm)
OCTOBER										
Limpopo	30	31	103	5	29	36	27	122	6	36
Mpumalanga	60	27	134	24	61	66	25	138	27	67
North-West	27	39	67	7	28	33	32	76	9	34
Northern Cape	7	74	27	0	6	11	46	35	0	10
Gauteng	53	11	103	31	54	60	10	104	37	60
Swaziland	68	24	125	36	67	75	22	130	40	73
Free State	40	38	100	9	38	47	34	108	15	45
KwaZulu-Natal	72	21	156	21	70	80	20	164	24	78
Lesotho	59	18	138	32	57	67	16	159	38	65
Eastern Cape	41	56	143	5	36	49	51	165	6	44
Western Cape	19	82	175	1	13	24	70	188	1	18
NOVEMBER										
Limpopo	66	25	176	16	63	72	23	195	17	69
Mpumalanga	101	21	207	38	102	107	19	209	45	106
North-West	48	42	118	11	46	54	36	122	15	53
Northern Cape	9	83	51	0	7	15	55	56	0	13
Gauteng	93	10	129	49	94	99	9	134	50	100
Swaziland	102	26	175	56	99	109	22	176	64	108
Free State	56	37	137	15	55	63	34	144	20	62
KwaZulu-Natal	96	19	226	28	94	103	17	231	31	102
Lesotho	80	19	238	42	77	86	17	238	48	84
Eastern Cape	50	57	167	4	44	58	51	186	7	52
Western Cape	15	85	143	0	12	21	67	172	1	18
DECEMBER										
Limpopo	82	25	267	20	81	90	25	291	23	89
Mpumalanga	110	21	236	51	106	116	21	264	63	110
North-West	61	35	123	16	57	67	32	125	20	63
Northern Cape	10	109	57	0	6	17	67	65	0	14
Gauteng	99	8	148	34	99	103	8	156	35	104
Swaziland	116	22	201	58	112	124	20	208	71	120
Free State	62	38	168	15	59	67	34	171	22	65
KwaZulu-Natal	111	21	258	30	109	118	19	271	33	115
Lesotho	94	19	239	51	91	99	18	251	53	95
Eastern Cape	53	62	195	4	42	61	54	197	5	50
Western Cape	10	108	119	0	8	17	71	139	0	14

5.4 DAILY RAINFALL

5.4.1 The Variability of Daily Rainfall

The temporal and spatial patterns of daily rainfall are much more variable than those of monthly rainfall, and these are again more variable than those exhibited by the annual totals of rainfall. The data from three rainfall stations (Figure 42) are selected to illustrate these trends. The Royal Observatory station has the longest rainfall record in South Africa, Roberstvlei records of the highest rainfall and Vioolsdrift, on the other hand, records of the lowest rainfall.

