RADAR REFLECTIVITY INFILLING TECHNIQUES

by

Stephen Malcolm Wesson

Submitted in fulfilment of the academic requirements for the degree of

Masters of Science in Engineering

in the

Civil Engineering Programme

University of KwaZulu-Natal, Howard College

2005

Durban

December 2005

ABSTRACT

Weather radar provides a detailed spatial representation of rainfall over a large area and in a real time basis. It has proven to be a valuable tool for hydrologists, agriculturalists and organisations that require accurate and real time information of rainfall and overcomes many of the disadvantages associated with the traditional raingauge estimate. However one of the shortcomings of the rainfall estimates supplied by weather radars is that there are quality problems associated with radar rainfall images that include ground clutter, beam blocking and anomalous propagation to name a few. To obtain the best rainfall estimate possible, techniques for removing ground clutter (non-meteorological echoes that influence radar data quality) on two-dimensional (2D) and three-dimensional (3D) radar rainfall image data sets were developed in this study.

The chosen method for estimating the "true" values behind the contaminated data was Kriging, which is considered to be the optimal technique for the spatial prediction of Gaussian data. Kriging has various advantages and disadvantages, which need to be taken into consideration in this type of application. For the radar rainfall images to be repaired in real time a computationally fast and efficient method of estimating the missing contaminated data was needed. This is achieved by exploiting the various properties associated with Kriging.

In South Africa, radar volume scan data are currently only available on one-kilometre horizontal grids at one-kilometre intervals above the earth's surface. This may not be an accurate representation of the rainfall that actually reaches ground level. To provide an estimate of the "true" rainfall reaching the earth's surface, an algorithm has been developed that extrapolates the radar data down to ground level. The extrapolation is carried out using a combination of 3D Universal and Ordinary Kriging which is preceded by a rainfall classification algorithm developed and calibrated in this study.

The techniques proposed for ground clutter infilling and the extrapolation of radar data to ground level have been tested for their effectiveness and efficiency on a wide range of selected rainfall events and indicate that the methodology is practically useful. The South African Weather Service (SAWS) has recently installed the software to "cleanse" the radar data as it is received in real time.

PREFACE

I, Stephen Malcolm Wesson hereby declare that the whole of this dissertation is my own work and has not been submitted in part, or in whole to any other University. Where use has been made of the work of others, it has been duly acknowledged in the text. The research work was carried out in the Department of Civil Engineering, Surveying and Construction, University of KwaZulu-Natal, Howard College, Durban, under the supervision of Professor G.G.S Pegram.

Date: .....

Signature: .....

As the candidate's supervisor I have approved this dissertation for submission.

Date: .....

Name: .....

Signature: .....

PREFACE

Weather radar can provide a detailed spatial representation of rainfall in real time and therefore provides a valuable tool for hydrologists, disaster managers, agriculturalists and any organisation that depends on the accurate and timely prediction of rainfall. Unfortunately data quality issues exist with the radar rainfall product which can result in inaccurate rainfall measurements.

This thesis presents an algorithm to improve the rainfall estimates supplied by weather radars. The structure of this thesis differs from the standard as a large proportion comprises two articles, one published and one that is currently under review. While conducting the research it was decided that the initial work involving the instabilities of the Kriging equations was important and novel enough to warrant publication. The paper was written and accepted for publication in 2004 in *Hydrology and Earth System Sciences* (HESS) and is included in Chapter 3 of this thesis.

The second paper, which is currently under review, resulted from some of this work done in this study was presented at the European General Sciences – General Assembly in Vienna, Austria in 2005. An invitation was extended to the participants of the session to submit a paper to a special edition of *Natural Hazards and Earth System Sciences* (NHESS). The paper has been submitted and is undergoing the review process; the paper is included in Chapter 4 of this thesis.

ACKNOWLEDGMENTS

I would like to express my gratitude to the following people and organisations:

- Professor Pegram for his constant support, encouragement and guidance throughout the research.
- Scott Sinclair for providing valuable advice and assistance, especially in regard to computer programming.
- Pieter Visser and Deon Terblanche for providing invaluable meteorological advice and data throughout the research.
- The Water Research Commission (WRC) for providing financial assistance for this project.
- The financial assistance of the Department of Labour (DoL) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily attributed to the DoL.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BTNN	-	Border Tracing Nearest neighbours
CAPPI	-	Constant Altitude Plan Position Indicator
CDF	-	Cumulative Distribution Function
CSIR	-	Council for Scientific and Industrial Research
DEM	-	Digital Elevation Model
FLOPS	-	Floating-Point Operations per Second
GRAHI	-	Group of Applied Research on Hydrometeorology
HESS	-	Hydrology and Earth System Sciences
ITNN	-	Individual Target Nearest Neighbours
K-S	-	Kolmogrov-Smirnov
MAVPR	-	Mean Apparent Vertical Profile of Reflectivity
mdv	-	Meteorological Data Volume
MSSE	-	Mean Sum of Square of Errors
NHESS	-	Natural Hazards and Earth System Sciences
NOWRAD	-	Nowcasting of Precipitation with Radar
NPRP	-	National Precipitation Research Programme
SAWB	-	South African Weather Bureau
SAWS	-	South African Weather Service
SSE	-	Sum of Square of Errors
SVD	-	Singular Value Decomposition
WAR	-	Wetted Area Ratio
WRC	-	Water Research Commission

CHAPTER 1
INTRODUCTION

Water is becoming an increasingly limited and valuable resource as populations rapidly grow and demand increases; this is especially true for developing countries such as South Africa. Extreme rainfall events can cause damage and loss of life making the accurate forecasts of these events vitally important. To address the above concerns the measurement of rainfall in real time and the ability to capture its high spatial variability is vital. Weather radar provides a detailed spatial representation of rainfall over a large area and in a real time basis. It has proven to be a valuable tool for organisations that require accurate and real time information. Currently the South African Weather Service has a network of eleven weather radars distributed over South Africa to measure rainfall.

At present however there is various data quality problems associated with the rainfall estimates supplied by weather radar. These can include ground clutter, beam blocking and anaprop to name a few; the result is that large regions within the radar volume scan contain contaminated data where no information of the rainfall intensity can be obtained. The data supplied by the weather radars are also only available from one-kilometre above ground level and there are regions within the volume scan where the rainrate is unknown. The rainfall estimates at ground level are of more interest and importance than the measurements aloft, which are unlikely to be an accurate indication of the rainfall at ground level since precipitation tends to be affected by a variety of atmospheric phenomena before reaching the earth's surface. The motivations for this study were to obtain the most accurate estimate of rainfall possible from weather radars by overcoming the various data quality problems associated with the estimates supplied.

The objectives of the study were to improve the quality of radar rainfall estimates at ground level on a real time basis. The first aim of this study was to provide a computational method to infill ground clutter contamination in a radar volume scan by use of Kriging. The next objective of the study was to provide an extrapolation of the radar volume scan data to ground level. The overall guiding intention of the study was to provide a fast and computational efficient algorithm written as a computer program which could be utilised by the South African Weather Service (SAWS). The dissertation has the structure and contents which are described below.

Chapter 2 of this dissertation contains an overview of two instruments, raingauges and weather radar, used to measure rainfall with the advantages and disadvantages associated with both are presented. An overview of how weather radars operate and measure rainfall is presented.

Chapter 3 of the thesis presents various techniques to provide computational stability and efficiency when applying the Kriging algorithm to radar reflectivity data. A theoretical and numerical justification of the 'screening effect' is provided that ensures computational efficiency. Two techniques are tested to infill ground clutter regions in two-dimensional (2D). The first of two papers is included as part of this chapter.

The next chapter, Chapter 4, provides a description of the operational method to infill all the unknown and contaminated data contained in a radar volume scan and in the process provides an estimate of the rainfall at ground level. A rainfall classification algorithm is described and a bright band correction procedure presented. Climatological semivariograms associated with the various rainfall types and the use of Ordinary and Universal Kriging are described together with their application in areas of classified rainfall. The use of Cascade Kriging to progressively infill unknown information to ground level is presented. The second of the two papers is included in this chapter.

The results associated with the testing of the algorithm are presented in Chapter 5 which includes a detailed statistical analysis of the testing undertaken to determine the effectiveness and efficiency of the proposed algorithms to infill ground clutter and also providing an estimate of the rainfall at ground level.

The final section of the thesis provides a summary of the work presented in the thesis and also recommendations for further work to be carried out on the presented algorithms. The algorithms presented in this thesis have been coded in C++ and the computer programme is currently being implemented to run on the computer system of the SAWS to infill regions contaminated with ground clutter and to provide an estimate of the rainfall at ground level.

CHAPTER 2
RAINFALL MEASURMENT BY RADAR

2.1 INTRODUCTION

In this chapter two instruments for measuring rainfall, raingauges and weather radar, are discussed and an overview of their limitations and advantages is presented. The majority of the chapter provides a review the data quality problems of the weather radar with emphasis on the problem of ground clutter echoes. An overview of how various researchers and weather bureaus have dealt with ground clutter are discussed. Finally an overview of rainfall estimation at ground level using radar rainfall data will be given.

The accurate measurement of precipitation is an inherently difficult task since the spatial and temporal variability of the rainfield is high and the true rainfield can never be completely known. No single instrument can give an error free estimation of rainfall. However many sectors such as agriculture, disaster managers and catchment managers (to name a few) rely on (hopefully) accurate and timely estimates of rainfall.

In a country such as South Africa, which is predominantly semi-arid and has a below average rainfall in terms of a global average, careful and accurate measurement of rainfall is required. The rainfall experienced in South Africa, especially in the interior of the country, is predominantly convective (Tyson and Preston-Whyte, 2000: 205-207) and has the potential to cause flash floods and damage to property. Convective rainfall also has the tendency to have high spatial and temporal variability that makes the measurement and forecasting of severe rainfall a difficult task.

To provide the best estimates of rainfall, the traditional method of measuring rainfall by raingauges, may result in unrealistic estimates, especially with convective rainfall. Remote sensing techniques, such as weather radars, do provide an alternative method and often improve the spatial accuracy.

The primary instruments used to measure rainfall at present can be listed as raingauges, radars and satellites. Efforts are currently underway in South Africa, in the form of a Water Research Commission (WRC) project, to combine the three estimates obtained from these instruments to provide a daily rainfall map over the country (Pegram *et al*, 2005).

These measurement devices all have their limitations and advantages. A brief review of raingauges in the next sub-section will now be given.

2.2 RAINGAUGE DATA

Raingauges are the traditional tool used for the recording of rainfall and are often regarded as the "true", or reference, rainfall estimates at ground level. Raingauges have provided a valuable source of information and raingauge measurements are used to calibrate and check that remote sensing techniques, such as radar, are returning "correct" estimates of rainfall values (Collier, 1986). A brief overview of raingauge data collection in South Africa is presented in the following section.

2.2.1 Raingauge Network in South Africa

Currently in South Africa there are approximately 152 manual standard rain gauges and 120 telemetering tipping bucket raingauges. The 152 manual standard raingauges return data read by an observer on a daily basis. The 120 telemetering tipping bucket raingauges work on an automatic basis where cellular phone technology is used to communicate data via SMS to the SAWS offices. Figure 2.1 provides an illustration of a solar powered automatic tipping bucket raingauge. The raingauge can measure rainfall at 15 second intervals and records every 0.2 mm of rainfall that is recorded and the time of the 0.2 mm accumulation. At 15 minute time intervals the recorded 0.2 mm rainfall accumulations are transmitted via SMS.

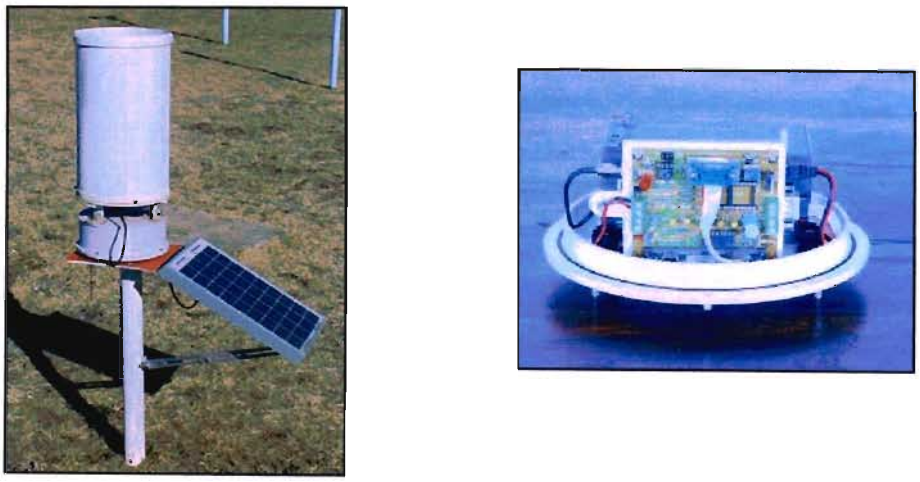


Figure 2.1: Photograph of a telemeter tipping bucket raingauge located in the Liebenbergsvlei catchment, South Africa.

The locations of the 120 automatic tipping bucket raingauges are indicated in Figure 2.2 by the yellow dots and the 152 manual standard raingauges by the red dots. South Africa is approximately 1.27 million square kilometres in size and the number of existing raingauges does not provide the sampling resolution required to accurately monitor precipitation values in the country. As can also be seen in Figure 2.2 the raingauges are spaced at irregular intervals that also add to the difficulty of determining an accurate representation of the rainfall across the entire country (Deyzel, 2003).

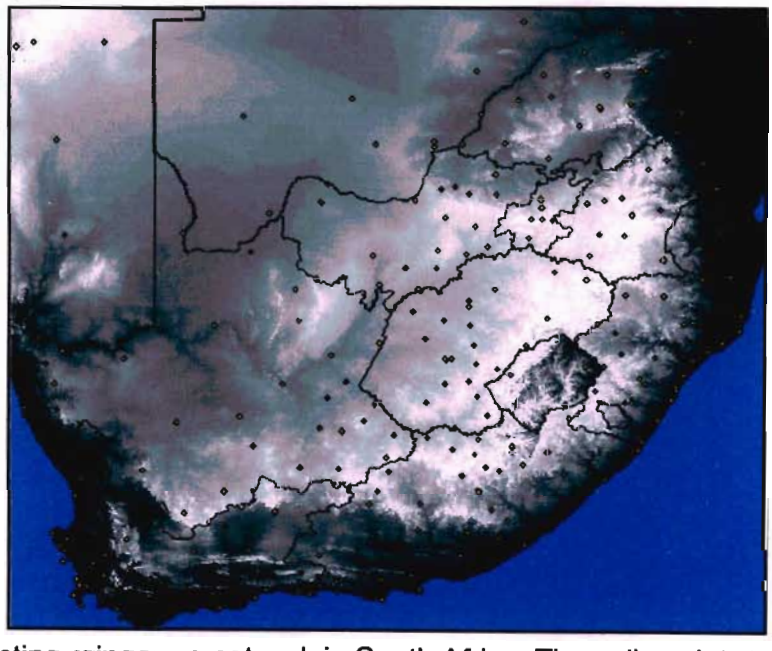


Figure 2.2: Existing raingauge network in South Africa. The yellow dots indicate the 120 automatic tipping bucket raingauges where the red dots indicate 152 manual standard raingauges.

2.2.2 Advantages

The advantages of raingauges can be listed as follows. They provide a direct measurement of rainfall depth at a point and measure accumulated rainfall. The data returned from raingauges is also easily understood and interpreted. Raingauges are also relatively cheap, easy to install and maintain. There are also comparatively long data records associated with raingauges and were the only rainfall measurements prior to the deployment of weather radars in the 1980's and weather satellites in the 1990's.

2.2.3 Limitations

Limitations do exist with the raingauge data. In South Africa the number of reliable daily reporting raingauges is falling. Raingauges provide an accurate accumulation

measurement of the point rainfall and fail to capture its spatial variability; even with a dense network of raingauges it is difficult to interpolate and extrapolate gauge data in any significant detail or accuracy especially, at short time scales. This is especially relevant in South Africa where the interior of the country predominantly receives very local convective rainfall that has a high temporal and spatial variability that raingauges can fail to measure in any significant accuracy. Raingauges also tend to underestimate the rainfall in situations of heavy rainfall by as much as 20 to 40% (Wilson and Brandes, 1979).

The types of errors associated with raingauges such as under measurement in heavy rainfall, can be corrected by the use of Equation (2.1) as discussed by Todini (2001).

$$I = \alpha \cdot I_G^{\beta} \qquad (2.1)$$

Where I is the actual rain intensity (mm/hr), I_G is the raingauge measured intensity (mm/hr) or tipping rate and α and β are the raingauge specific calibration parameters where the values are determined by special equipment or supplied by the manufacturer.

Random errors are also present which can be demonstrated by the difference in measurement obtained between closely situated raingauges. The random errors that exist between raingauges can be substantial when the local rainfall intensity is high and the timescales are short (Ciach, 2002). Habib and Krajewski (2002) also demonstrated that random errors exist between raingauges situated within a distance of 1 metre from one another, and highlighted the discrepancies that can occur between radar and raingauge accumulations

2.3 WEATHER RADAR DATA

Radar was initially developed in the late 1930's for the purpose of detecting enemy aircraft during World War 2; a brief historical account of the history of radar is given by Austin (Austin, 2001). On its inception it was noticed that a wide variety of echoes were returned to the display including birds, returns from the earth itself and rainfall.

2.3.1 Weather Radar Development and Network in South Africa

The first radars in South African originated in 1943 during World War 2 to monitor shipping movements. After the war in 1948 a weather radar was installed in Johannesburg by the

Council for Scientific and Industrial Research (CSIR) to monitor rainfall and the results of cloud seeding experiments. From the cloud seeding experiments, studies were conducted on hailstorms and the storms that produced them. During the 1960's weather radars were installed at Pretoria, Durban and Port Elizabeth by the South African Weather Bureau (SAWB) for upper air balloon tracking and weather observations. During the 1980's and 1990's research was focused on cloud seeding experiments especially at Bethlehem. In 1994 the MRL-5 C band weather radar was installed 20 km outside of Bethlehem to assist in the National Precipitation Research Programme (NPRP) (Pegram and Terblanche, 2001).

At present the SAWS radar network consists of eleven radars as illustrated in Figure 2.3. The radars are situated at Irene, Bloemfontein, Durban, Port Elizabeth, East London, Cape Town, Ermelo, De Aar, Bethlehem, Umtata and Polokwane. The radars are of C-Band type with 5 cm wavelength. The exception to this is the research radar located at Bethlehem which is an MRL–5 dual wavelength research radar, 10 cm S-Band and 3 cm X-Band, (Terblanche *et al*, 2001) and from which the vast majority of the data for this study was obtained. The majority of the SAWS radars are situated in the Eastern half of the country so as to cover the regions that have higher population densities and higher annual rainfall.

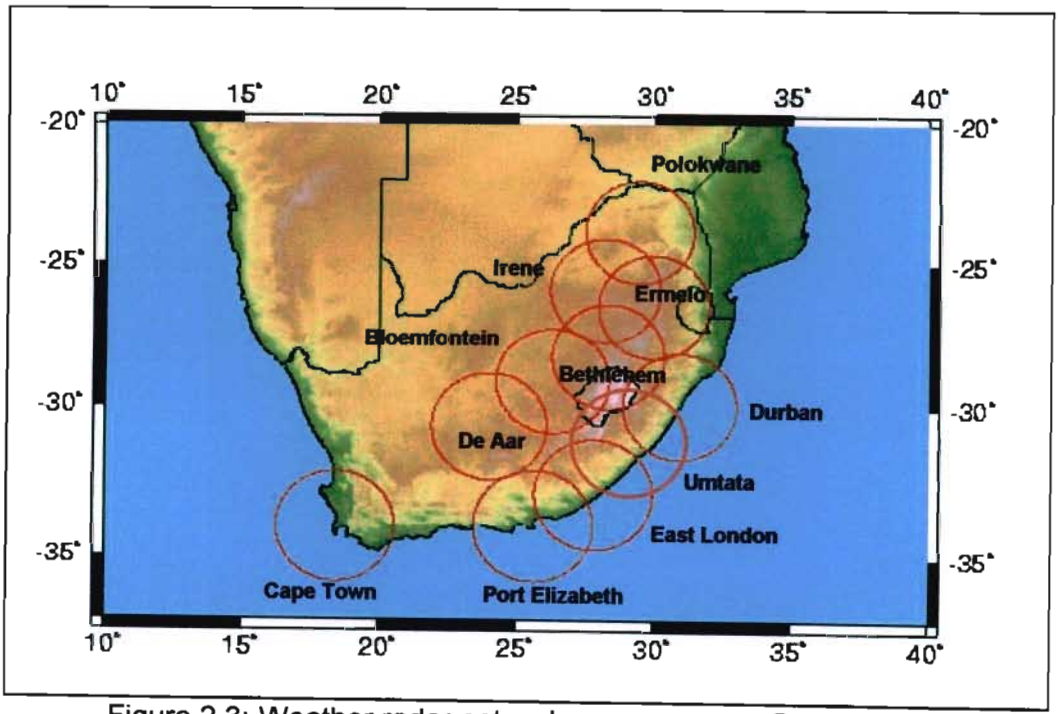


Figure 2.3: Weather radar network coverage over South Africa.

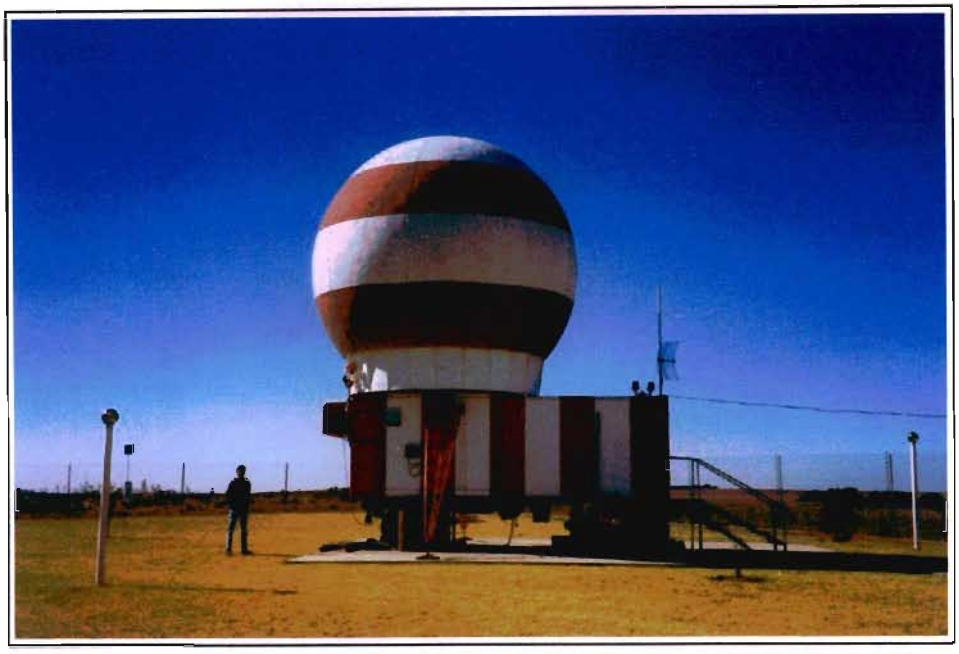


Figure 2.4: The MRL5 radar near Bethlehem.

2.3.2 Background Information and Terminology

A weather radar can be described as a remote sensing instrument that transmits and receives reflected electromagnetic energy. Short pulses of microwave energy produced by the transmitter in the radar are focused into a narrow beam. The pulses of energy travel through the atmosphere at the speed of light. Any targets that are encountered by the pulse of energy, such as raindrops, scatter the energy which is directed back to the radar. The reflected (a minute fraction of the emitted energy) energy is returned to the radar where a receiver system processes the returned data. Reflectivity is measured in units of dBZ which correspond to the intensity of rainfall. The reflected intensity is related to the number of raindrops per unit volume (Interpreting Weather Radar Information, 2005).

To convert reflectivity to rainrate, where the units used are mm/hr, an appropriate transform is used, such as the Marshall-Palmer formula (Marshall and Palmer, 1948) given in Equation (2.2) and used routinely by SAWS, where Z is the reflectivity (dBZ) and R is the rainrate (mm/hr).

$$Z = 200 \cdot R^{1.6} \qquad (2.2)$$

Figure 2.5 illustrates a typical instantaneous reflectivity image taken from the Bethlehem (South Africa) weather radar on 25th February 2003. The grey portions of the image indicate where no data was available and the black regions indicate regions of marked ground clutter.

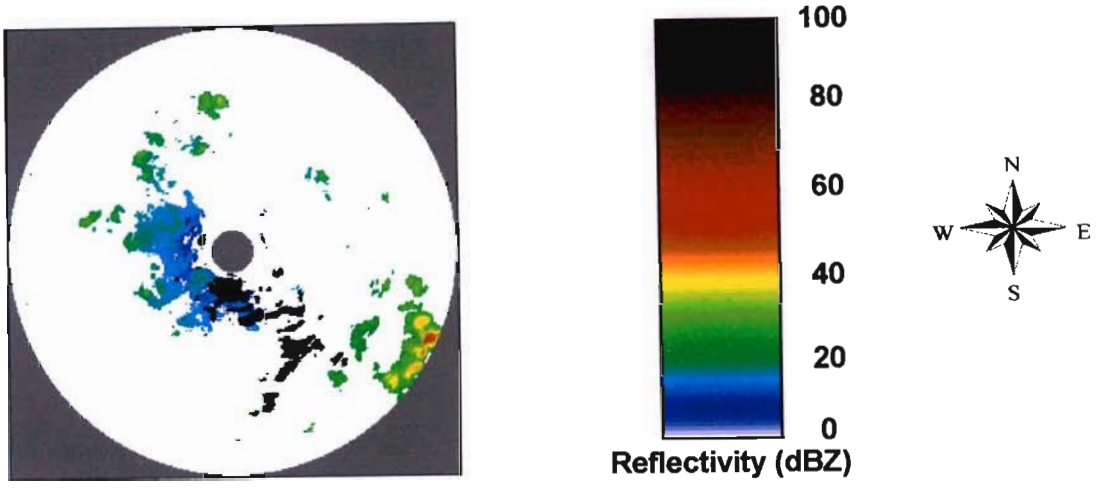


Figure 2.5: Typical radar reflectivity image taken from the Bethlehem (South Africa) weather radar on the 25 February 2003. The grey portions of the image indicate where no data are available and the black regions indicate ground clutter. The image is 300 km square.

A reflectivity image such as Figure 2.5 is obtained in the following manner. The radar performs 18 full rotation incremental elevation sweeps; the first elevation sweep begins at an angle of 1.5^0 and progressively works upwards to an angle of 55^0. The lowest elevation sweep of 1.5^0 is known as the base scan with the full set of scans known as the radar volume scan which is illustrated in Figure 2.6.

The raw reflectivity data returned from the weather radars in South Africa are in the form of a three-dimensional radar volume scan data specified in spherical co-ordinates. These are converted into Constant Altitude Plan Position Indicators (CAPPIs) to represent the data. CAPPIs are two-dimensional fields set out in pseudo Cartesian co-ordinates whose height is measured relative to mean sea level and with a local datum at the radar level. In South Africa the CAPPIs are at 1 km square resolution and of size 400 km by 400 km. The CAPPIs are spaced at 1km intervals from radar level and range in height from 1 km to 18 km above the radar level following the earth's curvature, as illustrated in Figure 2.7. The radar volume scan data are produced at approximately five-minute intervals.

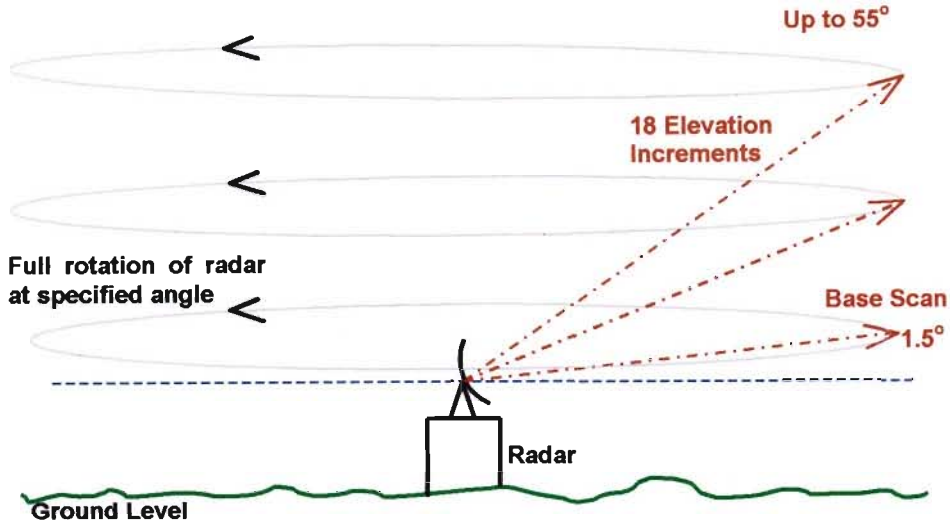


Figure 2.6: Illustration of the weather radar scanning procedure to obtain a radar volume scan information.

In order to convert the three-dimensional radar volume scan data from spherical to two-dimensional CAPPI co-ordinates two methodologies have been investigated for South African weather radars. The first method proposed by Seed in 1992 utilises a simple projection technique as is illustrated in Figure 2.7. The algorithm involves projecting the nearest bin value in the volume scan data to the CAPPI. Although this algorithm is very fast it has the effect of creating discontinuities on the CAPPI data.

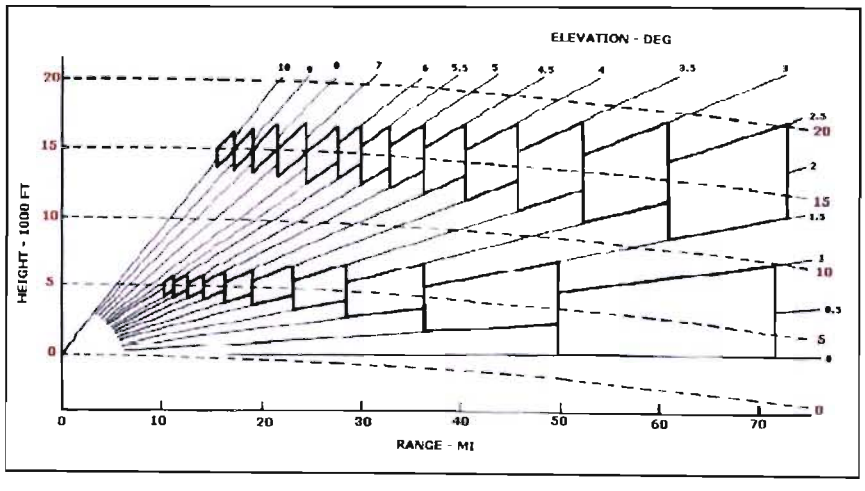


Figure 2.7: Illustration of how CAPPIs are obtained via use of a projection technique (Battan, 1973).

The second and improved approach devised by Mittermaier and Terblanche (1997), and which is currently routinely used to convert the polar data to Cartesian CAPPI data, works as follows. The method computes the linear weighted average from the eight nearest bin reflectivities, where 4 are used from the scan above and 4 from the scan below, surrounding the point on the CAPPI which is being determined. In the averaging process the range, azimuth, and elevation angle of the target and the eight bin points are taken into account. The result is that a far smoother CAPPI image is produced.

Once the data have been converted from spherical to Cartesian co-ordinates, the CAPPI data, which now consist of a three-dimensional matrix set of 18 levels of 400 by 400 grids for one volume scan, are converted and stored in a Meteorological Data Volume (mdv) file. The mdv file format is specifically designed to compress and store the large volumes of data which are returned from weather radars.

2.3.3 Advantages

There are numerous advantages attached to the rainfall data that are returned from weather radars which makes it an attractive alternative to the traditional raingauge data. Many of the disadvantages associated with raingauge data are overcome by weather radar, especially in the context of the problems in South Africa. Instantaneous rainfall images show rainfall in comprehensive spatial detail and complexity and can provide valuable information on the spatial structure of rainfall. The coverage of weather radars is also large and can range up to a 200 kilometre radius. The data returned from the radar is in real time and the spatial detail of the rainfall makes it ideal for the prediction of severe weather patterns and flood forecasting for disaster managers.

2.3.4 Limitations

There are, however, various data quality issues associated with the data returned from weather radars. An outline of a few of these limitations are presented below:

BRIGHT BAND

As snow and ice crystals drop through the 0°C isotherm, they begin to melt and are surrounded by a thin layer of water. The result is that, from the viewpoint of the radar, the melting snow and ice crystals resemble large blobs of water and the reflectivity values for the melting level are enhanced greatly, providing an overestimation of rain (Sanchez-Diezma *et al*, 2000). This level is called the bright band, which can be identified by the

unexpected increase in reflectivity that occurs in the vertical reflectivity profile, in South Africa typically at the 2 km CAPPI, and in plan can also be identified as rings of abnormally high reflectivity values. The bright band presents problems for the accurate prediction of rainfall at ground level and most often results in an underestimate of the rainfall at ground level. At Bethelehem in South Africa the bright band generally occurs at 2 km altitude as determined by Mittermaier (1999). A typical image with regions affected by bright band is shown in Figure 2.8.

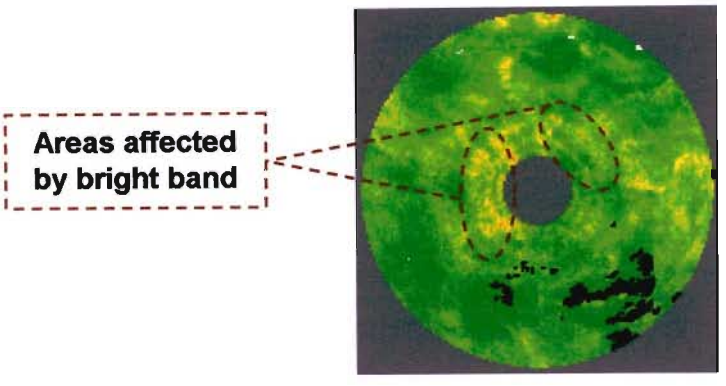


Figure 2.8: Radar reflectivity image taken from the Bethlehem weather radar, 17 December 1995. Indicated on the 2 km CAPPI image are selected regions which show evidence of being affected by a bright band. The dimensions of the image are 150 km square.

ANOMALOUS PROPAGATION

As the electromagnetic pulse from the radar travels through the atmosphere it does not always travel in a straight line. The atmosphere frequently consists of different density gradients created from layers of different temperature and moisture. This has the effect of dramatically refracting and bending the electromagnetic wave as it travels through the atmosphere. The beam can be refracted and bent so that it eventually strikes the surface of the ground; this takes the shape of a concentric circular band. This will return a signal to the radar of a region of high reflectivity values. The regions of anomalous propagation are not permanent and fluctuate in size as they are dependent on atmospheric conditions. In the interior of South Africa anomalous propagation generally occurs during the winter months and is associated with periods where there is no rainfall and clear skies.

BEAM BLOCKING

Beam blocking occurs when a physical obstruction such as a tower or building partially blocks the electromagnetic energy pulse. This creates a gap in the weather radar CAPPI data where no rainfall is recorded. An example of this is illustrated in Figure 2.9 where the instantaneous reflectivity image is taken from the Durban weather radar on the 11 December 2000. Illustrated on the image is a region in the South West quadrant affected by beam blocking; this contaminates all of the lower CAPPIs of the radar volume scan. No rainfall data can be detected in a zone of beam blocking which can lead to substantial segments of the radar never recording rainfall.

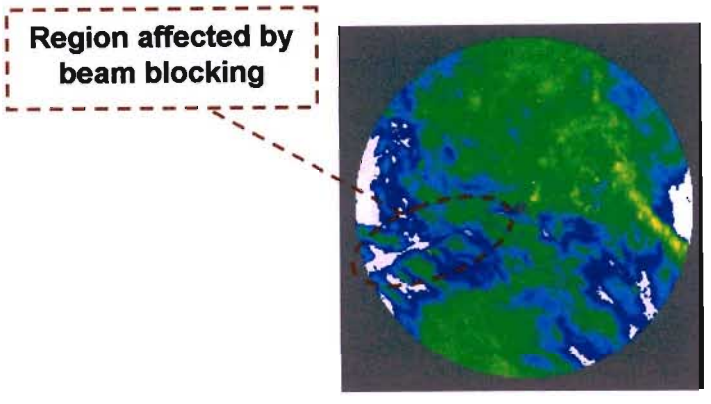


Figure 2.9: Radar reflectivity image taken from the Durban weather radar, 11 December 2000. Illustrated on the image is a region affected by beam blocking in the South West quadrant. The size of the image is 250 km square.

GROUND CLUTTER

A major source of problems associated with radar data quality is ground clutter contamination. Ground clutter refers to the scenario when the radar beam strikes the earth's surface or any other non-meteorological land based object and returns a signal to the weather radar. Ground clutter contamination generally occurs when weather radars are situated in mountainous regions or areas with high lying terrain and can be identified as permanent returns to the weather radar, however the location of ground clutter can vary due to refraction of the weather radar beam in the atmosphere. The returns to the weather radar generally consist of high reflectivity values (e.g. +60 dBZ). The South African weather radar network experiences ground clutter contamination at the majority of the sites that make up the weather radar network. At the Bethlehem weather radar site in South Africa ground clutter contamination occurs in the South East sector of the weather radar due to the Maluti and Rooiberge Mountains. The majority of this dissertation deals

with the problem of removing ground clutter, but the techniques can be applied to repair other contaminated regions once they have been identified.

2.4 GROUND CLUTTER REMOVAL

The accurate detection and removal of ground clutter is important for a variety of reasons. The returns of ground clutter reflectivities to the radar usually exhibit abnormally high values which can lead to severe overestimates of rainfall. At present a variety of organisations and the general public have access to radar rainfall images, via the internet (e.g. http://metsys.weathersa.co.za/RadarGeneral.htm), who do not have specialised knowledge of radar meteorology. The presence of ground clutter echoes could lead to misinterpretation and confusion of expected and current rainfall intensity. An example of an image available from the Innovation and Research Division of SAWS website appears in Figure 2.10. The image is from the Durban weather radar (14 July 2005) on a day which experienced no rainfall. The dark portions of the image indicate ground clutter contamination.

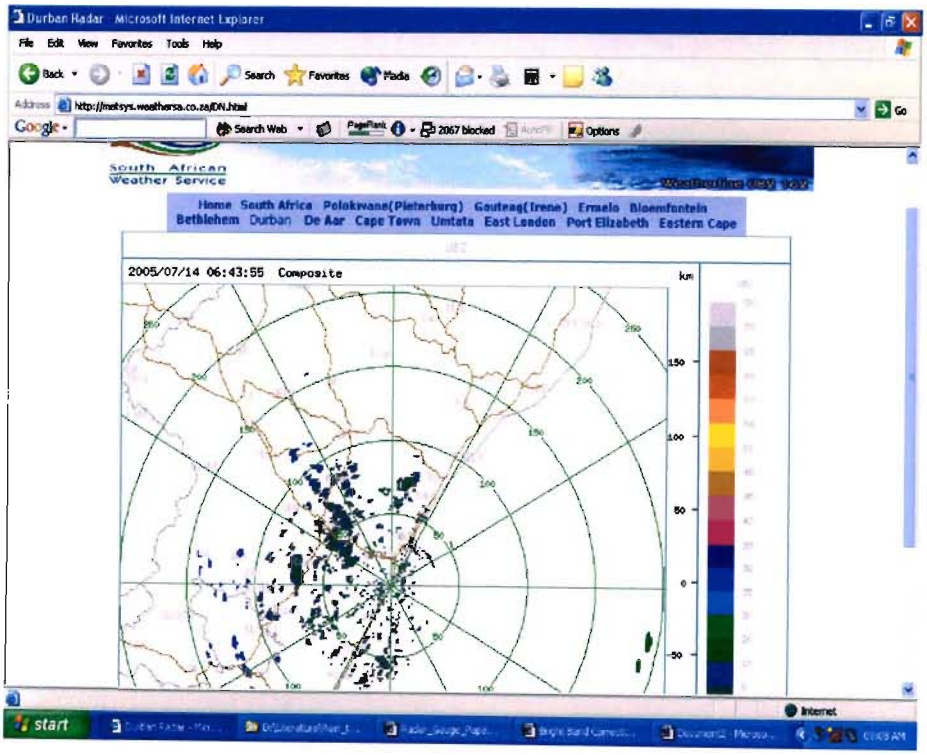


Figure 2.10: Instantaneous reflectivity image taken from the Durban weather radar on the 14 July 2005 (available at: http://metsys.weathersa.co.za/DN.html) on a day which experienced no rainfall. Dark portions of the image represent regions contaminated with ground clutter.

2.4.1 Methodologies to Identify Ground Clutter Regions

Ground clutter removal can be viewed as a two-step process, firstly the ground clutter regions need to be identified and secondly the ground clutter removed and an estimate of the rainfall inserted as a replacement. A brief overview of the current methodologies adopted by various metrological institutes and researchers to identify and remove ground clutter echoes is presented below.

FIXED GROUND CLUTTER MAPS

One of the techniques used to identify the regions contaminated by ground clutter is by use of a clutter map. Fixed ground clutter maps can be obtained by recording the location of radar echoes obtained from days with no rainfall. A disadvantage of this technique is that ground clutter is not entirely stationary and location can vary according to the atmospheric conditions present and is also affected by the scanning algorithm. This can result in areas where rain is occurring being hidden by the ground clutter map (Riedl, 1994).

DOPPLER GROUND CLUTTER MAPS

Radars that possess Doppler capabilities can detect regions of ground clutter by examining the radial velocities of returned radar signals above a given threshold. In this instance it is assumed that ground clutter regions will return near zero radial velocities since the ground clutter is near stationary over long periods of time. If the radial velocity of the returned signal is close to zero, and does not exceed a particular threshold value, then it is marked as a ground clutter location. This technique does possess the shortcoming that near stationary or slow moving rainfall such as stratiform rain may be incorrectly classified as ground clutter. Not all weather radars have Doppler capability; this particular method of identifying ground clutter is not always possible.

ADAPTIVE GROUND CLUTTER MAPS

A statistical based scheme can be utilised to determine the location of ground clutter. The assumption is made that the variability of the ground clutter is less than that of precipitation over each image. Over a period of time the clutter echoes will be highly correlated to one another whereas the weather echoes are uncorrelated. Once the computed correlation, which can be calculated for each pixel in the radar volume scan, exceeds a predetermined threshold value, that location is marked as ground clutter. This

method of identifying ground clutter is currently utilised by Meteo-France, the UK Met Office (Sugier, 2002) and SAWS.

2.4.2 Methodologies to Infill Ground Clutter Regions

NEAREST PIXEL

A simple technique to provide an estimate of rainfall in a ground clutter region is to simply substitute the ground clutter point with the nearest uncontaminated rainfall value as is done by Crespi *et al* (1994) in treating ground clutter for a weather radar situated in Italy at Monte Grande. Although this technique is straightforward and computational efficient it will most probably provide poor estimates since the nearest point is highly unlikely to be a good indication of the rainfall in the ground clutter pixel of the actual rainfield. By using only the nearest pixel the possibility also exists that undetected ground clutter values will be used as an the estimate of the rainfall.

INTERPOLATION SCHEME

An example of an interpolation scheme is that proposed by Seed and Pegram (2001) who used 3D Kriging to estimate marked ground clutter locations. The 3D Kriging was tested using two different approaches, using firstly a nearest neighbourhood approach where 5, 10, 20 and 40 nearest neighbours were used to estimate ground clutter points on an individual basis; a second approach investigated was to infill all the missing data simultaneously where Fast Fourier Transforms were utilised so as to cope with the large data sets. These two approaches were tested on radar reflectivity data from the Melbourne, Australian weather radar and were shown to be effective in removing ground clutter contamination in a real time basis. The ideas presented in that paper have been a source of some of the ideas for this thesis.

2.4.3 Specific Methodologies Adopted in Other Studies to Identify and Infill Ground Clutter

GROUP OF APPLIED RESEARCH ON HYDROMETEOROLOGY (GRAHI)

The Group of Applied Research on Hydrometeorology (GRAHI) in Spain at Catalunya was founded to provide scientific and technological support in the field of hydrometeorological management and forecasting.

In order to detect ground clutter and anomalous propagation, fuzzy-logic techniques are utilised, the algorithm is dynamic and works on a real time basis. Probability values based on values of variables measured at each point are then determined in order to identify possible ground clutter locations. The fuzzy-logic technique works as follows. The data input into the scheme are reflectivity and Doppler data. From these data, fields are generated of reflectivity, radial velocity, vertical gradient, echotop and texture which are determined for every point in the field; the fields are known as membership functions in the algorithm. A weighting is applied to each of the membership functions and a final probability field is computed for every pixel in the volume scan. If the probability is greater than 0.5 in the field then it is marked as a ground clutter location. Testing on typical rain events revealed that the above algorithm identified the majority of ground clutter points but did falsely identify certain pixels as ground clutter locations (Berenguer *et al*, 2005).

The ground clutter substitution algorithm works on polar co-ordinate radar volume scan data. The first step of the algorithm involves a classification of the radar volume scan data into stratiform and convective zones. The algorithm employed relies on using a combination of both horizontal and vertical interpolation based on the rainfall type. The horizontal interpolation is applied in regions of stratiform rain (regions of low reflectivity gradients) and vertical interpolation in regions of convective rainfall (regions of high reflectivity gradients). Initially horizontal interpolation is performed using Delaunay triangulation to identify the nearest non-contaminated points and a linear interpolation then employed. All the substituted values that exceed a threshold value of 45 dBZ are then reselected for vertical substitution where the nearest non-contaminated value in the vertical column is then substituted. This process is then repeated for all radar levels (Sanchez-Diezma *et al*, 2001).

The method takes into account the natural structure of the rainfall by means of identifying stratiform and convective rainfall. However use of pure vertical substitution for the convective rainfall may lead to an under estimate of the areas considered to be convective, since convective rainfall tends to show a steep vertical profile of reflectivity and often intensifies as ground level is approached (Steiner *et al*, 1978). The proposed methodology returns more accurate infilling results than the ground clutter substitution method employed by the Spanish National Institute of Meteorology. The results returned for the GRAHI infilling procedure indicate a relatively accurate method for infilling ground clutter regions.

SOUTH AFRICAN WEATHER SERVICES (SAWS)

Currently the Innovation and Research Division of SAWS at Bethlehem employs an auto-correlation technique to identify the regions in the radar volume scan which are contaminated by ground clutter. The assumption is made that the variability of rainfall is greater than that of ground clutter since ground clutter is predominantly stationary. The ground clutter detection algorithm works as follows; where Equation (2.3) is applied to each pixel:

$$E_T = X_T \cdot X_{T-1} \cdot x + E_{T-1} \cdot (1 - x) \qquad (2.3)$$

where E_T refers to the auto-correlation of radar reflectivity of a given pixel at time T, X_T refers to the reflectivity of a given pixel at time T and x refers to the extinction coefficient. The extinction coefficient (x) is set, after checking a range of options, to 0.03 so that slow moving rainfall is not flagged as ground clutter. When the auto-correlation E_T exceeds a set threshold that location is automatically flagged as containing ground clutter, the threshold value is set to 37.4 dBZ obtained after 40 volume scans. The horizontal and vertical auto-correlations are also examined as a secondary test; if the vertical difference of the time series exceeds a set threshold (22.3 dBZ over a period of 40 volume scans) or if the horizontal difference exceeds a set threshold (14.14 dBZ also over a period of 40 volume scans) then that pixel is marked as ground clutter (Visser, 2003).