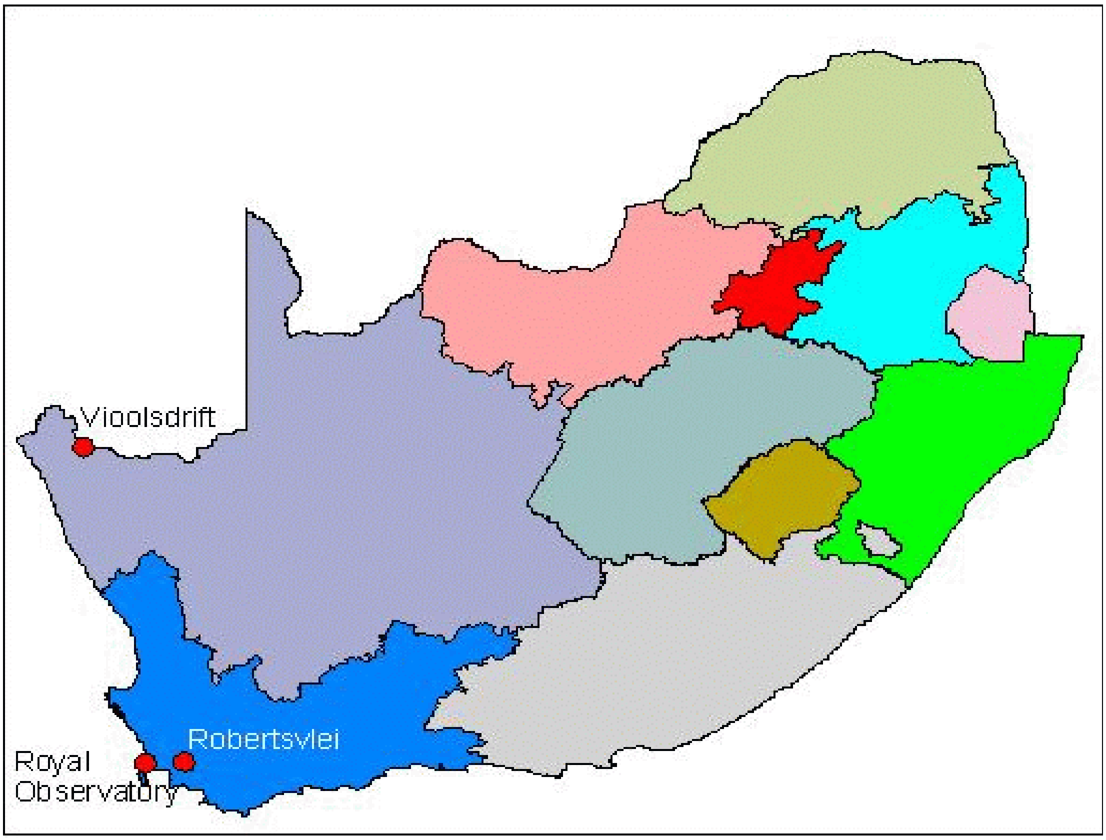


Figure 42 Location of the 3 rainfall stations that have of the highest, longest and lowest rainfall annual totals

Roberstvlei recorded its lowest annual precipitation of 1 293.7 mm in 1974, the median annual precipitation of 2 043.0 mm was achieved in 1990 and the highest annual precipitation of 3 222.2 mm was produced in 1978. The Royal Observatory recorded its lowest annual precipitation of 418.9 mm in 1930, the median annual precipitation of 609.6 mm was achieved in 1913 and the highest annual precipitation of 1 037.7 mm was produced in 1892. Vioolsdrift recorded its lowest annual precipitation of 1.7 mm in 1979, the median annual precipitation of 41.4 mm was achieved in 1988 and the highest annual precipitation of 128.1 mm was produced in 1997. The summary of the statistics for these three stations (Table 19) highlights the fact that the temporal variation of daily rainfall is much more variable than those of monthly and annual rainfall totals.

Table 19 Comparative rainfall statistics at the three selected rainfall sites

Statistic	Time-step	Robertsvlei	Royal Observatory	Vioolsdrift
Sample size	Daily	13 460	51 470	16 675
	Monthly	431	1 531	546
	Annual	30	113	43
Minimum (mm)	Daily	0.0	0.0	0.0
	Monthly	0.0	0.0	0.0
	Annual	1 293.7	418.9	1.7
Maximum (mm)	Daily	440.0	99.0	80.0
	Monthly	975.1	344.8	107.7
	Annual	3 222.2	1 037.7	128.1
Mean (mm)	Daily	5.5	1.7	0.1
	Monthly	166.1	50.4	4.2
	Annual	2 087.9	618.8	46.6
Median (mm)	Daily	0.0	0.0	0.0
	Monthly	105.0	36.9	0.0
	Annual	2 043.0	609.6	41.4
Standard Deviation (mm)	Daily	18.1	5.1	1.6
	Monthly	173.9	46.8	10.1
	Annual	437.3	123.0	31.8
Coefficient of Variation (CV%)	Daily	328.3	303.6	1 158.7
	Monthly	104.7	92.9	238.9
	Annual	20.9	19.9	68.2

One of the initial aims of this project was to produce rasters of daily rainfall at a spatial resolution of one arc minute. One of these rasters of daily rainfall values takes up roughly 0.5 MB of hard disk space, which means that 365 rasters occupies about 182 MB. This will imply that a CD-ROM will hold approximately 3.5 years of daily rasters. An example of 366 rasters, with their point shape files, for 1980 is included on the accompanying CD-ROM for perusal. The rasters were created using IDW. A test was performed at the one arc minute pixel, represented by station 0152482 W, and the 366 pixels produce an annual total of 781.11 mm whereas the observed total for 1980 is 781.6 mm. This slight difference can be attributed to rounding errors and the IDW approach.

5.4.2 Issue of Spatial Scale

The present output from some of the General Circulation Models (GCM) is at a spatial scale of 3.75 x 2.50 degrees. The question that is often asked is, can this GCM information be down-scaled to be usable at a pixel size of 15 x 15 arc minutes?

An analysis was performed on the daily rainfall database to determine which day produced a large variation in rainfall over South Africa and 29 September 1987 was found to include zero and large rainfall amounts (Table 20). Heavy rainfall was concentrated in the KwaZulu-Natal area and this area was subsequently declared a disaster area.

Table 20 Descriptive statistics of the daily rainfall over South Africa on 29 September 1987

Sample size (stations)	3 512
Min rainfall (mm)	0
Max rainfall (mm)	576
Mean rainfall (mm)	29
Median rainfall (mm)	1
CV (%) rainfall	225

All the daily rainfall values, for each 3.75 x 2.5 degree GCM pixel, were averaged and this average value was assigned to the GCM pixel. A similar process was followed to determine the average daily rainfall, for 29 September 1987, for each block describing the 1:250 000 and 1:50 000 topographic maps (2x1 degree and 0.25x0.25 degrees respectively). Finally, the daily rainfall, at the rain gauge positions, was used to produce a Thiessen polygon maps representing the status of rainfall on this particular day (Figure 43).