Once a pixel is identified as containing ground clutter, rainfall is not estimated at that particular pixel location for the rest of the day. A final filter is also applied to the image to ensure no isolated ground clutter pixels are left as rainfall values and the filter consists of examining a 5 by 5 grid area. An example of a derived clutter map for the Bethlehem weather radar on the 5 December 2003 is illustrated in Figure 2.11 taken during a time period that experienced no rainfall. As illustrated, extensive ground clutter is located in the South East region of the image where the Maluti and Rooiberge mountains are located.

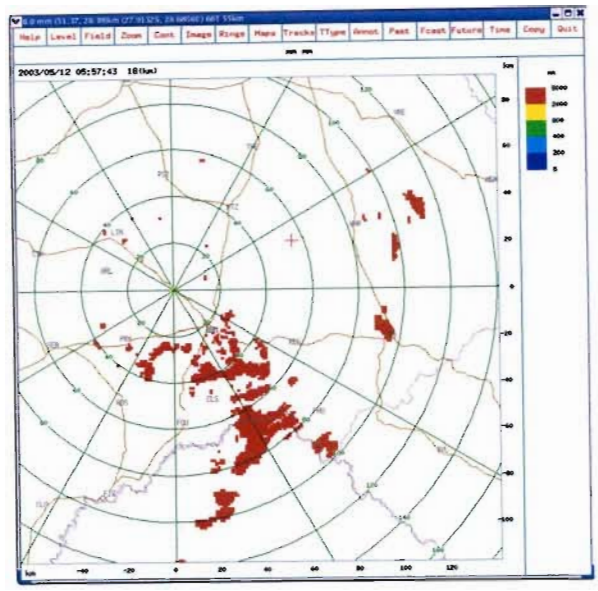


Figure 2.11: Reflectivity data taken from the Bethlehem weather radar from a time period that experienced no rainfall on the 5 December 1995. As illustrated in the image extensive ground clutter is situated in the South East region of the image.

Current practice is that once all the ground clutter locations have been located, the contaminated pixels are masked out in the volume scan. No replacement or estimation of the possible rainfall in these locations of the contaminated pixels takes place at present.

SWISS METEOROLOGY BUREAU (MeteoSwiss)

At present the weather radar network of MeteoSwiss consists of three operational radars and six research radars. The project NOWRAD (NOWcasting of precipitation with RADar) places strong emphasise on data quality control to ensure accurate precipitation products are returned from weather radars. Outlined in this sub-section are the steps performed in the data control algorithm to detect and infill all radar volume scan data contaminated by ground clutter (Joss and Lee, 1995).

In order to detect if a reflectivity measurement is a ground clutter point, a decision tree classification system is used. At each step in the decision tree an attempt is made to classify the point as rain, ground clutter or system noise. The steps in the algorithm are as follows:

- Minimum Detectable Signal Test – if the received signal is below a given threshold, which is set slightly above the system noise level, the pixel is set a precipitation rate of zero.

- Wide-Band Noise Test – the wide band noise test is performed in the Doppler channel; if the coherence is below a certain threshold, then the signal is marked as precipitation otherwise the signal is marked as a ground clutter location.
- Velocity Algorithm Test – If the signal velocity is above a set threshold, in this instance 2 m/s, the received signal is classified as a significant precipitation value since this indicates movement of the signal. If the signal is classified as precipitation then the ground clutter points on the map are deleted at that location. If the signal velocity is below the threshold the signal is marked as possible ground clutter.
- Statistical Clutter-Filter Algorithm Tests – the test at this point is for signals that are coherent but have approximately zero velocity. The test determines the probability that the signal is precipitation, which is based on the signal probability distribution. If the statistical distribution of the signal indicates the pixel value is an outlier, the pixel is marked as ground clutter.
- Gradient Algorithm Test – In this test the change in returned signal power with height is examined. A large change in gradient, greater than 10 dB/deg, indicates a ground clutter or anomalous propagation location, so the pixel for that location is then marked as a ground clutter location.

If none of the steps in the decision tree conclusively determines if the returned signal power is a ground clutter or precipitation pixel, then an adaptive ground clutter map is consulted. If the clutter map indicates that the position in question has experienced significant clutter in the recent past then the measurement is marked as ground clutter, otherwise it is marked as precipitation.

Once all points in the volume scan have been marked as containing precipitation or ground clutter, the clutter is then eliminated from the volume scan by use of a nearest neighbours averaging technique. The raw polar data has 83-metre resolution, so each 1 kilometre zone consists of 12 reflectivity measurements. All ground clutter free measurements in each 1 km window are averaged to produce a single reflectivity estimate for that 1 km range bin (Joss and Lee, 1995).

2.5 RAINFALL ESTIMATION AT GROUND LEVEL

The radar reflectivity data contained in the South African CAPPIs are only available from one-kilometre above radar level and there are regions within the volume scan where the

rainrate is unknown. In applications such as disaster management, hydrology and agriculture, the rainfall estimates at ground level are of more interest and importance than the measurements aloft, which are unlikely to be an accurate indication of the rainfall at ground level (Jordan *et al*, 2000) since precipitation tends to be affected by a variety of atmospheric phenomena before reaching the earth's surface. An example of the bottom three CAPPIs of radar reflectivity data supplied by SAWS is indicated in Figure 2.12.

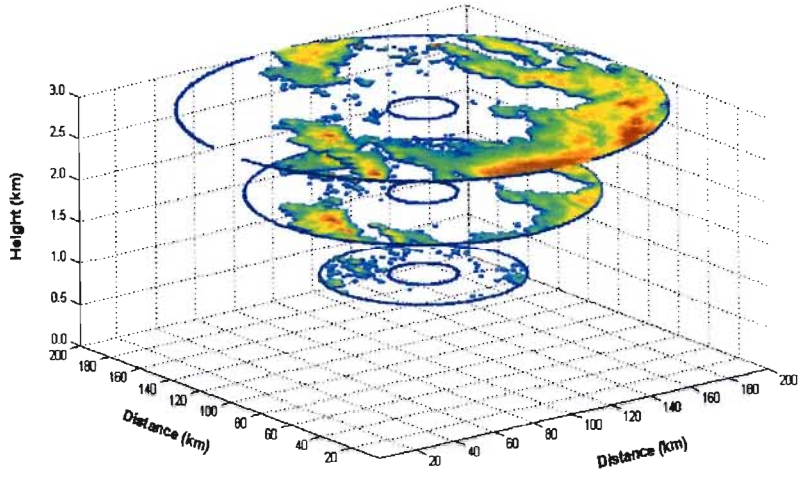


Figure 2.12: 3D Graphical illustration of radar volume scan data up to 3 km above ground level. Data are from the Bethlehem weather radar (14 February 1996).

As indicated in Figure 2.12 the rainfall is unknown at ground level and large portions of the volume scan exist where no rainfall data are available. Numerous organisations have made attempts to extrapolate the above given radar reflectivity above to ground level in order to provide a reasonable estimate of the instantaneous rainrate at ground level. Some of the methodologies employed by the various organisations are outlined in the next sub-section.

2.5.1 Present Methodologies

GRAHI

The methodology proposed by GRAHI to extrapolate radar reflectivity data contained aloft to ground level is based upon estimating the shape of the vertical profile of reflectivity. The algorithm employed involves identifying the vertical profile of reflectivity close to the

radar's location. The identified profile is then applied at range from the radar to obtain an estimate of the rainfall at ground level up to a distance of 100 km from the weather radar.

The algorithm works by firstly computing the Mean Apparent Vertical Profile of Reflectivity (MAVPR). The mean profile shape is firstly computed from reflectivity data contained in close proximity to the radar, the profile is then normalized with reference to a chosen reference height uses the mean profile. Once the MAVPR is obtained the Vertical Profile of Reflectivity can then be obtained in any other region of the volume scan by multiplying the MAVPR by a fitting factor.

Problems identified in this algorithm were related to the bright band, which occurs in stratiform rainfall, and the highest inaccuracies were associated with this zone. The algorithm also relies on significant rainfall being situated close to the radar. The algorithm is simple and computationally efficient which makes it suitable for real time applications. However the effect of the bright band is left uncorrected in this algorithm that provides a degraded estimate at ground level especially at distance from the radar (Franco *et al*, 2002).

LACK & FOX

Wind drift can have a drastic effect on the accuracy of rainfall estimates at ground level. The radar reflectivity values are contained aloft at distances in excess of 1 km above ground level as the rainfall falls towards ground level wind can laterally advect the rain over large distances. The reflectivity values contained directly above a point on the grounds surface may not be an accurate indication of the rainfall at ground level at that point due to wind drift.

The algorithm proposed by Lack and Fox (2004) uses input data of reflectivity and; u, v and w wind components which are obtained by using a Doppler radar. The algorithm works by computing how much of the precipitation is advected into adjacent cells when the precipitation is set out on a Cartesian grid. Droplet fall speed is computed (w) and the horizontal vectors (u and v) also computed for each grid cell.

The contribution of the rainfall in a pixel to an adjacent pixel or pixels can then be calculated by examining the horizontal and vertical wind velocity components. The contribution is determined by calculating a critical radius for each pixel, the critical radius

then gives an indication of which adjacent pixels are affected and the fraction of the contribution that is received at a pixel. A new grid is then derived based on the sum of the fractions of contribution. The new estimate at ground level is then obtained by examining the lowest volume scan data that is available.

SWISS METEOROLOGY BUREAU (MeteoSwiss)

The algorithm utilised by MeteoSwiss to provide an estimate of the rainfall at ground level works on firstly considering the sources of errors in the horizontal and vertical directions and then using a mean vertical profile to provide an estimate of the rainfall at ground level in a manner similar to the algorithm employed by GRAHI.

The horizontal contribution of errors can include the effects of the earth's curvature, ground clutter and beam broadening with distance as examples of a few of the errors that may influence radar data quality. The errors are corrected and any contaminated data (e.g. ground clutter) is then estimated for each volume scan.

The vertical contribution of the radar to the rainfall at ground level is then considered. The vertical profile of rainfall is highly variable and dependent on rainfall type. The mean vertical profile is computed in real time for each volume scan, within a radius of 70 km of the radar, where a weighting is applied so that variables located closer to the radar have greater emphasis. The mean vertical profile is then used to provide an estimate of the rainfall at ground level with distance from the radar.

The shortcomings of this algorithm are that in order for a mean profile to be computed significant rainfall needs to be present within the 70 km range. Bright band could also introduce inaccuracies into the estimates at ground level and no distinction is made between rainfall types; convective and stratiform (Joss and Lee, 1995).

SOUTH AFRICAN WEATHER SERVICE

In order to provide an estimate of the rainfall at ground level for the Daily Rainfall Mapping Project in South Africa (Pegram *et al*, 2005) the following approaches have being investigated. Initially the rainfall estimate at ground level was taken as the maximum vertical radar reflectivity value. However when the weather radar accumulation estimates were compared to raingauge accumulation estimates it was shown by Visser (2003) that

this provides serious overestimation of the rainfall at ground level even for convective rain events.

The alternatives investigated were to use an average value of the vertical profile of reflectivity, which was shown for selected stratiform and convective rain events to provide an improved estimate in terms of matching the raingauge accumulation values more closely. The method worked well for convective rainfall but tended to provide a poor estimate of stratiform rainfall. The poor estimates of stratiform rainfall may be attributed to the effects of the bright band. The averaging method is currently implemented and replaced the method of using the maximum value in the vertical profile.

The alteration of the Marshall-Palmer formula (Equation (2.2)) was also investigated to provide an improved estimate of the rainfall at ground level when utilising the averaging method. It was determined that a more appropriate relationship to use was $Z = 300 \cdot R^{1.5}$ which provided a far improved estimate than the Marshall-Palmer relationship that is currently utilised (Visser, 2003).

At present the averaging method seems to provide a reasonable estimate of the rainfall at ground level and is computationally efficient and fast. An alternative and improved method of providing rainfall estimates at ground level is presented in the Chapter 4 of this thesis.

2.6 CHAPTER SUMMARY

In this chapter an overview of the advantages and disadvantages attached to two instruments used to measure rainfall, raingauges and weather radar. A brief history of the development of the weather radar network in South Africa was outlined; background information on how weather radars operate and detect rainfall was discussed.

The various disadvantages and data quality problems associated with weather radar were reviewed. These include bright band, beam blocking, anomalous propagation and ground clutter. The problem of ground clutter was focused upon and the various methodologies that have been developed by researches and meteorological organisations to identify and remove regions of ground clutter were outlined. The problem of rainfall estimation at ground level from weather radar volume scan data was discussed as were the various techniques developed to provide an estimation of the rainfall rate at ground level.

The remaining chapters of this thesis attempt to address the problems of infilling all the unknown and contaminated radar volume scan data and provide an estimate of the rainfall at ground level on a real time basis. The algorithm developed is designed specifically so as to be utilised by SAWS in order to improve the quality of radar rainfall estimates in South Africa.

CHAPTER 3
KRIGING METHODS USED FOR IMAGE REPAIR

3.1 INTRODUCTION

In this chapter an overview of Kriging, which is used in the estimation of missing data, will be examined. The problem with computational efficiency and stability associated with Kriging will be discussed along with the various techniques developed to overcome these difficulties. These techniques, which will be elaborated on in the remainder of the chapter, include exploiting the 'screening effect' and matrix rank reduction in combination with Singular Value Decomposition (SVD). Included in this chapter is a copy of the journal paper, which was entitled *Radar rainfall image repair techniques*, published on the topics reviewed in this chapter.

3.2 OVERVIEW OF ORDINARY KRIGING THEORY

Kriging is a geostatistical technique that is named after a South African mining engineer, D.G. Krige, who in the 1950's developed a methodology to determine true distributions based on sample distributions for mining applications (Cressie, 1993: 106). Further research based on Krige's work resulted in the development of the equations of optimal spatial linear prediction; the Kriging equations. Although extensively used in mining applications it has becoming increasingly used in fields such as hydrology and other environmental fields where optimal spatial linear prediction is required (Bras and Rodriguez-Iturbe, 1985: 385-386).

Kriging was chosen as the computational method to infill unknown reflectivity data or data that are contaminated by ground clutter. Kriging is a minimum mean squared error method of spatial prediction and is the best linear unbiased technique when estimating missing data from a Gaussian field (Cressie, 1993: 106).

An example of computed histograms from an instantaneous reflectivity (dBZ) and a rainrate (mm/hr) image is shown in Figure 3.1. The images are from the Bethlehem weather radar (24 January 2002) with the selected CAPPI 4 km above the radar and includes convective and stratiform regions. Only the northern half of the image is considered to ensure that no ground clutter data values are included in the computation of the histograms. The image has the dimensions of a width of 250 km and depth of 125 km.

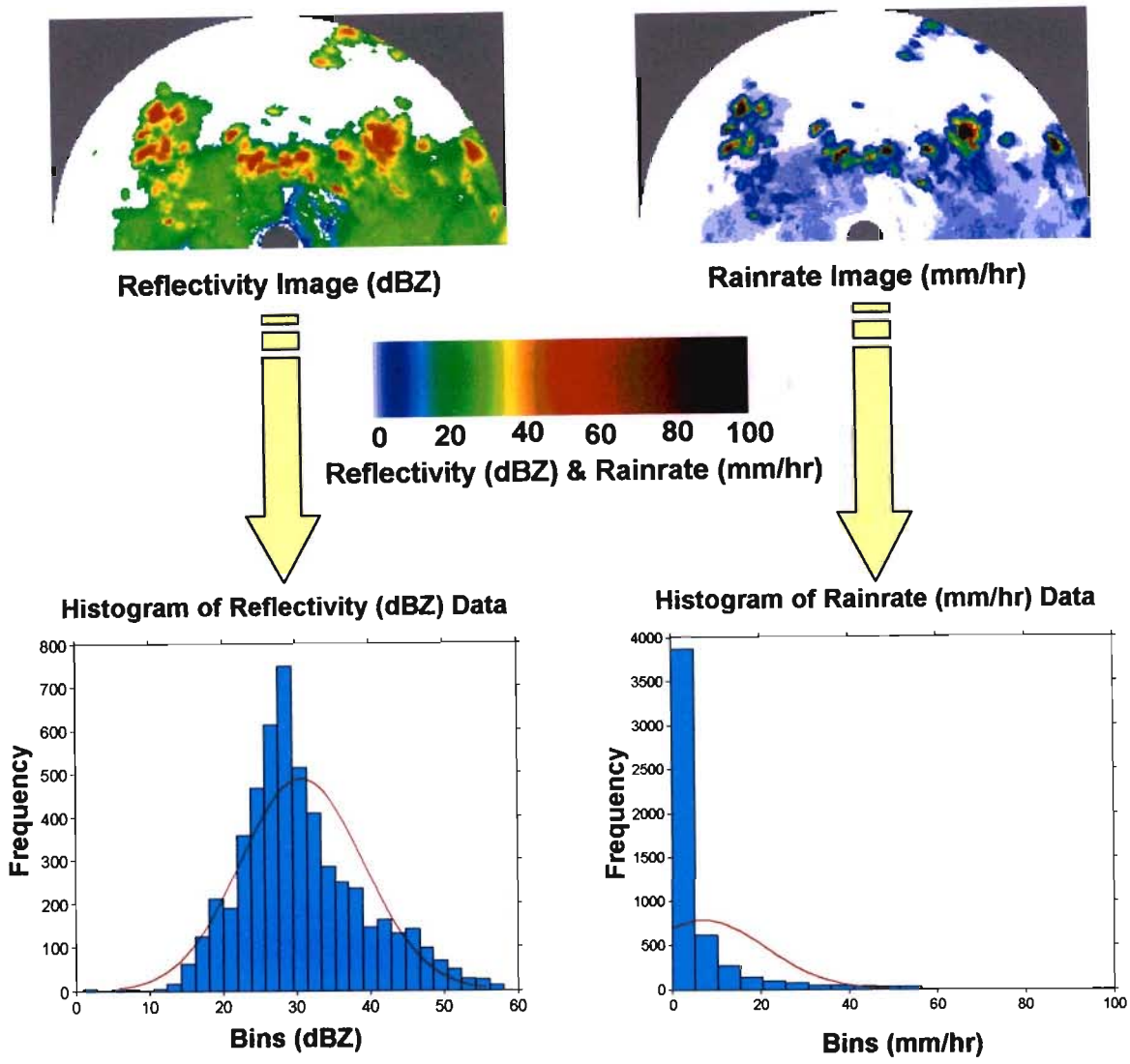


Figure 3.1: Comparison of distributions returned from instantaneous reflectivity (dBZ) and rainrate (mm/hr) images from the Bethlehem weather radar. The red line indicates normal distributions fitted by the method of moments.

As illustrated in Figure 3.1 the reflectivity data returns an approximately normal distribution whereas the rainrate data return a highly skewed lognormal distribution, as has been shown in previous studies (Bell, 1987; Pegram and Clothier, 2001). The left histogram approximates a normal distribution with two modes, possibly due to the convective and stratiform rain. The threshold of 35 dBZ minimizes the area under the tails of either distribution being in this zone. The superimposed ideal normal distribution corresponds relatively well to the distribution computed for reflectivity data. For this reason it was decided to work with data in the reflectivity domain to ensure that a near Gaussian data set is used in optimizing the Kriging computations.

In Ordinary Kriging the mean is assumed constant and unknown in the field and is implicitly estimated from the data. The Ordinary Kriging equation, used to estimate a missing (target) datum from intact (control) data, is given by Equation (3.1):

$$\hat{z}(s_0) = \lambda^T(s_0) \cdot z \tag{3.1}$$

where s_0 is the location at which the datum $\hat{z}(s_0)$ needs to be estimated and is referred to as the target, $\lambda(s_0)$ is a vector of computed weighting values and z the vector of known data or control. The above computation can be carried out to compute the estimate of one target at a time or in matrix form a set of points simultaneously.

The weighting values assigned to the controls are computed by the matrix Equation (3.2):

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda(s_0) \\ \mu(s_0) \end{bmatrix} = \begin{bmatrix} g(s_0) \\ 1 \end{bmatrix} \tag{3.2}$$

where G is a matrix of semivariogram values (in this instance between the controls), u a unit vector of ones, $\mu(s_0)$ is the Lagrange multiplier which is included to ensure that the weighting values sum to unity and $g(s_0)$ is the vector of semivariogram values linking the target and the controls. A simple illustration of a scenario involving three controls and one target is shown in Figure 3.2 and demonstrates how the λ_i are assigned to controls and the G_{ij} and g_i values are determined.

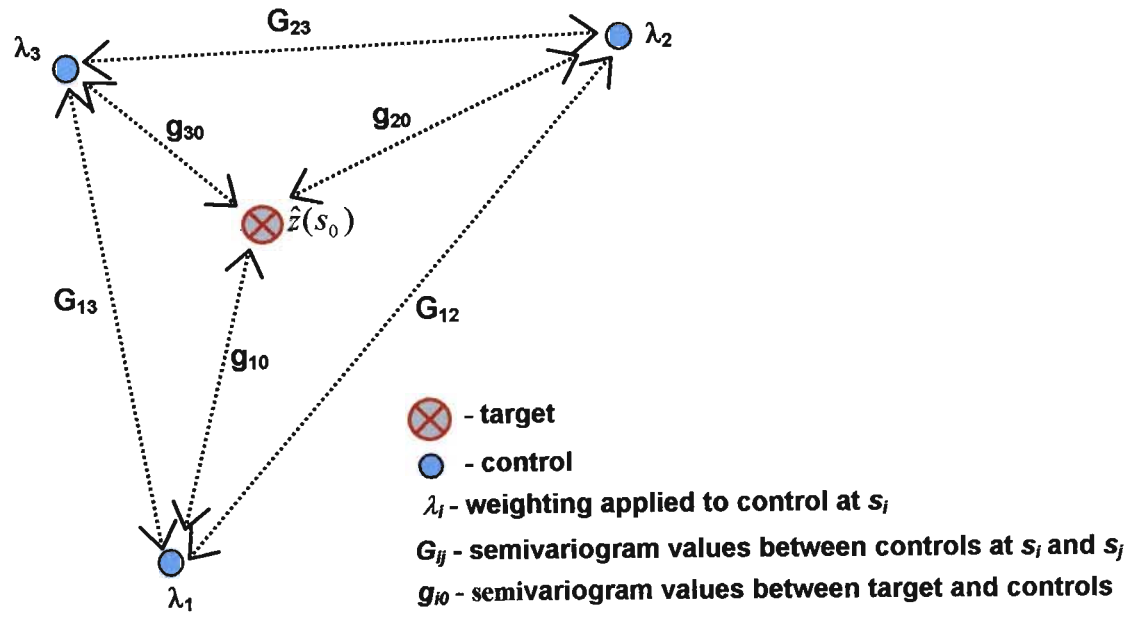


Figure 3.2: Simple illustration of Ordinary Kriging scenario with three controls and one target to be estimated.

In a random field, Ordinary Kriging is carried out using a semivariogram function. The isotropic semivariogram model chosen is the 2-parameter exponential model specified in Equation (3.3) and is used in all Kriging computations in this study.

$$g(s) = 1 - \exp[-(s/L)^{\alpha}] \tag{3.3}$$

where s is the Euclidian distance between data points, L is a scale parameter termed the correlation length and α is the exponent or shape parameter, which lies in the range $0<\alpha\leq 2$; the model is thus Gaussian when $\alpha = 2$ and exponential when $\alpha = 1$.

3.3 METHODS FOR IMPROVING COMPUTATIONAL EFFICIENCY

One of the disadvantages of the Kriging technique is that it relies on the solution of a linear system of equations whose size is proportional to the number of selected controls. For large systems of equations this can be time consuming and computationally burdensome. Previous authors such as Creutin and Obled (1982) have also made reference to the computational burden associated with Kriging.

For a matrix inversion to be performed the number of Floating-Point Operations per Second (FLOPS) performed by a computer is proportional to N^3, where N is equal to the size of the coefficient matrix.

3.3.1 The Screening Effect

There is an unexpected saving which can be exploited to ease the computational burden in a dense data set, because radar data which typically occur on a lattice, not all of the controls need to be used in estimating the missing data points. This is because a "screening effect" occurs where the significantly large Kriging weights associated with the controls are concentrated around the target (Chiles and Delfiner, 1999: 202-206). This can be used advantageously to significantly reduce computation time. For each individual target, or set of targets in a cluster, a Kriging neighbourhood in its vicinity can be selected to reduce the dimensionality, N, of the problem and hence N^3.

A theoretical justification of the screening effect in one-dimensions was derived in Wesson and Pegram (2004) which follows contained on pages 45 to 47, which demonstrates that observations only adjacent to opposite the missing data contain significant weighting values and have significant influence on the infilling of missing data. This should also apply in 2D and 3D. A computational justification of the "screening effect" can also be quite easily demonstrated. An example of the screening effect is demonstrated on a 2D data set by Wesson and Pegram (2004), on page 48, which illustrates that the significant Kriging weights, are situated in close vicinity to the target.

Negative Kriging weighting values do occur with distance from the targets particularly when the exponent in Equation (3.3) increases from 1 to 2. An option does exist to constrain the Kriging weights to be non-negative ($\lambda_i \geq 0$) such as done by Mazzetti (2004: 15-17) when Block Kriging raingauge to radar estimates. However this does have the disadvantage of then constraining the Kriged estimate to lie within the minimum and maximum values of the selected controls (Cressie, 1993: 143). In this application it is more advantageous to allow the Kriged estimates to range outside of the minimum and maximum values of the controls. It can also be regarded as a unnecessary constraint (Cressie, 1993: 143) and using a technique to eliminate non-negative weights such as the one proposed by Barnes and Johnson (Chiles and Delfiner, 1999: 224) is computationally time consuming and in a real time application may prove to be impractical.

The question of the optimum number of controls to select when infilling a target, so that necessary accuracy is maintained and computational efficiency is ensured, will be addressed in Sections 3.4.1 and 3.4.2. The methodology employed and the results obtained will be discussed in detail in these two sub-sections.

3.3.2 Ill-Conditioning of the Coefficient Matrix

In Ordinary Kriging, a set of linear equations needs to be solved in order to determine the weighting values to be applied to the selected controls. The solution to this is illustrated by Equation (3.4), obtained by inverting the coefficient matrix of Equation (3.2).

$$\left[\lambda^T(s_0) \quad \mu(s_0)\right] = \left[g^T(s_0) \quad 1\right] \cdot \begin{bmatrix} G & u \\ u^T & 1 \end{bmatrix}^{-1} \tag{3.4}$$

An unexpected problem which was encountered when applying the Ordinary Kriging technique to infilling ground clutter on instantaneous radar reflectivity images is that the coefficient matrix can be highly ill-conditioned. The ill-conditioning of the coefficient matrix is related to a number of factors which are outlined by Wesson and Pegram (2004) on pages 48 to 52.

In this instance conventional fast methods of computing the inverse of the coefficient matrix cannot be used such as Gauss-Jordan or LU Decomposition. The method of Singular Value Decomposition (SVD) was therefore employed to solve the problem of ill-conditioning. SVD involves the decomposition of the matrix into a column orthogonal matrix U, a diagonal matrix W which contains the singular values along its diagonal and the transpose of an orthogonal matrix V. The singular values that are predominantly close to zero can then be trimmed in order to eliminate the problem of ill-conditioning. The inversion is obtained by simply inverting the remaining singular values and computing the product $U \cdot W^{-1} \cdot V^T$.

In order to determine the appropriate number of singular values (w_i) to trim the following methodology was adopted. The cumulative sum of the squared ranked w_i^2 values is firstly computed. An example of the cumulative sum of the squared singular values (Σw_i^2) is illustrated in Figure 3.3. This was computed for a coefficient matrix of size 150 by 150 for typical α values and a constant L=11 pixels, the first 25 values are shown for illustration

purposes. Figure 3.3 indicates that the first few terms of the coefficient matrix contribute significantly to the Σw_i^2 with the remainder contributing very little.

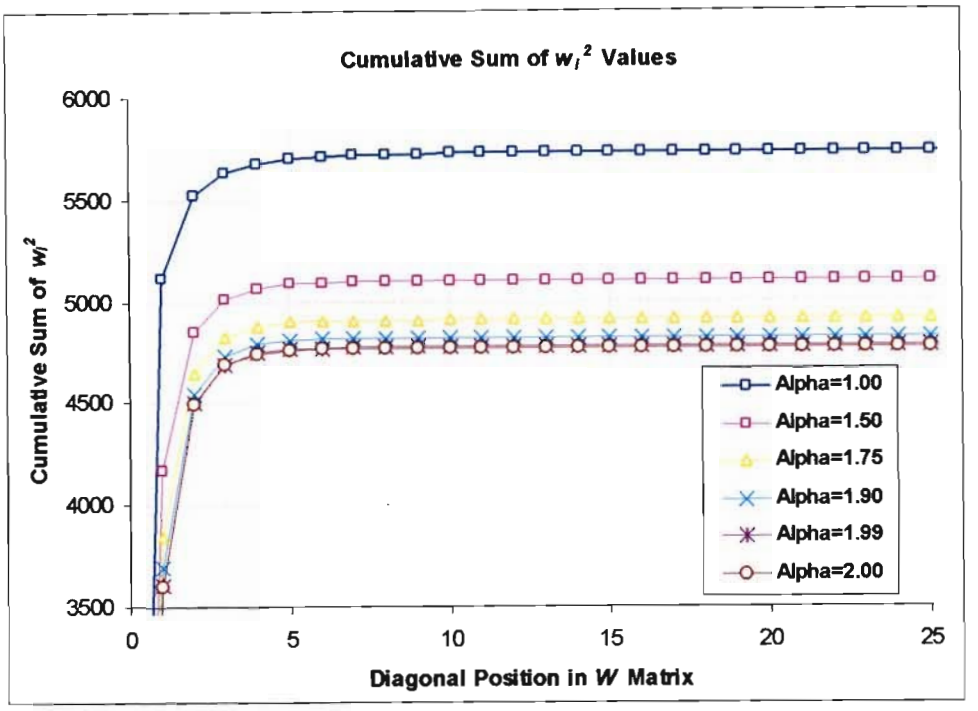


Figure 3.3: Cumulative sum of the w_i^2 and corresponding diagonal position in the W matrix for different values of α and a constant L = 11 pixels.

The appropriate number of w_i values to remove is determined by computing the cumulative sum of all the w_i^2's values. The percentage of the total cumulative sum of the w_i^2 is then computed for each position in the W matrix; once a certain percentage of the total cumulative sum is reached the remainder of the w_i can then be considered significantly small and unlikely to contribute very much. The correct percentage of the total cumulative sum needs to be determined that eliminates the singular values that cause instabilities but also maintain the important w_i values in the W matrix.

To determine the correct percentage value to set as a cut off a test was undertaken which first consisted of selecting a 30 km square portion of an instantaneous reflectivity image. A portion of this field was then removed to simulate a ground clutter scenario and then infilled with a moving nearest neighbours approach strategy, where the number of controls to infill each target consisted of 30 controls. The targets were infilled several times with a range of α and L values. The w_i singular values were then trimmed at certain percentage values of the cumulative sum, and the corresponding Sum of Square of Errors (SSE)

between the observed and estimated data computed. The cumulative sum of singular values, s_j, is computed by Equation (3.5).

$$s_j = \sum_{i=1}^{j} w_i^2 \tag{3.5}$$

The SSE and the s_j for a practical range of α and L values are shown in Figure 3.4. As can be seen in Figure 3.4, once the percentage value of s_j equals or exceeds a value of 99.995% the SSE suggests a poor quality estimate is returned and that too many of the singular values in the W matrix have been removed.

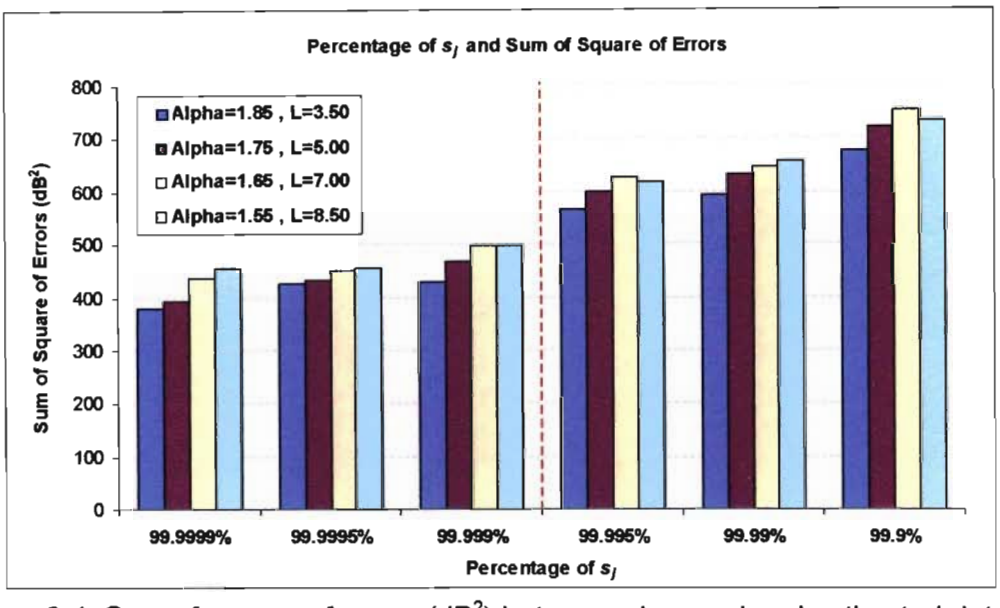


Figure 3.4: Sum of square of errors (dB^2) between observed and estimated data for selected α and L values for different trimming of singular values from W matrix at set percentage values of s_j.

For the test described above the sum of the Kriging weighting values ($\Sigma\lambda_i$) were also computed. If the coefficient matrix in Equation (3.4) is stable the sum of the weights should sum to unity. As shown in Figure 3.5 the results indicate that the weights no longer sum to unity once the singular values in the W matrix are trimmed at and beyond a value of 99.995%.

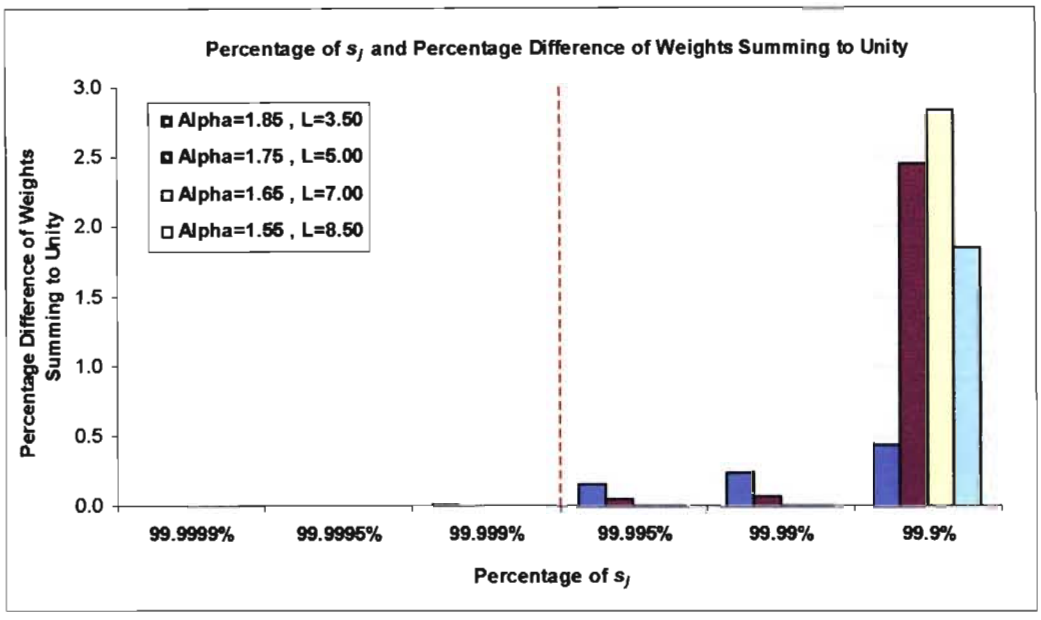


Figure 3.5: Percentage difference of Kriging weights summing to unity and the percentage of s_j that correspond to the point where the singular values are trimmed.

The trimming of the singular values at the 99.995% cumulative sum value corresponds to an approximately 50% reduction in size of the coefficient matrix as indicated by Figure 3.6. This has the effect of eliminating the singular values that cause the ill-conditioning of the coefficient matrix but still retaining the w_i values which have significant magnitude.

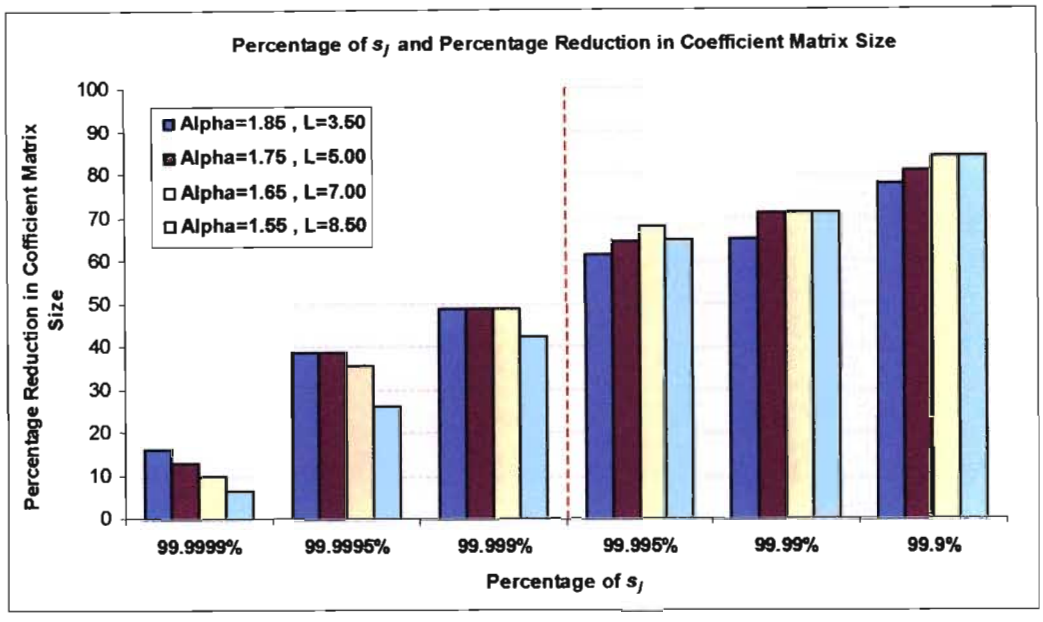


Figure 3.6: Percentage of s_j and corresponding percentage reduction in the size of the coefficient matrix.

Based on the above tests it was decided in the 2D infilling procedures that once a cumulative sum of percentage of the w_i^2's values reaches or exceeds a value of 99.995% the remainder of the singular values are removed to ensure computational stability and efficiency. The removing of the singular values represents an approximately 50% reduction in the matrix size and a consequent approximate 8-fold reduction in computation time.

3.4 INFILLING GROUND CLUTTER IN TWO DIMENSIONS

There are several approaches that can be adopted for Kriging target data, some of which are:

- compute all target values together.
- separate the targets into contiguous sets and identify their boundary controls.
- use neighbourhood Kriging to estimate each target value individually.

The first alternative of computing all the targets together is not practical as it requires an exorbitant number of controls which would result in a numerically unstable coefficient matrix and lead to an impractically long computation time. The other two alternatives were investigated on 2D data sets, the details of which will be outlined in sub-sections 3.4.1 and 3.4.2.

In using a moving neighbourhood of selected controls or selecting controls within a certain range from the border of a region of a ground clutter segment, the number of controls selected is a crucial factor and the number needs to be kept as small as possible without decreasing the accuracy of the Kriging estimates.

To determine the optimum number of controls the following testing procedure was undertaken. A 30 km square section was selected from the 4 km CAPPI from an instantaneous reflectivity image from the Bethlehem weather radar. From the selected segment a portion was removed to simulate a situation where ground clutter exists. The segment of missing data was then infilled using different combinations of α and L values representative of rainfall types ranging from convective to stratiform and also with different combinations of controls ranging from a using a large number of controls to very few. The SSE between the original and infilled data was then computed to determine the optimum number of controls. Figure 3.7 presents an illustration of the testing procedure employed. The results for each nearest neighbour approach tested will be outlined in the following two sub-sections.

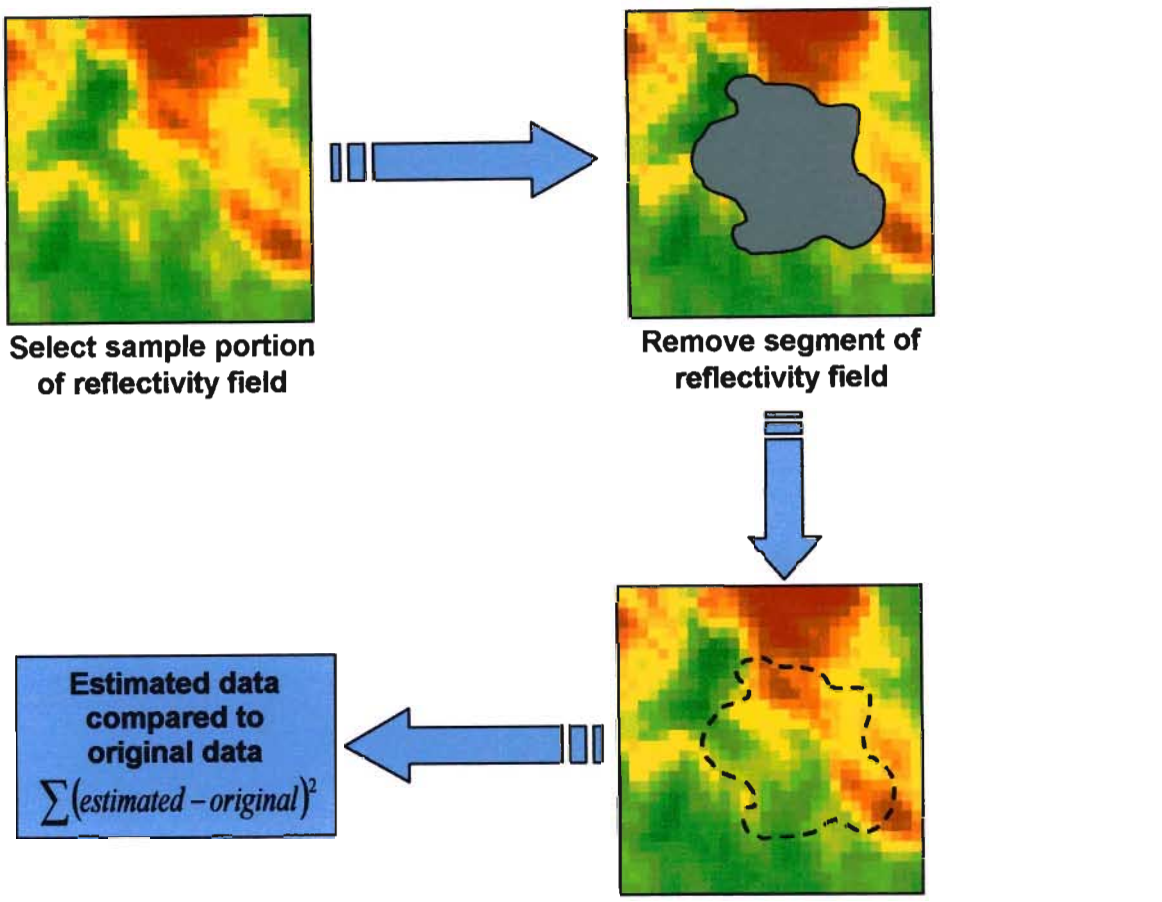


Figure 3.7: Flow chart of one step in the test to determine the optimum number of controls in the BTNN and ITNN techniques in two dimensions.

3.4.1 Individual Target's Nearest Neighbours (ITNN) Technique

The full details of the ITNN technique are described in Wesson & Pegram (2004), contained in Section 3.6, page 52. The technique works by infilling individual targets in turn using a moving neighbourhood approach where once a target pixel has been identified the 20 nearest controls are identified and used to infill the target. Constant semivariogram parameters of α = 1.5 and L = 11 pixels were used to infill all ground clutter pixels.

The testing procedure described at the beginning of Section 3.4 was used to determine the optimum number of controls for the ITNN ground clutter infilling algorithm. This was carried out with a range of semivariogram parameters representing rainfall ranging from convective to stratiform α = 1.85, 1.75, 1.65 and 1.55 with the corresponding L = 3.50,

5.00, 7.00 and 8.50 pixels. The numbers of controls used to infill individual targets were of the order of 44, 36, 28, 20, 12, 8 and 4. In each instance the SSE between the observed and infilled data was computed. Figure 3.8 indicates that there is no significant difference in the Kriged estimates when using 44, 36, 28 or 20 controls; however the SSE increases dramatically when using less than this number (using 12, 8 or 4 controls). It was therefore decided that the optimum number of controls to use was 20 in order to ensure computational efficiency and accuracy.

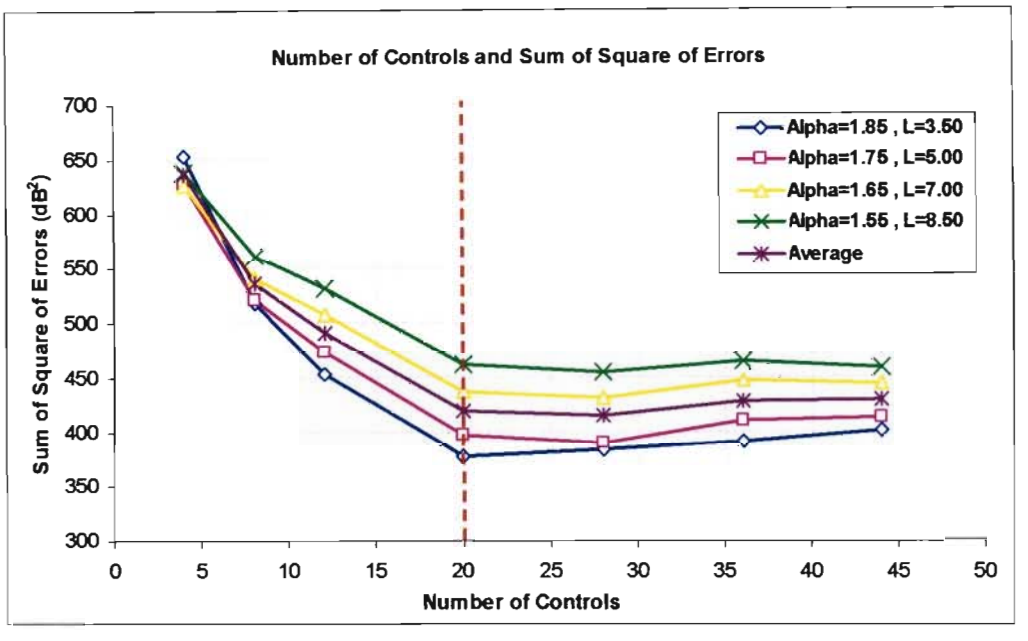


Figure 3.8: SSE between observed and infilled reflectivity data for different combinations of α and L semivariogram parameters and numbers of controls.

The decrease in computation time corresponding to the decrease in the number of controls used to infill individual targets is indicated in Figure 3.9. The time to infill this particular ground clutter segment consisting of 122 targets was 0.47 seconds when using 44 controls. The time decreases somewhat exponentially as the number of controls decreases; using the computed optimum number (20) of controls to infill the missing data corresponds to 0.07 seconds, which is nearly a 7-fold decrease in computation time.