Perusal of Table 21 reveals the following with regard to the amount of rainfall that occurred spatially over South Africa on 29 September 1987 and these include:

- the pixel averaging procedures yield substantially less area with no rainfall, and
- the four techniques yield similar percentage areas regarding the high rainfal amounts, i.e. greater than 50 mm.

No pixels with missing rainfall are generated when using the GCM, 1:250 000 and Thiessen procedures, however, there are many 1:50 000 blocks that do not have any rain gauges and this results in large areas with no rainfall values.

Table 21 Percentage of South Africa receiving different amounts of rainfall on 29 September 1987

Rainfall (mm)	GCM	1:250 000	1:50 000	Thiessen
	% of South Africa receiving rainfall on 29 September 1987			
0	47.8	52.7	53.3	62.0
0-10	27.3	24.2	18.6	14.3
10-30	7.9	3.6	11.1	9.0
30-50	9.5	9.9	4.6	4.9
50-100	0.0	5.0	6.0	5.3
100-200	7.4	1.5	4.2	2.7
>200	0.1	3.1	2.2	1.8

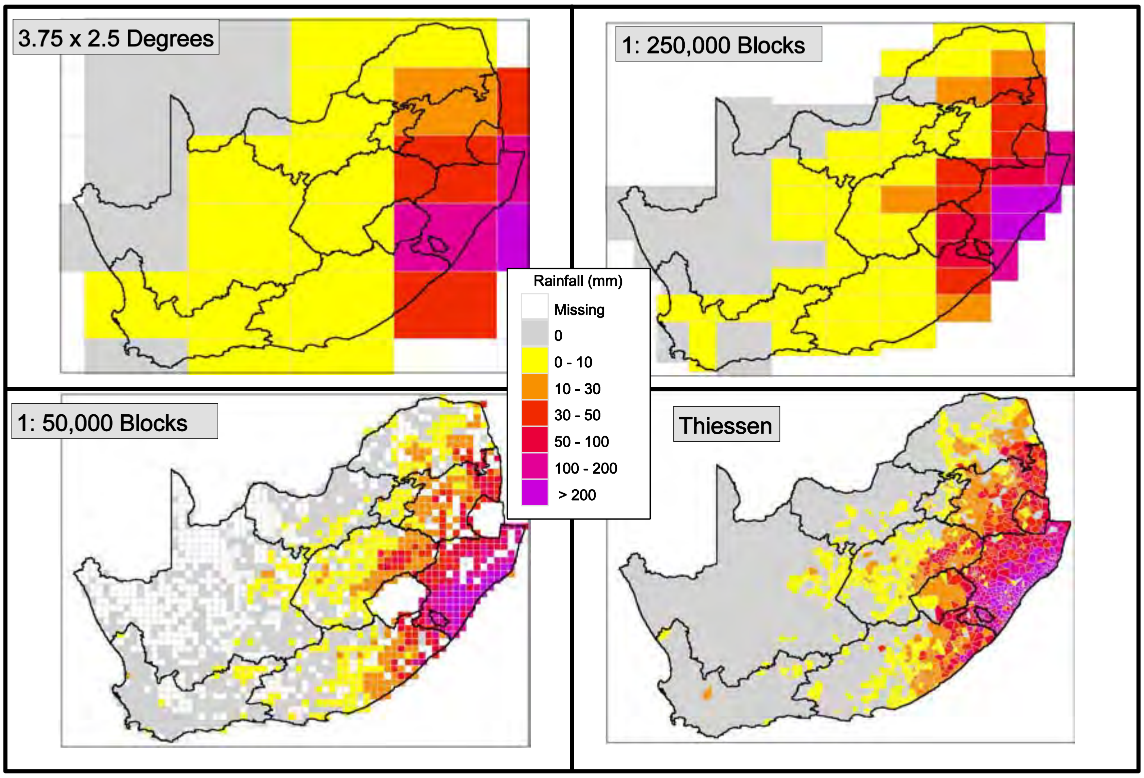


Figure 43 Spatial distribution of rainfall over South Africa using the four “averaging” techniques

CHAPTER 6

RAINFALL DATA MANIPULATION UTILITIES

The volume of the daily rainfall data used in this project made it impossible to use an off-the-shelf database or spreadsheet package. The fact that computer programs need to access and interrogate these large databases, in excess of 300 million records, resulted in the design and use of a very simple, but effective, ASCII file that is controlled by an index-file containing the position of the start and end records of each daily rainfall station in the large direct access file.

The majority of the computer programs used to create, analyse, interrogate and manipulate the rainfall data were developed in-house and Fortran90 is the preferred programming language.