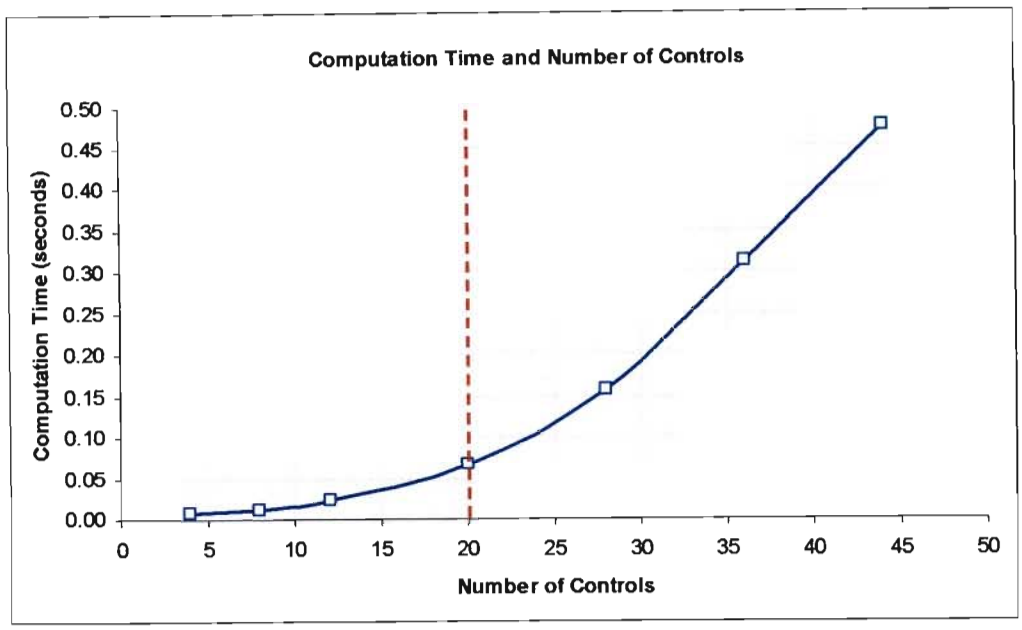


Figure 3.9: Computation time in seconds to infill all missing data with different numbers of controls when infilling individual targets.

3.4.2 Border Tracing Nearest Neighbours (BTNN) Technique

The alternative approach tested is also a technique that exploits the 'screening effect' like the ITNN approach, however it differs in that instead of individual targets being infilled in turn, contiguous segments are estimated at a time. The BTNN technique employs a border-tracing algorithm to trace contiguous segments of ground clutter and simultaneously select controls within a set radius of the edge of the ground clutter segment. Once the segment has been completely traced all the targets are infilled for that segment simultaneously. A more detailed explanation of the BTNN technique is described in Wesson and Pegram (2004) contained in Section 3.6, pages 52 to 54.

The testing procedure described at the beginning of Section 3.4 was used to determine the optimum radius to use when selecting the controls contained adjacent to targets selected. The same range of semivariogram parameters (α and L) was once again employed as used in the ITNN algorithm. The number of controls used to infill the data was selected in a range of 6, 5, 4, 3, 2.5, 2 and 1 pixel distance from the edge of the ground clutter segment to be infilled. In order to cope with the large coefficient matrices that require inversion, an Inversion by Partitioning Technique was employed which has the effect of reducing the inversion process from a N^3 process to $¼ \cdot N^3$ operation.

As illustrated in Figure 3.10 from a pixel radius of 6 to 2.5 there is no significant difference in the SSE between the observed and estimated data. Below the radius of 2.5 pixels however the SSE between the observed and estimated data does start to become significantly different. It was therefore decided that the optimum pixel radius to use in this instance is 2.5 pixels to obtain computational efficiency and maintain a significant degree of accuracy.

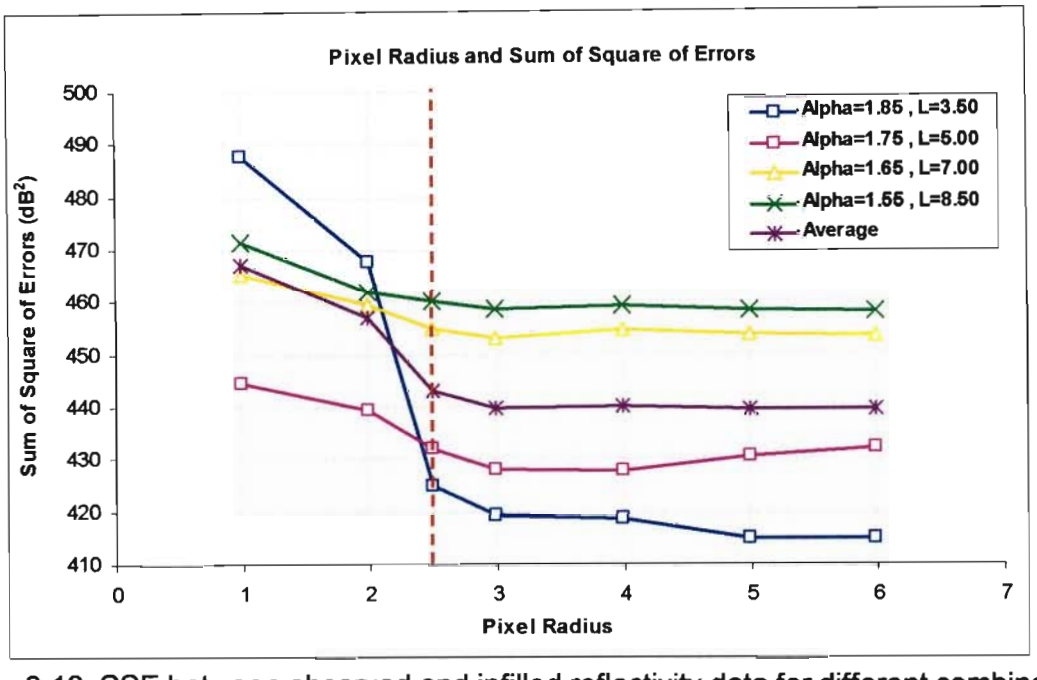


Figure 3.10: SSE between observed and infilled reflectivity data for different combinations of α and L semivariogram parameters and controls selected within a specified radius from the ground clutter segment edge.

The decrease in computation time corresponding to the decrease in the pixel radius and total number of controls is illustrated in Figure 3.11. There is a considerable decrease in computation time from the pixel radius of 6 to 2.5, approximately a 17-fold decrease. The computation times in this instance were found to be significantly higher than the corresponding times for the ITNN technique.

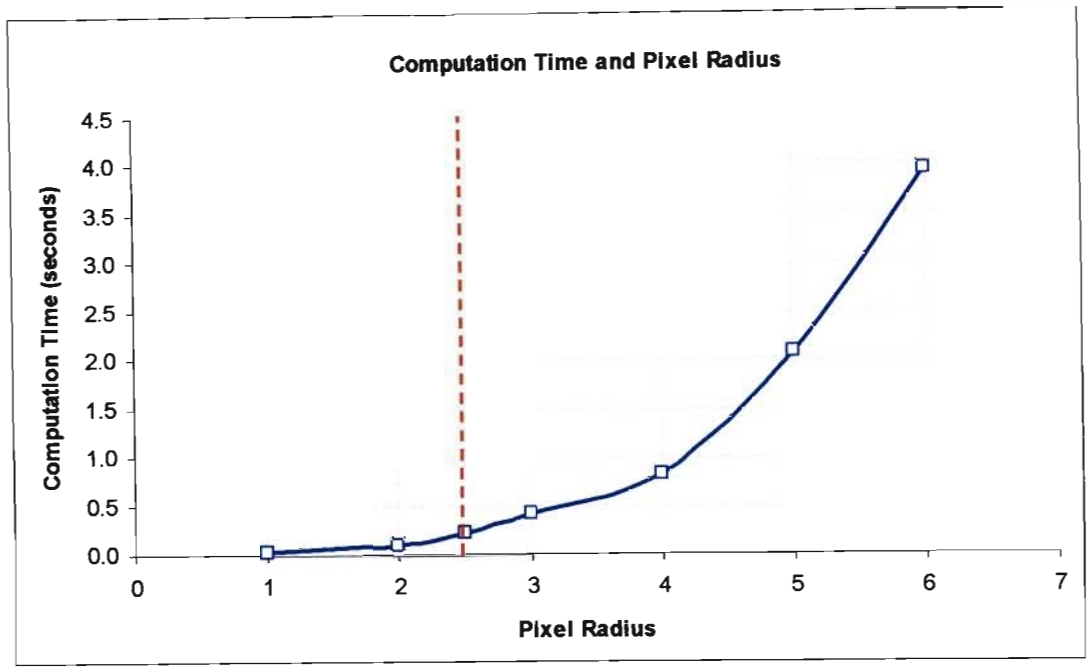


Figure 3.11: Computation time in seconds to infill all missing data with corresponding pixel radius used in selecting controls to infill all missing data contained in the ground clutter segment.

3.4.3 Comparison of the Two Nearest Neighbour Infilling Techniques

A comparison of the two techniques, ITNN and BTNN, was carried out on an instantaneous reflectivity image taken from the Durban weather radar (18 November 2000). The Bethlehem ground clutter map was then superimposed onto a region of the image where the reflectivity values were known. The ground clutter was then infilled with both the ITNN and BTNN techniques.

The infilled data was then compared to the original data and the SSE computed and time to infill the missing data recorded for both the ITNN and BTNN technique. It was found that the ITNN technique was considerably faster and more accurate. The details of the testing and results are contained in the paper in Section 3.6, pages 54 to 56.

It was decided, based on these results, that the ITNN technique would be adopted and developed further in order to infill ground clutter and missing data contained in radar volume scans due to the following reasons:

- The computation time is considerably faster than the BTNN technique.
- ITNN returns more accurate results than the BTNN technique.

- The ITNN technique can be far more easily adapted to work in 3D space, whereas the BTTN technique would require complex 3D border tracing techniques algorithms to be devised.
- The ITNN coefficient matrices are likely to be far more stable than the coefficient matrices used in the BTNN technique due to the larger number of controls used in the BTNN technique and hence larger coefficient matrices.

3.5 RAINFALL ESTIMATION AT GROUND LEVEL

A preliminary example of how Kriging can be used to extrapolate radar reflectivity data from aloft to ground level is presented in Wesson and Pegram (2004), page 56. In later sections of this document this topic will be reviewed in far greater detail and a validation between the raingauge accumulation estimates at ground level and radar accumulation estimates at ground level compared.

3.6 JOURNAL PAPER – *RADAR RAINFALL IMAGE REPAIR TECHNIQUES*

The journal paper published by Wesson and Pegram in *Hydrology and Earth System Sciences* (HESS) in 2004, volume 8(2), pages 220 - 234 is now presented.

Hydrology and Earth System Sciences, 8(2), *220–234* (2004) © EGU

Hydrology & Earth System Sciences

Radar rainfall image repair techniques

Stephen M. Wesson and Geoffrey G.S. Pegram

Civil Engineering, University of KwaZulu-Natal, Durban, 4041, South Africa

Email for corresponding author: wessons@ukzn.ac.za and pegram@ukzn.ac.za

Abstract

There are various quality problems associated with radar rainfall data viewed in images that include ground clutter, beam blocking and anomalous propagation, to name a few. To obtain the best rainfall estimate possible, techniques for removing ground clutter (non-meteorological echoes that influence radar data quality) on 2-D radar rainfall image data sets are presented here. These techniques concentrate on repairing the images in both a computationally fast and accurate manner, and are nearest neighbour techniques of two sub-types: Individual Target and Border Tracing. The contaminated data is estimated through Kriging, considered the optimal technique for the spatial interpolation of Gaussian data, where the 'screening effect' that occurs with the Kriging weighting distribution around target points is exploited to ensure computational efficiency. Matrix rank reduction techniques in combination with Singular Value Decomposition (SVD) are also suggested for finding an efficient solution to the Kriging Equations which can cope with near singular systems. Rainfall estimation at ground level from radar rainfall volume scan data is of interest and importance in earth bound applications such as hydrology and agriculture. As an extension of the above, Ordinary Kriging is applied to three-dimensional radar rainfall data to estimate rainfall rate at ground level.

Keywords: ground clutter, data infilling, Ordinary Kriging, nearest neighbours, Singular Value Decomposition, border tracing, computation time, ground level rainfall estimation

Introduction

Raingauges have traditionally been used for the recording of rainfall over catchment areas and are often regarded as providing the 'true' rainfall at ground level. They provide a direct measurement of rainfall depth at a point and are relatively easy and cheap to maintain. Limitations do however exist with the raingauge data. Raingauges provide only a point accumulation measurement of the rainfall and fail to capture its spatial variability; even with a dense network of raingauges it is difficult to interpolate and extrapolate gauge data in any significant detail or accuracy, especially at short time scales.

Radar images of instantaneous rain-rate address some of the problems and limitations associated with raingauge data. Rainfall can vary greatly in both space and time. Radar rainfall images can show the instantaneous rainfall rate in far greater spatial detail and complexity than is possible with raingauge data. With this type of information it is also possible to observe and predict severe weather patterns in a far more timely and efficient manner. At present in South Africa, the weather radar network of 11 radars provides instantaneous rainfall images of rain rate at approximately five-minute intervals and with one minute of arc (about 1.5 kilometres over South Africa) resolution.

There is a down side: there are various data quality issues associated with the radar rainfall data. Most importantly, radar provides an indirect measurement of the instantaneous rainfall rate. The reflectivity (dBZ) values returned to the radar are typically converted to a rainfall rate (mm hr^{-1}) with a variant of the Marshall Palmer formula (Marshall and Palmer, 1948). Errors in the radar rainfall data can include the following: ground clutter, beam blocking, bright band and anomalous propagation, to name a few. Ground clutter (non-meteorological echoes that influence radar data quality) is caused by the radar beam colliding with the earth's surface yielding abnormally high reflectivity values. This can result in large portions of the rainfall image being contaminated with highly reflective non-rainfall information; ground clutter is the main problem addressed herein.

The location of the ground clutter is identified by the South

African Weather Services Research Division, METSYS (Meteorological Systems and Technology), using an auto-correlation technique. In this technique the variability of reflectivity at ground clutter locations is considered to be much lower than that of precipitation. To detect this difference, the temporal auto-correlation is calculated at each pixel and once a determined threshold value of persistence is reached, the pixel is flagged as containing ground clutter. In this way the likely non-rainfall information can be marked and separated from the actual rainfall values (Visser, 2003). Figure 1 shows a typical radar image of widespread relatively low intensity rainfall at Bethlehem, South Africa, in summer. Where there was no rain recorded, the image background is white; the maximum observed reflectivity is given as yellow and does not exceed 30 dBZ, which is equivalent to approximately 3 mm hr^{-1}. The estimated ground clutter pixels are flagged, appearing black in this image.

To remove the ground clutter and estimate the missing rainfall data in real time, a computationally fast and accurate technique is required. Kriging was chosen as the method for estimating the missing data. However, for Kriging to be used in a real time application, advantage needs to be taken of various factors such as the pattern of the Kriging weights in the 'good' or control data pixels surrounding the ground clutter. Hence, an effective and efficient method for estimating the missing target rainfall values can be provided.

Standard techniques to remove ground clutter include subtracting a 'clutter map', derived from a period of no precipitation, from the observed reflectivity values. Unfortunately ground clutter is not totally stationary and its location can vary depending on atmospheric conditions (Vezzani, 1994).

The Kriging technique mentioned briefly above, and described in more detail later, is proposed as an alternative. The methodology is presented in this paper along with relevant examples of its application. Preliminary results are encouraging, however the technique still needs to be fully validated on different types of rainfall.

Brief overview of Ordinary Kriging theory

Kriging was chosen as the computational method for infilling the contaminated rainfall data. Kriging is used to estimate the missing data as it is the optimal technique for the spatial interpolation of Gaussian data (Cressie, 1993: 106) if the sum of estimation errors squared is to be minimised. Reflectivity values of rainfall have been shown to be approximately normally distributed, because the rain rate on an image follows a truncated lognormal distribution (Bell, 1987; Pegram and Clothier, 2001). Thus, in this application, the estimation of the missing data by Kriging is carried out in the reflectivity rather than the rain rate domain. Kriging has several advantages over other interpolation techniques. The basis function used in Kriging is determined by the data set and its spatial structure; in this way the actual nature of the data set is taken into account. In other well known interpolation techniques such as multiquadrics, Fourier series or splines the basis function is merely chosen for computational convenience.

Ordinary Kriging is used here where the mean is assumed unknown and is implicitly estimated from the data. The Ordinary Kriging equations used herein are given in Eqn. (1).

$$z(s_0) = \lambda^T(s_0) \cdot z \tag{1}$$

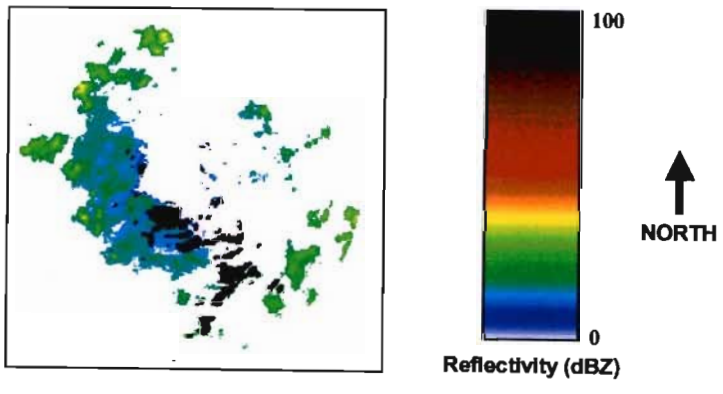


Fig. 1. *Typical instantaneous radar rainfall reflectivity image from the Bethlehem weather radar (20 February 2003). Black areas indicate contaminated data. The square surrounding the radar image has sides of length 270 kilometres.*

where $z(s_0)$ is the estimated data value at the target location $s_0 = (x_0, y_0)$, $\lambda(s_0)$ is the vector of Kriging weights associated with the location s_0 and z is the vector of known or control data. This computation is carried out to estimate the reflectivity at each target pixel in turn. The weights are computed from Eqn. (2):

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda(s_0) \\ \mu(s_0) \end{bmatrix} = \begin{bmatrix} g(s_0) \\ 1 \end{bmatrix} \quad (2)$$

where u is a unit vector of ones and μ a Lagrange multiplier. In Eqn. (2) $g(s_0)$ is the vector of covariances or semi-variogram values between the target location (data point to be estimated) and the control points (known data values z) and G is the matrix of covariances or semi-variogram values between the control points (the covariance or semi-variogram function is estimated from the data). Equation (2) is derived by minimising the estimation error variance and the Lagrange multiplier is included to constrain the sum of the weights to unity. The variogram is defined as the expectation of the square of the differences of the field variables as shown in Eqn. (3):

$$2 \cdot \gamma(x,h) = E\left\{ [Z(x) - Z(x+h)]^2 \right\} \quad (3)$$

the semi-variogram is defined as $\gamma(x, h)$ (Journel and Huijbregts, 1978: 11).

In a stationary random field, Ordinary Kriging can be carried out by using either a covariance or semi-variogram function, where one is the complement of the other with reference to the field variance. Mathematically they are equivalent in this context, however in some applications the variogram is more robust (Cressie, 1993: 70–73), so was used throughout in preference to the covariance. The isotropic semi-variogram model chosen is the two-parameter exponential model specified in Eqn. (4) and is used in all Kriging computations herein.

$$g(s) = 1 - \exp[-(s/L)^{\alpha}] \quad (4)$$

where s is the Euclidian distance between data points, L is a scale parameter called the correlation length and α is the exponent parameter, which lies in the range $0 < \alpha \leq 2$; the model is thus Gaussian when $\alpha = 2$.

The advantages of the Kriging technique can be listed as follows: it is considered to be the best linear unbiased spatial estimator (Journel and Huijbregts, 1978: 57, 304); in extrapolation the values converge to the mean of the field whereas other techniques do not necessarily exhibit this desirable behaviour; the technique can be easily used on 1-D, 2-D and 3-D data sets; the precision of the estimated data can be obtained. The precision is termed as the Kriging variance which, when using a semi-variogram function, is computed by Eqn. (5) (Cressie, 1993: 122–123):

$$\sigma_k^2(s_0) = \lambda^T(s_0) \cdot g(s_0) + \mu(s_0) \quad (5)$$

The main disadvantage of the Kriging technique is that it relies on finding the solution to a set of linear equations so that in large data sets the matrix algebra can become computationally burdensome and time consuming. Not as well recognised, and one of the main thrusts of this paper, is that the coefficient matrix can be highly ill-conditioned depending on the chosen parameters in the semi-variogram function and the data configuration. This can lead to inaccurate solutions and numerical instability. These matters will be dealt with in more detail in the sequel.

Methods for improving computational efficiency

In weather radar applications in South Africa where data sets are processed at a frequency of 5 minutes, the missing rainfall data need to be estimated in real time, thus a computationally fast and accurate technique is needed. The classical Kriging technique requires the solution of a linear system of equations whose size is determined by the number of control points. This can be time consuming, as the coefficient matrix becomes large. A radar image typically contains 120 000 pixels, leading to a coefficient matrix with dimensions greater then 100 000 by 100 000. Currently it is impractical to work with systems this size online.

There is an unexpected saving which can be exploited to ease the computational burden in a dense data set, as in radar data which typically occur on a lattice, not all of the control points need to be used in estimating the missing data points. A 'screening effect' occurs where the significant Kriging weights associated with the controls are concentrated around the target points (Chiles and Delfiner, 1999: 202–206). This can be used advantageously to reduce computation time significantly. For each individual target point, or set of target points in a cluster, a Kriging neighbourhood in its vicinity can be selected to reduce the dimensionality of the problem.

THEORETICAL JUSTIFICATION OF THE SCREENING EFFECT IN ONE DIMENSION

To give some theoretical justification to the 'screening effect', a one-dimensional analysis is presented which shows this very nicely. This does not easily extend itself to two dimensions or more, but intuitively and computationally the idea is justified by extension from a single dimension. It

turns out that, in one dimension, the screening effect is limited by the order of the model in cases where the co-variogram is that of an autoregressive process of order p. In fact only the first p intact (control) data nearest the point to be infilled (target) will have non-zero weights λ in the Kriging equation $z(s_0) = \lambda^T(s_0)\cdot\underline{z}$. A small one-dimensional example will suffice to explain the ideas.

One-dimensional AR(1) and AR(2) models

Suppose there is a sample $\{z_i, i = 1, 2, \ldots, 6\}$ with two missing values at positions $i = 3$ and 4 which are to be estimated. Assume the sample is a realisation of a zero-mean unit-variance Gaussian AR(1) process with parameter ϕ: $z_i = \phi z_{i-1} + a_i$ and (because of the reversibility of the AR(1) process) $z_i = \phi z_{i+1} + e_i$ where $\{a_t\}$ and $\{e_t\}$ are independent white noise processes with the same variance. To attempt to infill the missing data using a forecast from z_2, would give $\{\hat{z}_3, \hat{z}_4\} = \{\phi z_2, \phi^2 z_2\}$ and hindcasting from z_5, $\{\hat{z}_3, \hat{z}_4\} = \{\phi^2 z_5, \phi z_5\}$, with no prospect of agreement. Any compromise would be subjective and *ad hoc*, unless, for example, an iterative scheme embodying the EM algorithm were to be employed.