6.1 DESCRIPTION OF THE DATABASE STRUCTURE

The daily rainfall database consists of more that 300 million rainfall values for roughly 14 000 stations. The data in this database originate from many different organisations and individuals, each having their own structure and set of quality control codes. The author has an extensive library of Fortran based routines to analyse time series information. A flat direct access ASCII data structure is not the most economical format, but it allows for quick access to any record in the database. A suite of programs was developed to assist the user in extracting rainfall data and storing it in a number of common data formats.

6.2 DAILY RAINFALL DATABASE

The rainfall data are stored in a format that was initially used by the SA Weather Bureau (now the South African Weather Service) and which has been enhanced to suit the requirements of this project. The data are in a flat ASCII file with a record length of 171 characters, *viz*.

1-9	Rainfall station identifier (Character)
11-14	Year (Integer)
15-16	Month (Integer)
17-171	31 Fields of daily rainfall (Character)

Each month has 31 values, irrespective of the number of days in the month, and these 31 fields each consist of an alphanumeric value of 5 characters, *viz*.

- if the first character is non-numeric, then the remaining four characters are converted to integer and this is then the rainfall amount in tenths of a mm derived according to the code displayed in Table 22;
- if the first character is numeric then the rainfall amount (in tenths of a mm) is retrieved as follows;
- if e.g. the value is 90123 then it implies that 90000 has been added to the amount
- and the value of 12.3 mm cannot be used as it forms part of an accumulation period (Table 22);

- if e.g. the value is 01234 then it implies that it is a valid observed value of 123.4 mm.

The ASCII datafile acts as a direct access file via an "*index*" file. The structure of the index file is as follows;

1-9	Rainfall station identifier (Character)
11-20	Position, in the direct access file, where the above mentioned rainfall stations data starts.

The next record then points to the start location of that rainfall stations data. In other words, the data for a station i is located between records at position i and position (i+1)-1.

The quality code of each rainfall values means that the user of the database can interrogate the values and include or exclude infilled data if they so wish. If the user would prefer to use a different infilling technique they would extract all the observed data and then infill the missing rainfall amounts.

6.3 MONTHLY RAINFALL DATABASE

There are number of stations that only have monthly data available. All the daily rainfall values are converted into monthly summaries to allow the user fast access to a number of monthly and annual statistics pertaining to each station. These data are maintained in a direct access ASCII file of record length 96 characters. The structure of this file is as follows:

- The first record contains the alphanumeric station identifier (columns 1 through 9).
- The next section, depending on the number of years of available record, contains for each year the year followed by 12 monthly totals and the annual total at the end of each record. The quality codes that are used are described in Table 23.
- This section is followed by 9 records that contain the minimum, maximum, mean, media, standard deviation, skewness, kurtosis, coefficient of variation and the sample size in years.

These statistics are only calculated for complete months and for complete annual totals, i.e. if the station has a missing day it implies that the data for that month and year are excluded from the statistical calculations.

The ASCII datafile acts as a direct access file via an "*index*" file. The structure of the index file is as follows:

1-9	Rainfall station identifier (Character)
11-20	Position, in the direct access file, where the above mentioned rainfall stations data starts
22-28	the longitude, in degrees decimal, co-ordinate of the station
30-37	the latitude, in negative degrees decimal, co-ordinate of the station.

Table 22 Quality codes associated with the daily rainfall data

Code	OK	Explanation
	✓	rainfall recorded, no problems
00000	✓	drops, no measurable amount for day
-9999	✗	station open, but rainfall not measured
-8888	✗	part of an accumulation period, rained on day but unknown amount
-7777	✓	zero rainfall
-6666	✗	part of an accumulation period, not known if it rained on this day
-5552	✗	the station has 3 and more years where the annual total equals zero
-5553	✗	all the unexplained rainfall codes are replaced by this single code
-5555	✗	padding value (similar to -9999)
10000+	✗	rainfall reported orally, not reliable 0139228W 15/1/1932 10414
20000+	✗	no rainfall recorded, but most stations in the vicinity recorded reasonably large amounts 0004602AW 22/1/1987 20000
30000+	✗	it rained on this day, no accumulation total, the amount is unreliable 0002069W 30/4/1952 30005
40000+	✗	its not known if it rained on this day, an accumulation total was recorded, but is unreliable 0005730W 19/12/1982 40770
50000+	✗	it did not rain on this day, an accumulation total was recorded, but is unreliable 0174600AW 26/8/1979 50112
60000+	✗	it did rain on this day, an accumulation total was recorded, the amount is unreliable 0041713W 17/4/1983 60135
70000+	✓	it did not rain on this day, an accumulation total was recorded, the amount is reliable 0174600AW 16/9/1979 70083
80000+	✗	it rained on this day, an accumulation total was recorded, the amount is reliable
90000+	✗	its not known if it rained on this day, an accumulation total was recorded, the amount is reliable
-777	✓	a,b,c, d or z -7777 (zero rainfall value)
axxxx	✓	ARC data
bxxxx	✓	infilled using EMA (Smithers & Schulze, 2000)
cxxxx	✓	infilled using IDW (Meier, 1997)
dxxxx	✓	Monthly patched (Total $\leq$2 mm) (Dent *et al.*, 1989)
fxxxx	✓	Median, MAP ratio
zxxxx	✓	Zimbabwe data
e-999	✗	daily rainfall value exceeds 600 mm on that day

The next record then points to the start location of that rainfall stations data. In other words, the data for a station i is located between records at position i and position (i+1)-1.

Table 23 Quality codes associated with the monthly rainfall data

Code	OK	Explanation
	✓	No problem (monthly)
		Usable monthly
		Usable annual
+	✗	incomplete annual total
&	✗	Suspected error in data since there are more than 5 months of zero rainfall during wettest period of the year
Q	✗	MAP inconsistent with neighbouring stations or topography
M	✗	at least one daily value missing in the month
-99.9M	✗	all daily values for the month are missing
$	✓	contains daily data which were coded as accumulated data (total reliable) by the SAWB
*	✗	contains daily data which were coded as accumulated data (total unreliable) by the SAWB
O	✗	monthly total exceeds 999.9 mm
P	✓	synthesised value using the Zucchini technique (Model 5) as reported by Dent *et al*. (1989)
X	✗	mean monthly value for the month is substituted since the month has missing data and no nearby station is suitable for patching
#	✗	the monthly total contains one or more daily values which were coded 30 000+ in the SAWB codes
a	✓	one or more daily value originated from the ARC restricted data
b	✓	one or more daily value originated from the Smithers & Schulze (2000) patched data
c	✓	one or more daily value originated from the Meier (1997) patched data
d	✓	one or more daily value ($\leq$2 mm) originated from the Zucchini Model 5 data
N	✓	from the monthly Namibian data file
Z	✓	daily codes of a and/or b and/or c and/or d are present in one month
-999.9	✗	no data are available for the month (not even missing values, i.e. padding)
-99.9A	✗	Station has $\geq$3 years of zero rainfall
I	✓	from the monthly ICFR data file
J	✓	from the 2001 appeal for public stations
z	✓	Zimbabwe rainfall data

6.4 DESCRIPTION OF THE SOFTWARE

The majority of the software was developed using Fortran90 and a list and description of the key routines is outlined below. The source code will be contained on the compact discs containing the rainfall databases and extraction routines.

- ***br_sawb.f90*** is a routine to extract daily rainfall data, using different search criteria, from the direct access data file (rainfall.dat) and report the extracted rainfall values in user selectable formats.
- ***br_astats.f90*** is a routine to extract monthly rainfall information, using different search criteria, from the direct access data file (astats.mon).
- ***extract.f90*** extracts a host of different attributes from the monthly rainfall information file. For example, extracts to a single file the station id and the MAP, creates two

columns of data, one being the observed annual total and the other the annual total if infilled data appear for that particular year.

- ***sel_rstn.f90*** uses the standard file (file.stn) that contains the station ids to produce a file that can be pasted into an ArcView search dialogue.
- ***top10.f90*** interrogates the monthly rainfall data file to determine *inter alia* which stations have the highest MAP, CV, etc. These different maxima and minima are assigned a numeric score and the stations are ranked according to these scores.
- ***dtop10.f90*** is similar to *top10.f90* except that the daily rainfall file is used.
- ***cr_sawb.f90*** creates a SAWB 9-character station id based on the co-ordinates of the rainfall station.
- ***get_names.f90*** uses the standard file (file.stn) that contains the station ids to extract the name of the station, where available, and to create files that allow the user to import the station details into ArcView.
- ***best5.f90*** determines which are the 5 best neighbours given the station id of a rainfall station.
- ***dsurface_rainfall.f90*** uses the daily rainfall data file (rainfall.dat) and extract to a file all the usable daily rainfall data for a particular day. This routine is used to create a point shapefile of daily rainfall amounts for each day that can then be imported into ArcView and interpolated onto a rectangular raster to produce daily rainfall surfaces.
- ***analyse_rainfall.f90*** uses the daily rainfall data file (rainfall.dat) and tabulates useful information, *viz.* number of days with zero rainfall, number of infilled rainfall records.
- ***big.for*** uses the daily rainfall data file (rainfall.dat) and creates the monthly rainfall data file (astats.mon). All the monthly totals are flagged according to the quality of the daily rainfall data.
- ***com_astats.f90*** is used to create the final monthly rainfall data file. It combines all the monthly rainfall data files into one file.
- ***rcomb_rainfall.f90*** calls *comb_rainfall* a number of times to create the final daily rainfall data file. The multiple execution of *comb_rainfall.f90* is used to overcome a memory leak problem in Microsoft Fortran PowerStation 4 that is used by the author.
- ***eliminate.f90*** uses information generated by *extract.f90* to eliminate the suspect rainfall data from the analysis.
- ***pickup.f90*** assigns the explanatory variable information to the rainfall stations.
- ***for_jcs.f90*** appends additional locational information onto the selected rainfall stations data.
- ***dup_stns.f90*** eliminates the duplicate rainfall stations used a maximum length of record criteria. There is often more than one rainfall station representing the same pixel as all the stations co-ordinates are rounded to the nearest one arc minute.