As an optimal alternative, use Kriging to estimate the missing values. The correlation matrix between the known values is then formed from the usual 6×6 matrix and removing the two central rows and columns:

$$G_{11} = \begin{bmatrix} 1 & \phi & \phi^4 & \phi^5 \\ \phi & 1 & \phi^3 & \phi^4 \\ \phi^4 & \phi^3 & 1 & \phi \\ \phi^5 & \phi^4 & \phi & 1 \end{bmatrix}$$

whose inverse is:

$$G_{11}^{-1} = \begin{bmatrix} (1-\phi^6) & -\phi(1-\phi^6) & 0 & 0 \\ -\phi(1-\phi^6) & (1-\phi^8) & -\phi^3(1-\phi^2) & 0 \\ 0 & -\phi^3(1-\phi^2) & (1-\phi^8) & -\phi(1-\phi^6) \\ 0 & 0 & -\phi(1-\phi^6) & (1-\phi^6) \end{bmatrix} \div [(1-\phi^2)(1-\phi^6)]$$

$$= \begin{bmatrix} c & a & & \\ a & \beta & \alpha & \\ & \alpha & \beta & a \\ & & a & c \end{bmatrix}$$

To get the matrix of Kriging weights Λ^T one needs the correlation matrix between the two missing and four known values:

$$G_{21} = \begin{bmatrix} \phi^2 & \phi & \phi^2 & \phi^3 \\ \phi^3 & \phi^2 & \phi & \phi^2 \end{bmatrix}$$

so that because of the symmetry in this case

$$\Lambda^T = \begin{bmatrix} 0 & \phi(1-\phi^4) & \phi^2(1-\phi^2) & 0 \\ 0 & \phi^2(1-\phi^2) & \phi(1-\phi^4) & 0 \end{bmatrix} \div [1-\phi^6]$$

The first thing to note is that only the observations contiguous to the gap have non-zero weights, so are the only observed data involved in the estimation. The second thing to note is that the inverse of G_{11} is a tridiagonal matrix, whose form can be exploited to find explicit expressions for its elements in the general 1-D case.

In general, for a sequence of n complete observations from an AR(1) process, the correlation matrix is $R_n = \{\phi^{|i-j|}, i, j = 1, 2, \ldots, n\}$. Its inverse is a tridiagonal matrix whose upper and lower diagonals are filled with equal elements $a = -\phi/(1-\phi^2)$, the main diagonal has equal elements $b = (1+\phi^2)/(1-\phi^2)$, except for the first and last which equal $c = 1/(1-\phi^2)$.

It is not difficult to show that when a block of data of width w is missing after the m^{th} observed value, then the correlation matrix of the surviving variables becomes

$$R^*(n,m,w) = \begin{bmatrix} R_m & Q \\ Q^T & R_{n-m-w} \end{bmatrix}$$

where

$$Q = \begin{bmatrix} \phi^{m+w} & \phi^{m+w+1} & - & \phi^{n-1} \\ \phi^{m+w-1} & \phi^{m+w} & - & \phi^{n-2} \\ - & - & - & - \\ \phi^{w+1} & \phi^{w+2} & - & \phi^{n-m} \end{bmatrix}$$

and where R_m has the same form as R_n. The inverse of $R^*(n,m,w)$ will be a tridiagonal matrix of the same form as R_n^{-1} with identical diagonal elements except for the four elements either side of the gap, where instead of α and β, the elements will be a and b, which after some tedious but straightforward algebra are found to be equal to:

$$\alpha = (1-\phi^{2w+4})/[(1-\phi^2)(1-\phi^{2w+2})]$$
$$\beta = -\phi^{w+1}/(1-\phi^{2w+2})$$

which are seen to collapse to a and b when $w = 0$ (no gap).

When G_{11}^{-1} is pre-multiplied by G_{21}^T to give Λ^T, only the weights of the elements either side of the gaps will be non-zero. This happens in general and not only in the AR(1) example above. The behaviour extends to AR(p) models, where now the size of the screen is p elements deep either side of the gaps.

Other combinations of gaps in the data still produce a tridiagonal inverse whose elements either side of the gaps depend on the width of the gaps as well as the width of the intact elements between the gaps. A small-dimensional numerical example will demonstrate the ideas. In a sequence of 11 AR(1) values with $\phi = 0.5$, the targets are at $i = 3, 4, 6$ and 7. The G_{11} matrix, its inverse, G_{21} and the resulting matrix of Kriging weights Λ^T (where the solid lines indicate the gaps) is:

G_{11} =

element	1	2	5	7	8	10	11
1	1	0.5	0.0625	0.015625	0.007813	0.001953	0.000977
2	0.5	1	0.125	0.03125	0.015625	0.003906	0.001953
5	0.0625	0.125	1	0.25	0.125	0.03125	0.015625
7	0.015625	0.03125	0.25	1	0.5	0.125	0.0625
8	0.007813	0.015625	0.125	0.5	1	0.25	0.125
10	0.001953	0.003906	0.03125	0.125	0.25	1	0.5
11	0.000977	0.001953	0.015625	0.0625	0.125	0.5	1

G_{21} =

element	1	2	5	7	8	10	11
3	0.25	0.5	0.25	0.0625	0.0313	0.0078	0.0039
4	0.125	0.25	0.5	0.125	0.0625	0.0156	0.0078
6	0.0312	0.0625	0.5	0.5	0.25	0.0625	0.0312
9	0.0039	0.0078	0.0625	0.25	0.5	0.5	0.25

Λ^T =

element	1	2	5	7	8	10	11
3		0.4761	0.1904				
4		0.1904	0.4761				
6			0.4	0.4			
9					0.4	0.4	

The only non-zero weights are those corresponding to the observations contiguous to the gaps, as asserted.

The ideas carry over to an AR(p) model as expected. This is useful because an AR(2) model is quite flexible for the purposes of describing the correlation models encountered in some random fields, particularly rainfields measured by radar. For example, given 14 values (with gaps at $i = 2$, 8 and 9) sampled from an AR(2) model: $z_i = \phi_1 z_{i-1} + \phi_2 z_{i-2} + a_i$ with $\phi_1 = 1.10$ and $\phi_2 = -0.32$, the inverse of G_{11} is tridiagonal as asserted and the Kriging weight matrix Λ^T is:

element	1	3	4	5	6	7	10	11	12-14
2	0.4977	0.6570	-0.1448						
8					-0.2285	0.8932	0.4225	-0.1435	
9					-0.1435	0.4225	0.8932	-0.2285	

It will be noted that the weights in the second spaces from the gap are negative for the AR(2) case with these parameters. It is also noted that this behaviour (negative weights) is found when Kriging with Gaussian shaped correlograms. Nevertheless, even in the Gaussian case, the non-zero weights are confined to a zone close to the gaps.

An interim conclusion is that in the particular case of equally spaced data in one dimension, *only the observations contiguous to the gaps influence the infilling of the missing data*. This is likely to have greater economies in higher dimensions, as will now be demonstrated computationally.

DEMONSTRATING THE TWO DIMENSIONAL SCREENING EFFECT BY COMPUTATION

An example of the 'screening effect' is illustrated in Fig. 2, which depicts a small dimensional example on a set of 8 × 9 pixels with four (marked by an asterisks) missing in the interior. The semi-variogram model used had an α = 1,5 and a correlation length of L = 11 pixels. (Note in passing that the value of L = 11 pixels, of dimension 1.5 kilometres, is an average value determined from observed radar rainfield intensities for a wide range of weather types and this value will be used in the sequel; it is not the purpose of this paper to deal with estimation of co-variograms). The weights associated with the 68 control points for each of the four target points were computed one set at a time. These weights were then summed for all four targets and their totals appear in the shaded pixels. It is seen that the significant Kriging weights are concentrated in the immediate vicinity of the target points and rapidly diminish in magnitude with distance. At a range of greater than two cells from a target point the Kriging weights are nearly zero. As can also be seen in Fig. 2, negative weighting values do occur some distance from the target. The Kriging weights could be forced to be non-negative but a disadvantage in doing this is that the Kriged estimates are then constrained to lie within the minimum and maximum values of the selected control data (Chiles and Delfiner, 1999: 224), (Cressie, 1993: 143). In the rainfield infilling application it is more appropriate to allow for the Kriged estimates to range outside of the minimum and maximum values of the control data used.

By considering only the control points with significant Kriging weights a considerable reduction in size of the coefficient matrix can be made in cases where the target points are in clusters contained within an otherwise complete set of controls. With this reduction in size of the control set a concomitant decrease in computation time can be achieved with little or no loss in the accuracy of the final results.

0	0	0	0	0	0	0	0	0
0	0	-0.03	-0.07	-0.01	0.02	0.01	0	0
0	-0.04	-0.02	0.38	-0.17	-0.18	-0.03	0	0
0	-0.14	0.54	*	1.33	0.40	0.00	-0.03	0
0	-0.15	0.62	*	*	*	0.39	-0.08	0
0	-0.04	-0.04	0.61	0.75	0.49	-0.01	-0.03	0
0	0	-0.04	-0.16	-0.20	-0.14	-0.04	0	0
0	0	0.01	0.02	0.02	0.02	0.01	0	0
0	0	0	0	0	0	0	0	0

Fig. 2. *Sum of all Kriging weights concentrated around four unknown data points. The variogram used had an α = 1.5 with L = 11 pixels.*

ILL-CONDITIONING OF THE COEFFICIENT MATRIX

In the Ordinary Kriging computational procedure, a set of linear equations must be solved to determine the weights $\lambda(s_0)$. This solution is given by Eqn. (6), derived from Eqn. (2):

$$\left[\lambda^T(s_0) \quad \mu(s_0)\right] = \left[g^T(s_0) \quad 1\right] \cdot \begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix}^{-1} \qquad (6)$$

Solving for $\lambda(s_0)$ requires inversion of the coefficient matrix (usually the equations are solved using an efficient method based on LU decomposition). It was found that the coefficient matrix can be highly ill-conditioned depending on (i) its size, (ii) the chosen parameterisation in the semi-variogram function (iii) the layout of the data set on a lattice and (iv) the ratio (s/L) as used in the semi-variogram model, Eqn. (4), where a decrease in the ratio results in the coefficient matrix becoming more ill-conditioned. The ill-conditioning is most sensitive to the α-parameter, the exponent in the semi-variogram function defined in Eqn. (4). As α increases from a value close to zero towards a value of two (converging towards a Gaussian semi-variogram function) the coefficient matrix becomes increasingly ill-conditioned. When α = 2 in the semi-variogram function, and the coefficient matrix is above a certain size, (as small as 68 × 68, from the 8 × 9 region of points shown in Fig. 2) the coefficient matrix is essentially singular. As indicated in Fig. 3 the determinant of the coefficient matrix (plotted on a log scale) decreases steadily as $\alpha \rightarrow 2$. The determinant was computed for a relatively small 50 by 50 coefficient matrix, with the semi-variogram function having a correlation length of L = 11 pixels.

Although the Gaussian semi-variogram results in a near singular coefficient matrix, it is still commonly used in practical Kriging applications (e.g. Todini, 2001), albeit for sparsely distributed control points. It is difficult to determine at what size the coefficient matrix does become noticeably unstable since this appears to be largely determined by factors such as the layout pattern of the control data points. However for a coefficient matrix in excess of 40 by 40 in size, and with α = 2, the matrix can be ill-conditioned enough to return nonsensical weights. In such cases, conventional methods of computing the inverse coefficient matrix such as Gauss-Jordan and its derivatives cannot be used. The method of Singular Value Decomposition (SVD) can be employed with advantage, however, to determine an accurate and meaningful solution for the weights, despite the ill-conditioning problem described above.

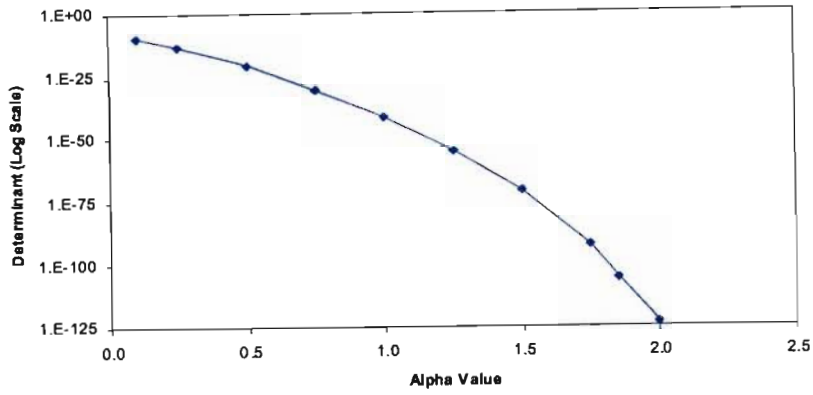


Fig. 3. *Semi-variogram exponent α and corresponding determinant of coefficient matrix (size 50 ×50) for L = 11 pixels.*

THEORETICAL JUSTIFICATION OF MATRIX RANK REDUCTION TECHNIQUE

Singular Value Decomposition involves the decomposition of a rectangular matrix into unique matrix 'factors'; an overview of the method is readily available in Press *et al.* (1992: 59–67) which is briefly summarised here. The matrix is decomposed into a column orthogonal matrix U, a diagonal matrix W that contains the non-negative singular values along its diagonal and the transpose of an orthogonal matrix V. The decomposition of the Ordinary Kriging coefficient matrix is given in Eqn. (7).

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix} = \begin{pmatrix} & U & \end{pmatrix} \cdot \begin{pmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_n \end{pmatrix} \cdot \begin{pmatrix} & V^T & \end{pmatrix} \tag{7}$$

To compute the inverse, one simply needs to invert the diagonal elements of W and multiply out the matrices in the order indicated by Eqn. (8):

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix}^{-1} = V \cdot \left\{ diagonal\left(\frac{1}{w_j} \right) \right\} \cdot U^T \tag{8}$$

When the coefficient matrix is ill-conditioned, the diagonal values along the W matrix need to be carefully considered because the singular values which are near zero define the (near) null space of the matrix. To obtain a meaningful pseudo-inverse, the diagonal elements of W^{-1} are set to $1/w_j$ where w_j is above some chosen near-zero threshold and to zero otherwise. This reduces the rank of the matrix by the size of the (near) null space.

To demonstrate the ideas, a coefficient matrix of size 150 by 150 was computed from a sample reflectivity field. The coefficient matrix was then decomposed as shown in Eqn. (7). This was done several times for α values ranging from one to two. The singular values obtained in the W-matrix were ranked then plotted in Fig. 4, where it is seen that the w_j values rapidly diminish in magnitude along the diagonal and become close to zero as α goes from one to two (Exponential to Gaussian semi-variogram); the threshold of 10^{-15} reflects machine precision.

Since many of the singular values are significantly close to zero relative to w_1 which is approximately 60, the most appropriate action is to eliminate these values and replace them (and their inverses) with zeros instead. This solves the problem of ill-conditioning in the inverted coefficient matrix, yields useful accurate solutions and, incidentally, reduces computation time.

The Euclidian norm of the coefficient matrix is equal to the sum of the squares of the singular values; if the coefficient matrix were a covariance matrix this sum would also be equal to the variance, as indicated by Eqn. (9):

$$\sum_{j=1}^{n} w_j^2 = \sigma_x^2 \tag{9}$$

The variance is predominantly contained in the first few terms of the W matrix with the others contributing very little. Only the largest of the singular values need to be considered in most cases since their sum closely approximates the variance.

Once the appropriate number of w_j values has been trimmed and replaced with zero values, only the corresponding rows and columns need to be retained in the U and V matrices; as shown in Eqn. (10) the columns of the U matrix and the rows of the V^T matrix can be trimmed. This speeds up the computation dramatically with little or no change in the accuracy of the final results.

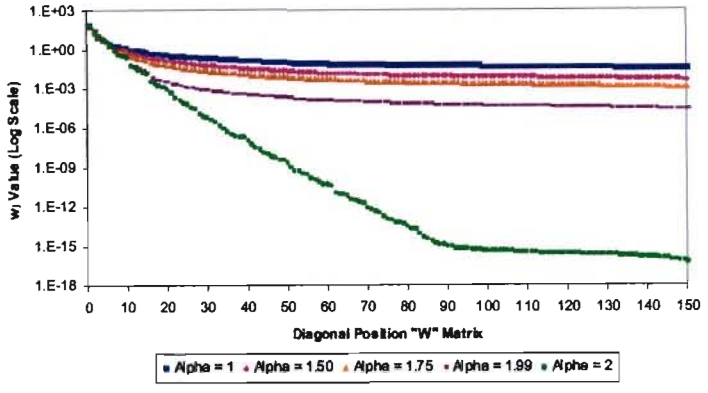


Fig. 4. *Singular values in W matrix for different values of semi-variogram exponent α; note the logarithmic scale. The coefficient matrix was 150 square and L was kept constant at 11 pixels.*

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix} = \begin{pmatrix} & U & \end{pmatrix} \cdot \begin{pmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_n \end{pmatrix} \cdot \begin{pmatrix} V^T \end{pmatrix} \qquad (10)$$

DEMONSTRATION OF ADVANTAGES OF MATRIX RANK REDUCTION BY COMPUTATION

As an example, a sample reflectivity field was selected and portions removed to simulate a ground clutter scenario. To infill the missing 57 target points, a 140 by 140 coefficient matrix had to be inverted. The semi-variogram parameters used were $\alpha = 1.5$ and L = 11 pixels. Initially, with none of the w_j values removed, the Kriging process (calculation of weights and estimation for each of the 57 target values) took approximately 5.6 seconds. The W matrix was then trimmed incrementally along with the corresponding rows and columns of the U and V^T matrix. It was found that up to 70% of the w_j values could be removed with no significant change in the final Kriged results (accords to the sum of the difference squared of hidden and computed targets) with a consequent decrease in time to 0.6 seconds as shown in Figs. 5 and 6. This corresponds to a nine-fold decrease in computation time with no concomitant change in accuracy achieved with the matrix rank reduction technique. The results would be more dramatic for a larger α.

An example of the numerical instabilities that can occur in the Kriging process is given in Fig. 7. A spatially correlated Gaussian random field (the upper part of Fig. 7) was generated with a known semi-variogram function on a 25 by 25 grid field. The semi-variogram parameters used were $\alpha = 1.5$, sill (ω) = 1, L = 1000 pixels (the one exception from L = 11, to emphasize the instability), nugget (p) = 0 and a variance (σ^2) = 1, as shown in Eqn. (11).

$$g(s_0) = \sigma^2 - \left\{ p + \omega\left[1 - \exp(-(s/L)^{\alpha})\right] \right\} \qquad (11)$$

Forty randomly scattered points were then selected from the correlated field as control points. Simple Kriging was then used to estimate the remaining points of the field from

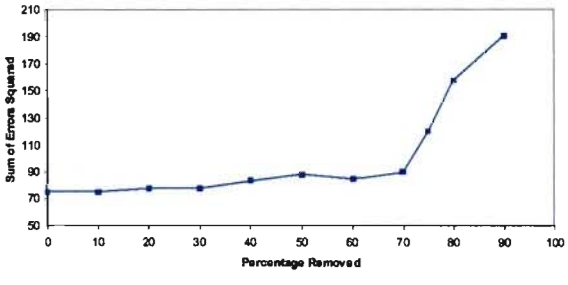


Fig. 5. *Sum of errors squared and corresponding percentage w_j removal, L = 11 pixels, α = 1.5, 140 controls (140 ×140 coefficient matrix).*

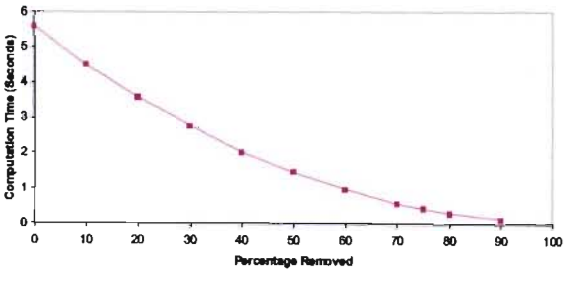


Fig. 6. *Computation time and corresponding percentage removal of singular values in a 140 × 140 coefficient matrix with L =11 pixels and α = 1.5.*

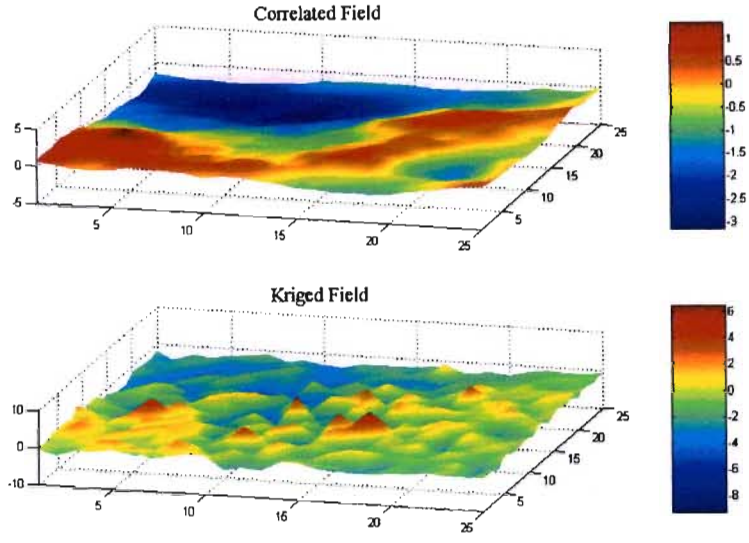


Fig. 7. *Simple Kriging without removal of near zero singular values. For definition of the variogram model, see text.*

the control points. As shown in comparing the two parts of Fig. 7, the Kriged field does not resemble the original correlated field at all due to the highly ill-conditioned coefficient matrix.

Figure 8 shows exactly the same scenario as in Fig. 7, however in this instance the trimming of the near zero singular values from the W matrix was carried out, as illustrated in Eqn. (10). The Kriged field is now a smoothed replica of the correlated field as desired. In Figs. 7 and 8 the 40 out of 250 points selected to perform the Kriging were randomly chosen. The α-value was midway between exponential and Gaussian, yet still the numerical instabilities occurred. This result is surprising and was thought to be important enough to publish as a warning to the unwary and to provide an antidote.

For typical radar reflectivity fields of diameter 250 pixels (about 50 000 data points), the above two techniques, nearest neighbours and matrix rank reduction need to be used in conjunction in order to achieve efficiency and stability. Image processing techniques also need to be taken advantage

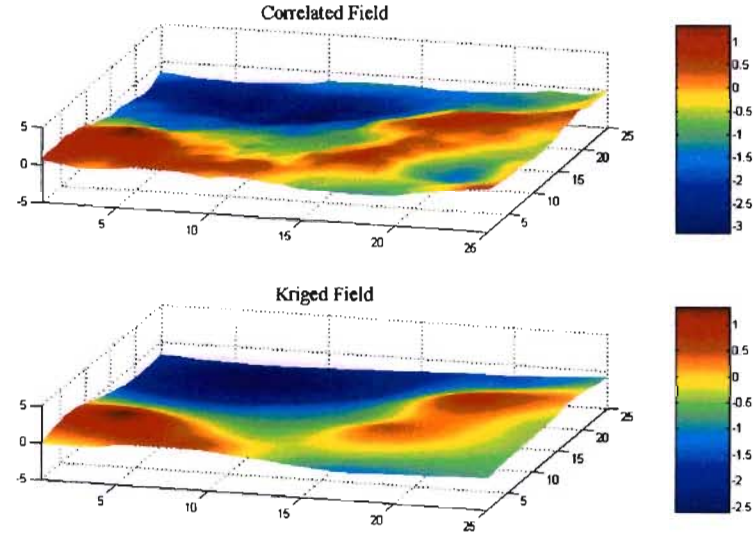


Fig. 8. *Simple Kriging with removal of near zero singular values.*

of so as to be able to manage the large data sets. Two techniques for ground clutter removal exploiting nearest neighbours will be compared in the next two sections; one works with individual target points in turn, the other by border tracing around a cluster of target points and solving the Kriging equations simultaneously.

Individual target's nearest neighbours (ITNN) technique

In the nearest neighbours technique applied to individual target points in turn (ITNN), the weighting distribution associated with Ordinary Kriging is taken into account in the sense that the 'screening effect' described in the previous section (where the significant weights at control points are concentrated in the immediate vicinity of the selected target data point) is exploited. In the algorithm suggested here, only the neighbours nearest to the target point are identified. This results in a considerable decrease in computation time and with no significant loss in accuracy. The SVD matrix rank reduction process is also carried out to reduce the computation time in finding the Kriging weights.

The ITNN algorithm can be described as follows for a two-dimensional data set. The radar reflectivity image is searched, pixel by pixel, starting at the top left-hand corner and working across each row then down the image, one row at a time, the ground clutter having been already flagged as –10 dBZ values. Once a ground clutter point is located, a search outwards from that point is initiated to locate the 20 nearest valid control points – the maximum would be 24 in a 5 × 5 square region if the target was an isolated pixel. Thus control values within two or so pixels of the target are selected. The search strategy proceeds as follows. Initially the rows directly above and below the point are searched, and then the columns on either side of that point. The search then moves incrementally outwards searching in turn rows and then columns until 20 uncontaminated neighbours are located. This strategy is illustrated in Fig. 9. The pixel marked with an asterisk is the target point initiating the first search on the image; other targets (yet to be infilled) are masked in grey.

Once the nearest neighbour controls are located they are tested to determine whether they are all zero: if they are, the target is set to zero. If the controls are all non-zero, Ordinary Kriging is carried out to estimate a value for the target.

Once the target point has been identified, an estimate of its value is calculated and then stored in a separate vector. The search resumes where it left off and continues searching across each row and then down the image, one row at a time, until the next ground clutter pixel is located. In this way the rainfall at each ground clutter pixel is estimated in a point by point manner unaffected by previous estimates until all the ground clutter pixels have been infilled. After the entire data set has been scanned the estimated reflectivity values, stored sequentially in a vector, are then inserted in place of the flagged ground clutter pixels.

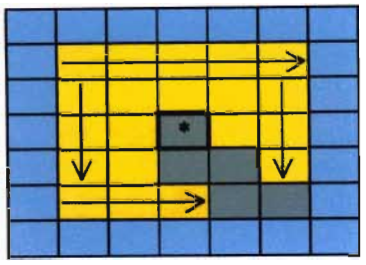


Fig. 9. *Nearest neighbours search strategy. Ground Clutter pixels marked as grey squares.*

As examples, two radar-rainfall reflectivity images in South Africa have been selected for repair, Fig. 10 from the Durban weather radar and Fig. 11 from the Bethlehem weather radar. On each of them the ground clutter pixels have been flagged with –10dBZ values, which appear as black segments in Figs. 10 and 11. The number of ground clutter targets is approximately 1400 (for both the Bethlehem and Durban data) and the time taken to estimate the missing data for one 400 by 400 pixel image is typically 0.5 to 2 seconds computed on a Pentium(R) 4 with a 2,40GHz CPU and 512MB of RAM. The images on the right of each pair are those which have been 'repaired', i.e. where the ground clutter values, or contaminated pixels, have been replaced with reflectivity values estimated by Kriging.

Border tracing nearest neighbours (BTNN) technique

An alternative for identifying segments which have to be infilled is the border tracing with nearest neighbours (BTNN) technique, where advantage is again taken of the distribution of the Kriging weights to reduce the computational load and improve computational efficiency by concentrating on the nearest neighbours around a cluster. In this method image processing algorithms were investigated to provide an alternative approach to infilling the missing data. In the ITNN technique the clutter was infilled point by point whereas in the BTNN technique the infilling is done a whole segment at a time, where a segment is one or more target pixels that are grouped together in a connected region.

Once again, the image is searched from the top left hand corner, row by row, moving downwards until a flagged

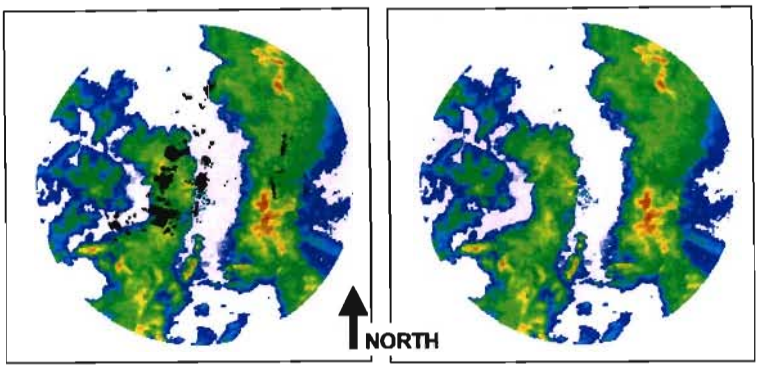


Fig. 10. *Durban radar rainfall reflectivity data before and after ground clutter removal (18 November 2000). Black areas indicate contaminated data. The square surrounding each radar image has sides of length approximately 270 by 270 kilometres.*

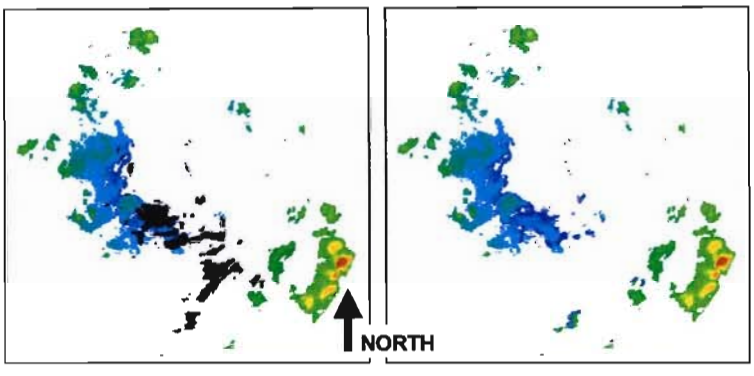


Fig. 11. *Bethlehem radar rainfall reflectivity data before and after ground clutter removal (25 February 2003). Black areas indicate contaminated data. The square surrounding each radar image has sides of length 270 kilometres.*

ground clutter pixel is located. An eight-connectivity border tracing technique (Sonka *et al.*, 1999: 142–145) is then implemented from that local origin. From that pixel the eight immediate surrounding neighbours are searched in an anticlockwise direction until another new ground clutter pixel is located. This becomes the new local origin and the process is repeated. The border is then traced in this manner until the original starting point is reached. Figure 12 (left) illustrates how the border is traced in an anticlockwise direction around the ground clutter segment, with the dashed lines indicating pixels tested and identified as controls in the process.

While the border is being traced, a Kriging neighbourhood two pixels deep from each border target is selected simultaneously. As indicated in Fig. 12, the control points within a range of 2.5 pixels are selected — this strategy neglects the corner pixels on a 5 × 5 square grid surrounding the target. All of the pixels selected in this manner are then identified as the Kriging control neighbourhood for the estimation of the cluster of target points in the particular segment. Once the border tracing has been completed, and the Kriging neighbourhood has been selected, an algorithm

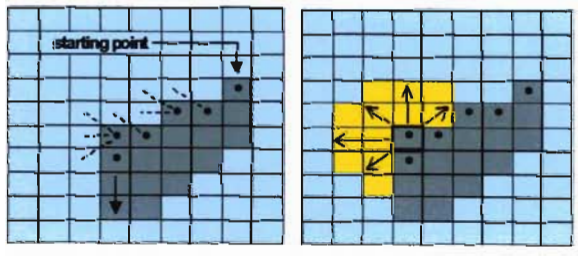


Fig. 12. Border tracing technique indicating trace direction and pixels tested in process (left image). Selection of Kriging neighbourhood for infilling of missing data (right image) (Sonka et al., 1999: 143).

identifies the positions of the remainder of the target points that are located in the interior of the traced contour.

The Kriging control neighbourhood is then tested to determine whether the controls are all zero. If this is the case, all of the target points in the segment are set to zero. If the data points are not all zero, ordinary Kriging is used to estimate values for all of the target points that make up that segment.

The solution to the Kriging equations is then computed by an inversion by partitioning technique (Aitken. 1967: 148) in conjunction with the SVD matrix rank reduction technique in an effort to decrease the computation time. The inversion by partitioning technique as employed here splits the coefficient matrix into four equal sub matrices, given by Eqn. (12).

$$A = \begin{bmatrix} P & Q \\ Q^T & S \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} \tilde{P} & \tilde{Q} \\ \tilde{Q}^T & \tilde{S} \end{bmatrix}$$

where:

$$\begin{aligned} \tilde{P} &= (P - Q \cdot S^{-1} \cdot Q^T)^{-1} \\ \tilde{Q} &= -(P - Q \cdot S^{-1} \cdot Q^T)^{-1} \cdot (Q \cdot S^{-1}) \\ \tilde{Q}^T &= -(S^{-1} \cdot Q^T) \cdot (P - Q \cdot S^{-1} \cdot Q^T)^{-1} \\ \tilde{S} &= S^{-1} + (S^{-1} \cdot Q^T) \cdot (P - Q \cdot S^{-1} \cdot Q^T)^{-1} \cdot (Q \cdot S^{-1}) \end{aligned} \qquad (12)$$

Further partitioning of the matrix into nine sub-matrices was tested but this did not result in a decrease in computation time. The inversion by partitioning technique was also found to be most effective for coefficient matrices with a size in excess of 30 by 30; for matrices below this size the SVD matrix rank reduction technique was computationally faster.

Once the target values for a cluster have been estimated they are inserted into the working data set immediately and the pixels flagged as known. The algorithm then carries on searching from the initial local origin that identified the now infilled target segment, row-by-row, moving across then down the image until the next Ground clutter pixel is located, at which point the process is repeated. Figure 13 is an example where the BTNN technique has been used to remove ground clutter, as before the left image contains the marked ground clutter and the right image the estimated reflectivity data.

To estimate the missing values in a typical radar rainfall image, the BTNN technique takes approximately 5 to 20 seconds. However, the time is highly dependent on the range used for selecting the Kriging neighbourhood for each segment. The greater the number of control points the greater the size of the coefficient matrix that needs to be inverted, where the computation time is dominated by the matrix inversion.

Comparison of the two nearest neighbour infilling techniques

An extended formal and detailed comparison of the two methods, ITNN and BTNN has yet to be done. However, as a preliminary example of such a comparison, on 18 November 2000, a radar reflectivity image from the Durban weather radar with no ground clutter but containing convective rainfall was selected. The Bethlehem ground clutter map (not coinciding with the Durban one) was then superimposed on this image, as indicated in Fig. 14. The missing reflectivity data were then estimated using both the ITNN and BTNN techniques. In this application of the ITNN technique the 20 nearest neighbours for each target were used and for the BTNN technique pixels within a range of 2.5 from the ground clutter border were selected. Ordinary

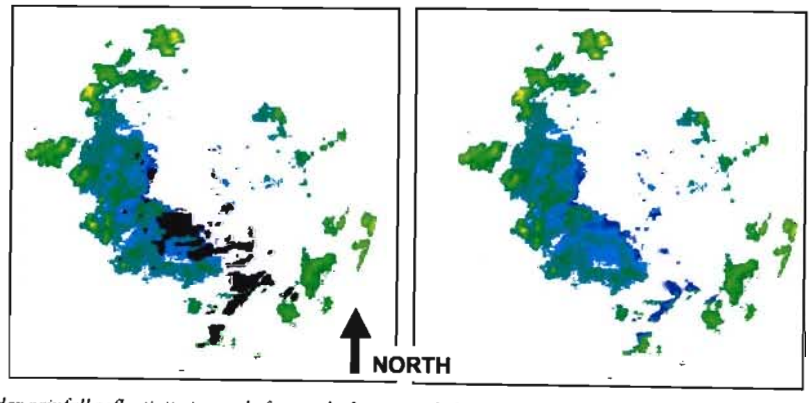


Fig. 13. *Bethlehem radar rainfall reflectivity image before and after ground clutter removal (25 February 2003). Black areas indicate contaminated data. The square surrounding each radar image has sides of length approximately 270 by 270 kilometres.*

Kriging was used to compute the missing reflectivity data with a semi-variogram function having L = 11 pixels and α = 1.5 being used in both cases. The estimated values were compared to the real reflectivity values and the square of the errors computed at each target point; these sets are shown in the lower pairs of diagrams in Fig. 14. The computation time was recorded for each technique; Table 1 gives a summary of the results.

For this particular radar reflectivity image, the ITNN technique performed better than the BTNN technique. Not only was the sum of the errors squared lower than the BTNN technique but the computation time was faster by a factor of ten. However, as stated earlier, a more comprehensive testing process needs to be undertaken to determine the most appropriate technique on a variety of rainfall types.

Limitations may exist for very large areas of ground clutter

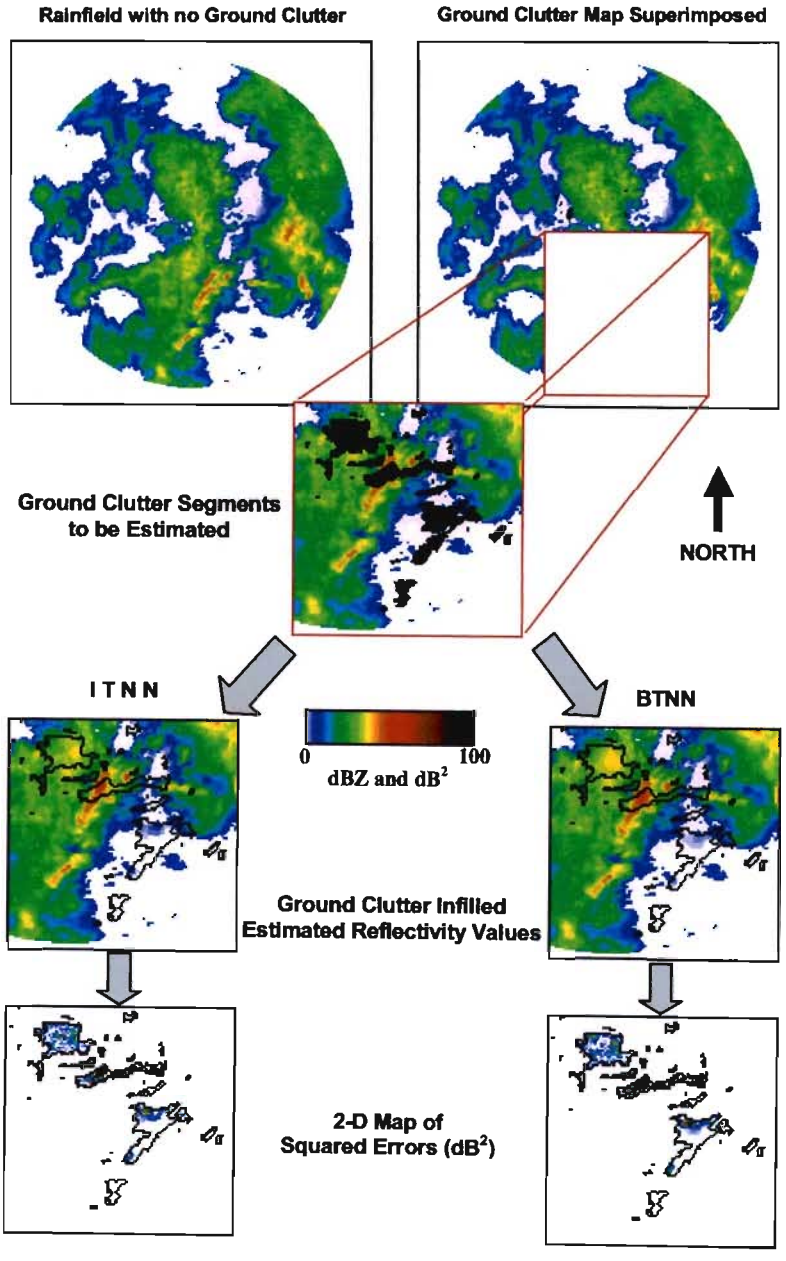


Fig. 14. *Comparison of ITNN and BTNN techniques.*

Table 1. Errors and computation time for techniques tested in the example in Fig. 14.

	Sum of squared errors (dB²)	*Computation time (seconds)*
ITNN	14 700	1.66
BTNN	19 000	16.8

where the best estimate of the rainfall near the centre of the ground clutter will be the mean value of the control points (if the distance exceeds the correlation length) which may result in errors of high magnitude. Large errors may also occur where ground clutter completely or partially hides highly convective rainfall which is surrounded by stratiform rain, this phenomenon occurs in one of the images shown in Fig. 14.

Rainfall estimation at ground level

The rainfall at ground level is of more interest than measurements aloft in Earth-bound applications such as hydrology and agriculture. Accurate rainfall estimates at ground level will also assist in the current Water Research Commission (WRC) research project in South Africa concerned with daily rainfall mapping over South Africa (Sinclair and Pegram, 2003). For ground level rainfall estimates, the Kriging technique can be extended easily from two dimensions into a three dimensional space, as illustrated by the following example.

In this algorithm the control points are the information in the pixels aloft which are Kriged onto target points in a 200-kilometre diameter circle at ground level. The Kriging is carried out with an ITNN approach, as suggested by Seed and Pegram (2001), with the 25 nearest control points to each target being selected.

An SVD matrix rank reduction technique was once again used to ensure computational efficiency. However, when Kriging in a 3-D data set, a 3-D semi-variogram model is used, as given by Eqn. (13), developed by and reported in Seed and Pegram (2001).

$$\gamma(h) = \sigma^2 \cdot [1 - \exp(-d^{2H})] \qquad (13)$$

where: $$d^2 = \left(\frac{r}{r_0}\right)^2 + \left(\frac{h}{h_0}\right)^2 \quad \rightarrow$$

$$d^2 = \left(\frac{x^2 + y^2}{r_0^2}\right) + \left(\frac{h}{h_0}\right)^2$$

where σ^2 is the field variance, r is the distance in the horizontal plane, r_0 the horizontal correlation length, h is the distance in the vertical plane, h_0 the vertical correlation length and H the scaling exponent.

An example application of this technique is illustrated in Fig. 15. The image is from the Durban weather radar, 9 November 2000. The following semi-variogram values were used in the Ordinary Kriging process: r_0 = 15km, h_0 = 12 km and H = 1,25 (Seed and Pegram, 2001).

Summary and conclusion

Radar rainfall data have importance in many fields of endeavour. However, the fact that errors persist in these data means that the rainfall estimates are not as accurate as they might be. These errors, noticeably ground clutter, can be detected and identified in the radar images. The technique proposed for estimating the missing data in this paper was Ordinary Kriging. Various methods to improve computational efficiency such as using nearest neighbours to a target point, matrix rank reduction techniques and inversion by partitioning have been implemented in order to make the Ordinary Kriging process computationally efficient in real-time applications.

Two techniques for removing ground clutter contamination have been developed which exploit the screening effect in Kriging. These are Individual Target Nearest Neighbours (ITNN) and Border Tracing Nearest Neighbours (BTNN) techniques. These methods still need to be tested thoroughly to determine which is optimal in terms of accuracy and appropriateness, but preliminary testing on typical sample sizes indicates that ITNN is ten times faster than BTNN and is more accurate. Both algorithms exhibit sufficient computational speed for on-line data cleansing, when supported by computationally efficient methods: (i) rank reduction using Singular Value Decomposition and (ii) matrix partitioning. It is anticipated that a further advantage of the ITNN Kriging approach would be to exploit local semi-variogram models evaluated dynamically; this has yet to be implemented but will be an advantage where radars sample convective and stratiform rainfall simultaneously.

The problem of rainfall estimation at ground level is of importance. The Ordinary Kriging technique was used in three dimensions to provide an estimate of rainfall at ground level. An ITNN approach was employed to extrapolate from control points aloft to targets at ground level. Further work needs to be carried out on the ground level rainfall estimation procedure to improve computational efficiency and validate the resulting rainfall estimates. This technique will be particularly useful where ground clutter targets are to be infilled, by importing information from aloft as well as laterally.

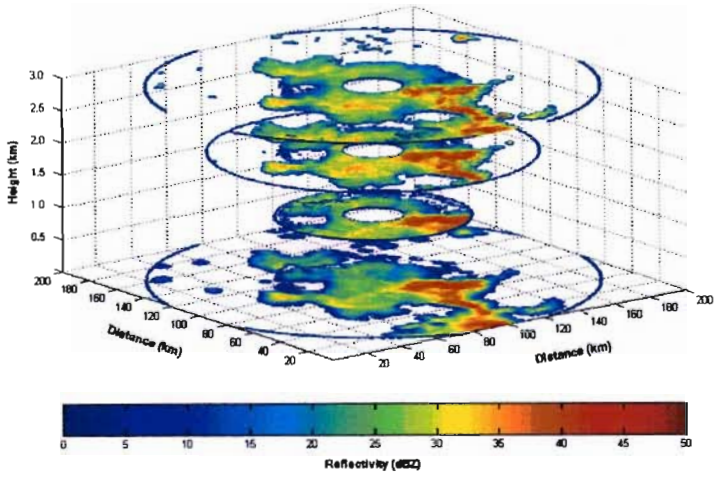


Fig. 15. *3D illustration of measured rainfall reflectivity data at 1, 2 and 3km above ground level and Kriged reflectivity estimate at ground level.*

The techniques presented in this paper show considerable promise and, as the ideas mature, should prove valuable in hydrometeorology.

Acknowledgments

This work is currently supported financially by the South African Water Research Commission (WRC) and the National Research Foundation (NRF). Thanks go to Scott Sinclair for Figs. 7 and 8 and work relating to them, also to Deon Terblanche and Pieter Visser of SAWS: METSYS at Bethlehem, South Africa for invaluable methodological input and the ground clutter maps. Finally, Alan Seed of the Australian Bureau of Meteorology is thanked for drawing attention to the reference to the 'screening effect' in Kriging and continual encouragement in the endeavour.

References

Aitken, A.C., 1967. *Determinants and Matrices.* Oliver and Boyd Ltd. London, UK.

Bell, T.L., 1987. A Space-Time Stochastic Model of Rainfall for Satellite Remote Sensing Studies, *J. Geophys. Res.*, **92**, D8, 9631.

Chiles, J.P. and Delfiner, P., 1999. *Geostatistics Modelling Spatial Uncertainty.* Wiley, New York, USA.

Cressie, N. 1993. *Statistics for Spatial Data.* Wiley-Interscience Publication. New York, USA.

Journel, A.G. and Huijbregts, Ch. J., 1978. *Mining Geostatistics.* Academic Press. London, UK.

Marshall, J.S. and Palmer, W.M., 1948. The Distribution of Raindrops with Size. *J. Meteorol.*, **5**, 165–166.

Pegram, G.G.S. and Clothier, A.N., 2001. High resolution space-time modelling of rainfall: the "string of beads" model. *J. Hydrol.*, **241**, 26–41.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1992. *Numerical Recipes in C The Art of Scientific Computing.* 2nd Edition. Cambridge University Press. Cambridge, UK.

Seed, A.W. and Pegram, G.G.S., 2001. Using Kriging to infill gaps in radar data due to ground clutter in real time. *Proceedings Fifth International Symposium on Hydrological Applications of Weather Radar Radar Hydrology.* Kyoto, Japan. 73–78.

Sinclair, D.S. and Pegram, G.G.S., 2003. Combining traditional and remote sensing techniques of rainfall measurement as a tool for Hydrology, Agriculture and Water Resources Management. *Proceedings 11th SA National Hydrology Symposium*, Port Elizabeth, South Africa.

Sonka, M., Hlavac, V. and Boyle, R., 1999. *Image Processing, Analysis, and Machine Vision.* 2nd Edition. PWS Publishing. Pacific Grove.

Todini, E., 2001. A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrol. Earth Syst. Sci.*, **5**, 187–201.

Vezzani, G., 1994. Ground clutter Suppression in Atmospheric Phenomena Observation. *COST 75 Weather Radar Systems International Seminar.* Brussels, Belgium. 281–295.

Visser, P., 2003, *Spatial interpolation and Mapping of rainfall: 2. Radar and Satellite products*, Draft final report (K5/1152) to the Water Research Commission for the period March 2000 to March 2003. Water Research Commission. Pretoria. 10–19.

3.7 CHAPTER SUMMARY AND WAY FORWARD

In this chapter a review of the infilling technique to remove ground clutter using Ordinary Kriging was presented and the advantages and disadvantages of Kriging discussed. Methods to improve computational efficiency were presented such as exploiting the 'screening effect' and the problem of ill-conditioning of the coefficient matrix was discussed. SVD, with the trimming of singular values, was proposed as a solution to the problem of ill-conditioning and was shown to provide computational efficiency and stability in the Kriging technique.