6.5 DESCRIPTION OF THE MOST COMMONLY USED DATA FILES

The majority of the data files used in this research project are flat ASCII files. The two direct access data files (rainfall.dat and astats.mon) can also be treated as flat ASCII files if necessary and would mean that access times would be slow as a result of the file sizes.

- ***rainfall.dat*** contains the daily rainfall database. See the beginning of this chapter for a detailed description.
- ***rainfall.idx*** contains the positions of the rainfall stations data in the direct access file (rainfall.dat). See the beginning of this chapter for a detailed description.

- ***file.stn*** is a flat ASCII file containing rainfall station ids. One 9-character station ids per record.
- ***crossref.all*** contains the co-ordinates and names of station, where available, of all the daily and monthly stations used in this research project. The format of the file is as follows; columns 1-9 = station id, columns 11-14 = station latitude (positive arc minutes of a degree), columns 16-19 = station longitude (arc minutes of a degree), columns 35-59 = station name (where available). The columns not mentioned here may contain information that were collected by other organisations and individuals and the author does not guarantee their correctness.
- ***astats.mon*** contains the monthly rainfall database. See the beginning of this chapter for a detailed description
- ***astats.idx*** contains the positions of the rainfall stations data in the direct access file (astats.mon). See the beginning of this chapter for a detailed description.
- ***output.br*** output from the daily extraction routine (*br_astats.f90*) and the format is dependent on what option the user selected.
- ***astats.br*** output from the monthly rainfall extraction routine (*br_astats.f90*). The format of this file is identical to that of the monthly rainfall data file (astats.mon). This is to allow subsets of the large data file to be extracted and analysed.

CHAPTER 7

RECOMMENDATIONS FOR FUTURE RESEARCH

Successful and effective research into, and use of, rainfall data depends largely on all the data being in a single database. There are many reasons for this. For example, if rainfall data were sought and acquired from different organisations, the data would first have to be re-formatted to a single database standard or form and the missing data would then need to be infilled. A single database, consisting of data that are collated from a host of different organisations and individuals, saves an enormous amount of valuable research time. A single database also allows the infilling procedures, for example, to make use of the best available temporal and spatial distribution of rainfall data. A plea is therefore made that there always be an organisation with a dedicated and interested individual for the collation and creation of an updated, quality controlled rainfall database.

The research into different interpolation and regression techniques to map annual rainfall have yielded success in the form of the Geographically Weighted Regression (GWR) technique (Brunsdon *et al*., 1996). The spatial interpolation and regression of daily rainfall surfaces, however, is still fraught with problems. The main factors are the volume of data compared to twelve monthly and a single MAP surface and the spatial variability on a day-by-day basis, both of which contribute substantially to the problem.

Lynch *et al*. (2001a) noted an increase in MAP from 1980 onwards for the rain gauges on the Irwin farm in an area south of Potchefstroom. Initially, this was regarded as an anomaly in the observations, but scrutiny of the original handwritten records and telephonic contact with Mr Neville Irwin suggests that the data are correct and homogeneous and that in that area there has definitely been an increase in MAP post 1980. A series of rainfall stations in the immediate area were analysed and a similar trend was found. Future research is suggested to determine how widespread this trend is and to determine if similar trends appear elsewhere in southern Africa and for different time periods.

There is a wealth of information in the literature pertaining to the occurrences of the El Niňo and La Niňa phenomena, but little in-depth research has been done successfully to determine if these phenomena really affect the South African rainfall amounts to the extent that it is believed to do. The rainfall database developed in this project will assist research into a possible link between El Niňo and La Niňa and the dry and wet spells experienced in South Africa.

CHAPTER 8

REFERENCES

Antevs, E. (1938). Rainfall and tree growth in the Great Basin. The American Geographical Society. pp 47-57.

Beven, K.J. (2000). Rainfall–Runoff Modeling: The Primer. Wiley and Sons, Sussex, United Kingdom.

Biswas, A.K. (1967). Development of rain gauges. Transactions of the American Society of Civil Engineers, 93 (IR3), 99–124.

Boughton, W.C. (1981). Rainfall variability in rainfall-runoff modeling. Civil Engineering Transactions, Australia, 23, 68-73.

Brunsdon, C. (2002). Personal communication. Reader in Spatial Analysis, Department of Geography, University of Newcastle-upon-Tyne, United Kingdom.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1996). Geographically weighted regression: A method for exploring spatial non-stationarity. Geographical Analysis, 28, 281–289.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1999). Some notes on parametric significance tests for Geographically Weighted Regression. Journal of Regional Science, 39, 497-524.

Burrough, P.A. (1986). Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford, United Kingdom.

Burt, J.E. and Barber, G.M. (1996). Elementary Statistics for Geographers (2nd Edition). The Guilford Press, London, United Kingdom.

Chief Directorate: Surveys and Mapping (1990). Private Bag X10, Mowbray, South Africa.

Chrisman, N. (1997). Exploring Geographic Information Systems. John Wiley and Sons Inc., New York, United States of America.

Clarke, G.P.Y. (1999). Personal communication. School of Mathematics, Statistics and Information Technology, University of Natal, Pietermaritzburg, South Africa.

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

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of Royal Statistics Society, B 39, 1-38.

Dent, M.C., Lynch, S.D. and Schulze, R.E. (1989). Mapping mean annual and other rainfall statistics over Southern Africa. University of Natal, Department of Agricultural Engineering, ACRU Report 27, Water Research Commission, Pretoria, South Africa. Report No 109/1/89.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2000). Quantitative Geography, Perspectives on Spatial Data Analysis. SAGE Publications, London, United Kingdom.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). Geographically Weighted Regression, the Analysis of Spatially Varying Relationships. John Wiley and Sons, London, United Kingdom.

Goodale, C.L., Aber, J.D. and Ollinger, S.V. (1998). Mapping monthly precipitation, temperature and solar radiation for Ireland with polynomial regression and a digital elevation model. Climate Research, 10, 35-49.

Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V. and Brown, C.J. (1997). The Atlas of Southern African Birds including Botswana, Lesotho, Namibia, South Africa and Zimbabwe. Birdlife South Africa, Johannesburg.

Hofierka, J., Parajka, J., Mitasova, H. and Mitas, L. (2002). Multivariate interpolation of precipitation using regularized spline with tension. Transactions in GIS, 6(2), 135-150.

Hughes, F. (2001). Geographically Weighted Regression for mapping mean annual precipitation in South Africa: An appraisal. Unpublished BSc (Hons) Project. School of Mathematics, Statistics and Information Technology, University of Natal, Pietermaritzburg, South Africa.

Hughes, F., Haines, L. and Lynch, S.D. (2001). Geographically Weighted Regression for mapping mean annual precipitation in South Africa: An appraisal. Poster paper included in the Proceedings, Tenth South African Hydrological Sciences Symposium, Pietermaritzburg, South Africa.

Hutchinson, M.F. (1998). Interpolation of rainfall data with thin plate smoothing splines: 1, Two dimensional smoothing of data with short range correlation. Journal of Geographic Information and Decision Analysis, 2, 152-167.

Hutchinson, M.F. and Gallant, J.C. (2000). Digital elevation models and representation of terrain shape. In: Wilson, J.P. and Gallant, J.C. (Eds) Terrain Analysis: Principles and Applications. John Wiley & Sons, Inc., New York, United States of America.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K. and Lucas, N. (2001). Using ArcGIS Geostatistical Analyst. ESRI, Redlands CA, United States of America.

Knisel, W.G., Renard, K.G. and Lane, L.J. (1979). Hydrologic data collection : How long is long enough? Proceedings of the Speciality Conference on Irrigation and Drainage in the Nineteen Eighties, American Society of Civil Engineering, Albuquerque, New Mexico, United States of America.

Kousky, V. (2002). Personal communication. Research Meteorologist, NOAA Climate Prediction Center, Camp Springs, MD, United States of America.

Lanza, L.G., Ramírez, J.A. and Todini, E. (2001). Stochastic rainfall interpolation and downscaling. Hydrology and Earth System Sciences, 5(2), 139-143.

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2001). Geographic Information Systems and Science. John Wiley and Sons, New York, United States of America.

Lynch, S.D. (2001a). Variable window size multiple regression of mean annual precipitation. Proceedings, Tenth Hydrological Sciences Symposium, University of Natal, Pietermaritzburg, South Africa.

Lynch, S.D. (2001b). A rainfall database for the new millennium. Proceedings, Tenth Hydrological Sciences Symposium, University of Natal, Pietermaritzburg, South Africa.

Lynch, S.D. (2002). Digital elevation models and spatial resolution. South African Journal of Science, 98, 219-222.

Lynch, S.D. and Dent, M.C. (1990). Appropriate record lengths for the estimation of mean annual and mean monthly precipitation in Southern Africa. Water S.A., 16, 93-98.

Lynch, S.D., Zulu, Z.T., King, K.N. and Knoesen, D.M. (2001a). 74 Years of rainfall recorded by the Irwins on two farms south of Potchefstroom. Water S.A., 27(4), 559-564.

Lynch, S.D., Zulu, Z.T., King, K.N., and Knoesen, D.M., (2001b). 74 Years of the Irwin's dedication in watching and recording rainfall on two farms south of Potchefstroom. Proceedings, Tenth Hydrological Sciences Symposium, University of Natal, Pietermaritzburg, South Africa.

Makhuva, T., Pegram, G.G.S., Sparks, R. and Zucchini, W. (1997a). Patching rainfall data using regression methods. 1. Best subset selection, EM and pseudo-EM methods: Theory. Journal of Hydrology, 198, 289-307.
Makhuva, T., Pegram, G.G.S., Sparks, R. and Zucchini, W. (1997b). Patching rainfall data using regression methods. 2. Comparisons of accuracy, bias and efficiency. Journal of Hydrology, 198, 308-318.
McNeill, L., Brandão, A., Zucchini, W. and Joubert, A.R. (1994). Interpolation of the Daily Rainfall Model. Report to the Water Research Commission, Pretoria, South Africa, WRC Report No. 305/1/94.
Meier, K.B. (1997). Development of a Spatial Database for Agrohydrological Model Applications in Southern Africa. Unpublished MSc Engineering dissertation, Department of Agricultural Engineering, University of Natal, Pietermaritzburg, South Africa.
Mills, G. and Hes, L. (1997). The Complete Book of Mammals of Southern Africa. Struik Publishers, Cape Town, South Africa.
Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). Applied Linear Statistical Models (Fourth Edition). McGraw-Hill, New York, United States of America.
Pegram, G.G.S. (2002). Personal communication. Department of Civil Engineering, University of Natal, Durban, South Africa.
Saveliev, A.A., Mucharamova, S.S. and Piliugin, G.A. (1998). Modeling of the daily rainfall values using surface under tension and kriging. Journal of Geographic Information and Decision Analysis, 2, 58-71.
Schultz, C. B. (1985). The sensitivity of output from a distributed hydrological model to rainfall input. Unpublished MSc. University of Natal, Pietermaritzburg.
Schulze, R.E. (2002). Personal communication. School of Bioresources Engineering and Environmental Hydrology, University of Natal, Pietermaritzburg, South Africa.
Schulze, R.E. (2003). Development of a Database of Gridded Daily Temperatures for Southern Africa. University of Natal, Pietermaritzburg, School of Bioresources Engineering and Environmental Hydrology. ACRUcons Report, 41. pp 82.
Seeber, R.J. (1983). A Southern African extreme-value model for the determination of the cumulative distribution of point rainfall rate. NITR, CSIR, Pretoria, South Africa. Report TEL 192.
Seed, A.W. and Austin, G.L. (1990). Sampling errors for rain gauge-derived mean areal daily and monthly rainfall. Journal of Hydrology, 118, 163–173.
Smithers, J.C. (2002). Personal communication, Ratio rainfall infilling techniques. School of Bioresources Engineering and Environmental Hydrology, University of Natal, Pietermaritzburg, South Africa.
Smithers, J.C. and Schulze, R.E. (2000). Long duration design rainfall estimates for South Africa. School of Bioresources Engineering and Environmental Hydrology, University of Natal, Pietermaritzburg, South Africa. Report to the Water Research Commission, Pretoria, South Africa, Report No. 811/1/00.
The Chief Director (1990). CAELUM-A history of notable weather events in South Africa: 1500-1990. The South African Weather Service, Private Bag X097, Pretoria, South Africa.
Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2), 234-240.
U.S. Geological Survey (1996). EROS Data Center, 47914, 252nd Street, Sioux Falls, SD, United States of America.

Ward, R.C. (1975). Principles of Hydrology, 2nd Ed. McGraw Hill, London, United Kingdom, pp 16–49.

Whitmore, J.S. (1967). The influence of locality factors on mean annual rainfall. Hydrological Research Division, Department of Water Affairs, Technical Report No. 40, pp 14.

World Meteorological Organisation (1974). Guide to Hydrological Practices. WMO, Geneva, Switzerland, No. 168, Ed. 3.

Zucchini, W. and Hiemstra, L.A.V. (1984). Part 1: Augmenting hydrological records. Water Research Commission, Pretoria, South Africa. WRC Report, 91/3/84.

CHAPTER 9

APPENDICES

The appendices are contained in separate Portable Document Files (PDFs) that can be located on the CD-ROM.

9.1	***calendar_yrs.pdf***	100 Years of above and below average rainfall years in South Africa using calendar years (1900-1999).
9.2	***hydrological_yrs.pdf***	101 Years of above and below average rainfall years in South Africa using hydrological years (1899/1900-1999/2000).
9.3	***mean and median monthly rfl.pdf***	Mean and median monthly rainfall.
9.4	***chance of rfl.pdf***	Percentage chance of more than a certain amount of rain falling on a particular day.
9.5	***required record length.pdf***	Record length required to calculate a representative MAP value.
9.6	***raindays.pdf***	The importance of research surrounding rain days in South Africa, and their application.
9.7	***resources required.pdf***	Resources required to generate the final MAP raster.