Two preliminary ideas to infill ground clutter in 2D were investigated, the ITNN and BTNN techniques. After comparison testing it was decided that the ITNN technique was more effective than BTNN to infill ground clutter and unknown data in radar volume scans due to its relative simplicity, accuracy and computational efficiency. A preliminary example of estimating rainfall at ground level from radar reflectivity volume scan data was also presented.

The issues still to be addressed in the remainder of this thesis are:

- Adapting the 2D ITNN technique to work in 3D.
- Investigating the effectiveness of using fixed precalibrated semivariogram parameters, α and L, rather than estimating the semivariogram for each target.
- Developing a method to extrapolate radar reflectivity data contained aloft to ground level.
- Overcoming the problem of the bright band and devising a correction procedure for this phenomenon.
- A full validation for the ground clutter infilling and radar reflectivity extrapolations to ground level will be reviewed.

CHAPTER 4
AN OPERATIONAL METHOD FOR INFILLING MISSING RADAR RAINFALL DATA

4.1 INTRODUCTION

Outlined in this chapter is the operational method devised for infilling all missing data contained in the radar volume scan and to provide an estimate of the rainfall at ground level. The steps taking in the algorithm are outlined in detail and the methodology and validation of each step is described. It is assumed in the sequel that contaminated data such as ground clutter have already been identified and that the pixels in the radar volume scan have been masked accordingly.

This chapter consists of the following sections. Firstly the details of the rainfall classification algorithm to distinguish convective and stratiform pixels are outlined and the characteristics of the classified rainfall described. A novel bright band correction algorithm based on local gradients is then introduced and tested. The next part describes the empirical computation of the robust semivariogram from radar reflectivity data and outlines the use and validation of climatalogical semivariograms based on rainfall type. Kriging theory is then briefly discussed, where the methods designed to ensure computational efficiency and stability in a 3D context are outlined. 3D Universal and Ordinary Kriging follow in their application to Cascade Kriging, a novel idea for extrapolation. Finally a statistical comparison between the temporally accumulated radar estimates and the Block Kriged raingauge estimates is carried out over matching areas, for selected rainfall events, to determine the quality of the rainfall estimates at ground level.

A copy of an invited journal paper, which has been reviewed and conditionally accepted for publication in NHESS and summarises the above topics, is also included. The sections of this chapter serve as preludes to the paper and discuss various aspects that were not included in the journal paper due to restrictions on its length.

4.2 RAINFALL CLASSIFICATION

The initial step of the operational algorithm for infilling the missing or inaccurate data contained in the volume scan, requires a simple classification and thresholding of the instantaneous radar volume scan data, pixel by pixel on each CAPPI. The classification

technique adopted was developed by examining already developed classification algorithms that have been used in practice.

Initially Mittermaier's (1999) rainfall classification algorithm was investigated, which involves classifying the rainfall into three separate zones – stratiform, intermediate and convective. The algorithm works by firstly examining the 4 km CAPPI; all pixels above 40 dBZ are immediately classified as convective. The remainder of the pixels are then examined in the horizontal plane on the 4 km CAPPI to determine if they are to be classified as intermediate or stratiform using local statistics. The second part of the algorithm involves examining the vertical structure of the rainfall by re-examining the vertical structure of the radar volume scan; if the average value of the non-zero values exceeds a predetermined threshold then the pixel may be reclassified as stratiform or convective. The vertical classification structure in this instance is given precedence over the initial horizontal step in the classification.

The above algorithm was applied to an instantaneous radar reflectivity image from the Bethlehem weather radar (30 December 2001), which is illustrated in Figure 4.1. The classified image (right hand side) shows that the rainfield has been classified into three distinct rainfall zones, a small percentage being classified as convective, the remainder as intermediate and stratiform.

Reflectivity Image **Classified Image**

Figure 4.1: Example of application of Mittermaier's (1999) classification algorithm applied to an instantaneous reflectivity image from the Bethlehem weather radar (30 December 2001), classified into three separate rain zones – convective, intermediate and stratiform. Sizes of images are 300 km square.

Although the algorithm effectively separates the rainfall into three zones, various problems were identified with the rainfall classification algorithm that made it unsuitable for the purposes of classifying radar volume scan data before the infilling of missing data and providing an estimate of the rainfall at ground level. The algorithm is firstly lengthy and for real time applications may result in an unnecessarily long computation time given to the classification part of the algorithm. The 2 km CAPPI also plays a pivotal role in classifying the image using the algorithm which results in only a small proportion of the total volume scan being classified because of the limited extent of the 2 km CAPPI.

Although the algorithm (from a meteorological perspective) accurately separates the rainfall into the three zones, in this study the classification was being developed with a view to providing climatological semivariograms based on the classifications. It was determined that the separation into three zones resulted in an overly elaborate classification of the rainfall, with little distinction or separation (in most instances) of the locally fitted semivariogram parameters into intermediate, convective or stratiform types. The classification was therefore too fine and needed to be simplified into a two-zone classification of stratiform and convective rainfall. Ehret (2002: 75-76) also simplified a rainfall classification algorithm he devised based on a fuzzy rule system after also determining that the classification was too fine.

The following classification scheme was devised which separates the rainfall pixels into only two types – convective and stratiform. The first step of the algorithm consists of applying a simple threshold where all reflectivity values equal to and below 18 dBZ are set to zero, equivalent to a rainrate of 0.5 mm/hr. If the values range from 18 dBZ to, but not including, 35 dBZ the rainfall is classified as stratiform and all values above and including 35 dBZ are classified as convective. The classification criteria are summarised in Table 4.1 for a pixel-by-pixel classification basis on each CAPPI.

Table 4.1: Rainfall classification and threshold criteria to separate valid radar volume scan data into stratiform and convective rainfall domains.

CRITERIA	CLASSIFICATION
Pixel ≤ 18 dBZ	No Rainfall: Pixel set to 0 dBZ
18 dBZ < Pixel < 35 dBZ	Stratiform Rain
Pixel ≥ 35 dBZ	Convective Rain

An example of the classification algorithm described in Table 4.1, and uses the same rainfall images as in Figure 4.1, is illustrated in Figure 4.2 where the CAPPI is classified into two separate zones – convective and stratiform rain. In comparison to Mittermaiers algorithm the intermediate zone has now been removed and a larger amount of the rainfall has now been classified as convective. Further details on the classification algorithm are contained in Wesson and Pegram (2005), which is presented in Section 4.7, pages 115 to 120.

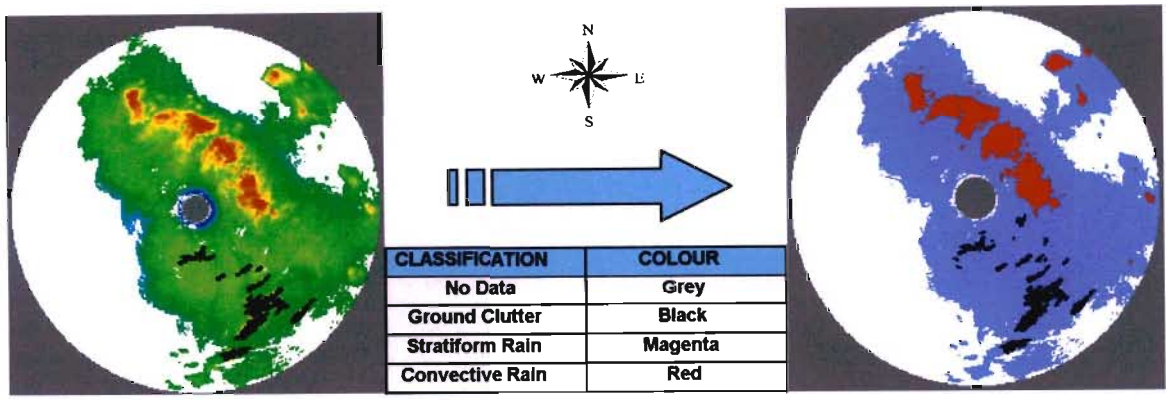


Figure 4.2: Example of classification algorithm results applied to an instantaneous reflectivity image from the Bethlehem weather radar (30 December 2001), classified into two rain zones – convective and stratiform. Sizes of images are 300 km square.

4.2.1 Characteristics of Classified Rainfall

The rainfall classification algorithm was tested on numerous images ranging over five different years (1995, 1996, 2000, 2001 and 2002) from two weather radars in South Africa: Bethlehem and Durban. The above two zone classification algorithm has the effect of separating the rainfall into distinct zones that have very different characteristics and statistical properties. This can be demonstrated by examining the climatological profiles of the classified rainfall and statistical properties such as the mean and standard deviation of

the rainfall. The climatological profiles and statistical properties are distinct in the two types of rain. This is elaborated on in more detail in Wesson and Pegram (2005) contained in Section 4.7 on pages 118 to 120.

4.3 BRIGHT BAND TREATMENT

As described in Section 2.3.4 various data quality problems do occur within weather radar volume scans, one of which is the bright band that causes an overestimation of rainfall when viewed by the weather radar due to ice crystals and snow melting once the 0^0C isotherm is reached. The bright band generally occurs at 2 km above ground level in the interior of South Africa (Mittermaier, 1999), as can be seen in the figure of climatological profiles of classified rainfall for the stratiform rainfall contained on page 120.

4.3.1 Effect of Bright Band on Ground Level Estimates

When applying an extrapolation algorithm such as Kriging to radar reflectivity data affected by bright band inaccurate estimations of stratiform rainfall may be obtained at ground level. This is illustrated in the left hand side image of Figure 4.3 where the trend of decreasing rainfall created from the 1 km and 2 km CAPPIs are extrapolated to ground level resulting in an underestimate of the rainfall at ground level. This results in using the 25 nearest neighbours ITNN technique, which typically uses controls only 2 layers above the target.

If a correction is applied to the 2 km CAPPI, as indicated in the right hand side image of Figure 4.3, the trend of rainfall is now seen to be increasing as ground level is approached and the vertical profile resembles rainfall which is not affected by bright band. If the rainfall data are now extrapolated to ground level from levels 1 to 2 this results in a more realistic estimate of the rainfall at ground level. Note that the profile was obtained from a column of data in the volume-scan only 1 km square in plan, so these data are measurements of individual reflectivities of single pixels.

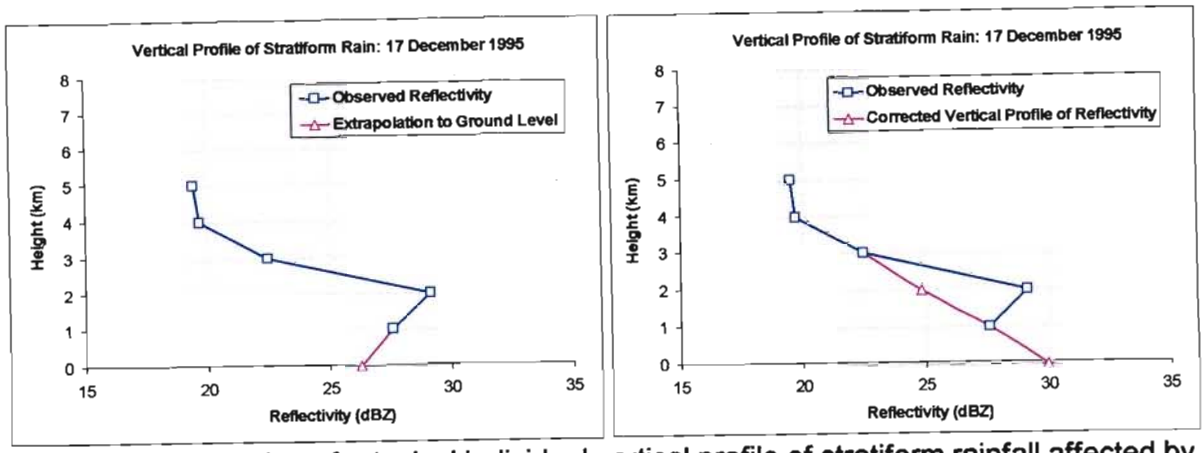


Figure 4.3: Illustration of a typical individual vertical profile of stratiform rainfall affected by bright band taken from the Bethlehem weather radar (17 December 1995) with an extrapolation to ground level. The right hand image represents a corrected image with extrapolation from the corrected data to ground level.

4.3.2 Bright Band Correction Algorithm

A correction algorithm for the bright band needs to be designed to be computationally efficient so as to work on a real time basis and on radar reflectivity data set out in Cartesian co-ordinates on CAPPIs. The algorithm also needs to effectively identify regions affected by bright band and have a low probability of adjusting reflectivity data not affected by bright band.

Initial attempts at correcting the bright band included investigating Simple Kriging as a possible tool. In Simple Kriging the mean is assumed known a prior and not stimated from the data by the computation. The Simple Kriging equations are given by Equation (4.1):

$$\hat{z}(s_0) = M(s_0) + \lambda^T(s_0) \cdot [z - M] \qquad (4.1)$$

Where $\hat{z}(s_0)$ is the estimated value of the target at point s_0, $M(s_0)$ is the "known", possibly pre-computed mean value of $z(s_0)$; once again $\lambda(s_0)$ is the vector of weighting values and $[z-M]$ is the vector of controls with their mean vector subtracted.

If a target is located at ground level and the controls are selected from the 1 and 2 km CAPPIs, the mean value for the controls at each height level is pre-computed and inserted in M. For want of data at ground level, stratiform controls in the mean value, $M(s_0)$, is the

mean value of the controls at the 1 km CAPPI; in this way it was expected to obtain a more realistic estimate of the rainfall at ground level.

A second alternative investigated was to apply a global correction to all stratiform controls on the 2 km CAPPI when affected by bright band. The adjustment factor is calculated by firstly computing the mean values of the stratiform rainfall on the 1 km and 3 km CAPPIs. The average of the two mean values is computed to determine the theoretical mean value of the stratiform rain in the 2 km CAPPI as if it was unaffected by bright band. The theoretical mean value is then divided by the observed mean value of the stratiform rainfall to obtain a correction factor which is applied to all stratiform rainfall in the 2 km CAPPI to obtain new corrected values for stratiform rainfall.

The above two techniques however did not provide improved estimates of stratiform rainfall at ground level. The Simple Kriging technique, although providing a slightly better estimate of the stratiform rainfall than Ordinary Kriging at ground level still provided an underestimate of the rainfall as indicated by the example shown in Figure 4.4.

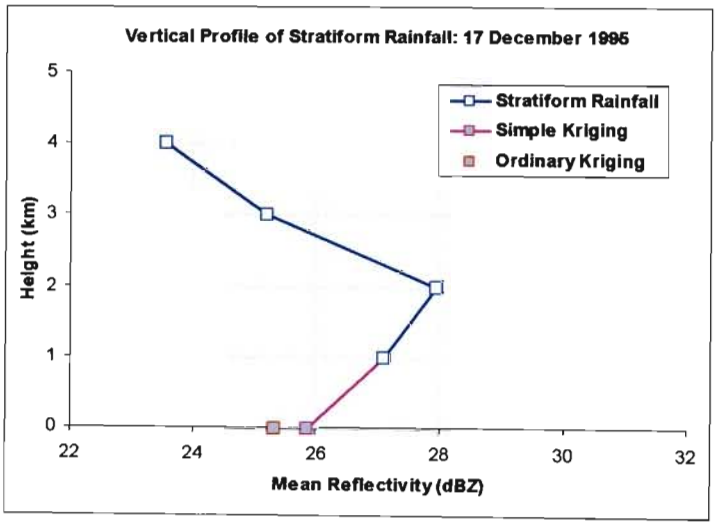


Figure 4.4: Stratiform rainfall extrapolations to ground level by both Ordinary and Simple Kriging for an individual profile in an instantaneous reflectivity volume scan from the Bethlehem weather radar (17 December 1995) affected by bright band.

The global correction technique (the second of the two suggested above) was also not adopted due to the fact that not all stratiform rainfall contained on the 2 km CAPPI is affected by bright band. A global correction scheme would therefore correct stratiform

rainfall values unaffected by the bright band and unnecessarily reduce their values below their true magnitude.

A technique was therefore devised to evaluate all vertical profiles in the stratiform zone on an individual basis by examining the local gradients to determine which stratiform pixels are affected by bright band. Stratiform reflectivity values identified as being affected by bright band are then corrected on an individual basis by examining the gradients computed from the individual column of pixels. A detailed explanation of the identification and correction of the stratiform rainfall pixels affected by bright band is contained in this paper Section 4.7, pages 120 to 125.

4.3.3 Testing the Bright Band Correction Procedure

To test the effectiveness of the proposed bright band correction procedure, a day of rainfall measured by the Bethlehem weather radar was selected (17 December 1995) that consisted predominantly of stratiform rainfall and showed clear evidence of a bright band.

Figure 4.5 is a scatter plot of the 24-hour radar accumulations estimated at ground level before and after the bright band correction procedure discussed in Section 4.3.1. The accumulations were computed for each pixel in a radius of 42 km from the radar, in the Northern half of the image, to avoid any possible ground clutter contamination, an area consisting of 2487 pixels. Figure 4.5 is a scatter plot of accumulation values before and after the bright band correction algorithm was applied. As can be seen on the figure the accumulations are, on average considerably higher by a shift of 7 mm in 24-hours after the bright band has been treated. This indicates that the correction produces a vertical profile, and corresponding extrapolation to ground level, similar to the right hand image of Figure 4.3.

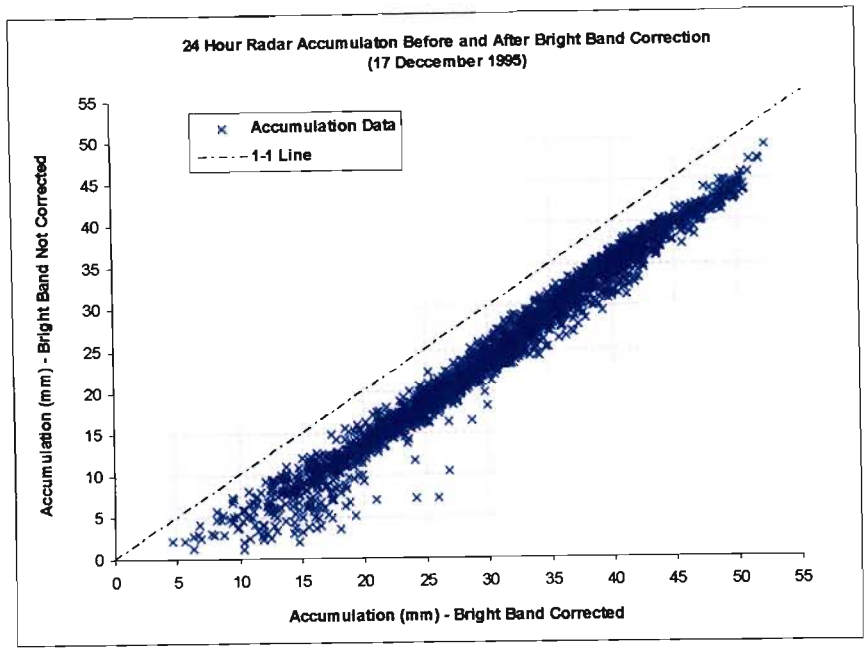


Figure 4.5: Scatter plot of a 24-hour accumulation of radar rainfall at ground level of 2487 pixels from the Bethlehem weather radar for 17 December 1995 before and after the bright band correction algorithm is applied. The results indicate that the bright band correction results in a higher accumulation of rainfall at ground level than without bright band correction.

To test the effectiveness of the bright band adjustment algorithm, a comparison was carried out between Block Kriged raingauge estimates and the radar accumulation estimates before and after the bright band correction procedure. The details of the raingauge network used for this test are contained in Section 4.7, pages 113 to 114. The comparison was carried out on the 24-hour rain event from the Bethlehem weather radar from the 17 December 1995 with a strong bright band evident. The details of the comparison between raingauge and radar accumulations will be discussed in Chapter 5.

The results of the comparison are illustrated in Figure 4.6 and summarised in Table 4.2. Figure 4.6 indicates that there is an underestimation of the uncorrected rainfall estimates by the radar in comparison to the raingauge accumulations with the corresponding correlation coefficient (r^2) returning a low value of 0.34. The radar accumulations after the bright band correction was applied to each instantaneous image throughout the 24-hours results in an improvement in the radar accumulation estimates at ground level. The radar accumulations more closely match the raingauge accumulation estimates and return a higher r^2 value and a lower Sum of square of errors (SSE) as indicated in Table 4.2.

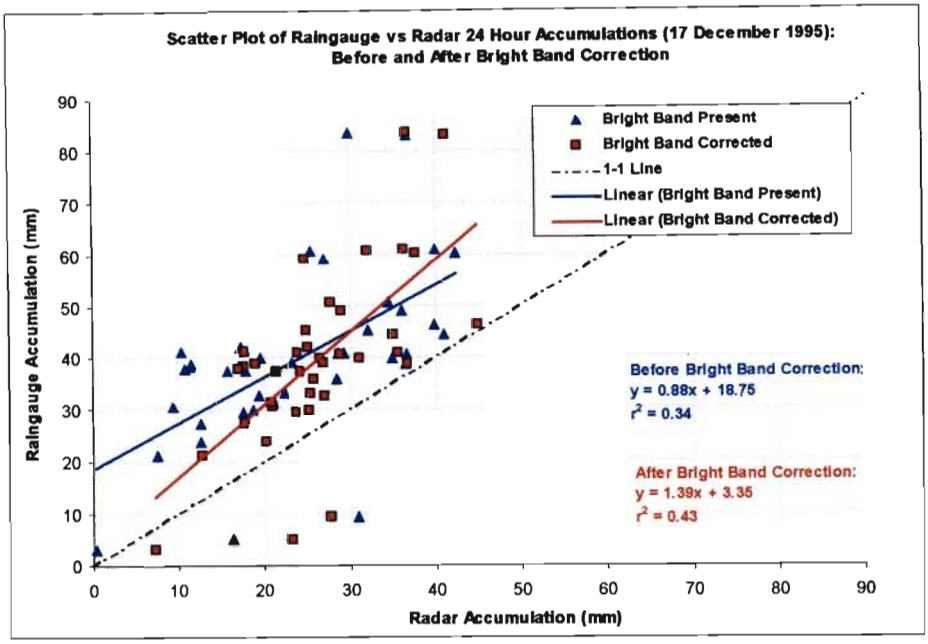


Figure 4.6: Scatter plot of a 24-hour radar and raingauge accumulation at ground level from the Bethlehem weather radar for 17 December 1995 before and after the bright band correction algorithm is applied.

Table 4.2: Sum of square of errors (SSE) and r^2 value between radar and raingauge estimates before and after bright band correction.

	Before Bright Band Correction	**After Bright Band Correction**
SSE	16 601	13 161
r^2	0.34	0.43

Based on the analysis described above and summarised in Table 4.2, although the improvement is not dramatic, the proposed bright band correction procedure provides a reasonable means of adjusting the 2 km CAPPI reflectivity values when affected by bright band so as to provide a better estimate of the rainfall at ground level. The bright band correction algorithm is computationally inexpensive and therefore suitable for real time applications.

4.4 SEMIVARIOGRAM ESTIMATION

The semivariogram provides a way of measuring the spatial statistical dependence between variables and is the first step in the Kriging process. An example of a typical idealised semivariogram is plotted in Figure 4.7. In a stationary random field, the semivariogram is the difference between the field variance and the spatial covariance

function. As indicated in the figure the semivariogram rises smoothly from the nugget of the graph; at the shorter distances the values are more highly correlated but with greater distance the values become less correlated until the sill is reached which is equal to the variance of the field. At this point the observations are considered to be uncorrelated; the range indicates the distance within which the observations are considered correlated.

- Range – the distance beyond which the data are considered to be uncorrelated.
- Sill – the point at which the empirically computed variance is equal to the variance of the field.
- Nugget – the nugget can represent either measurement error or the variance that exists within the field between closely spaced observation points.

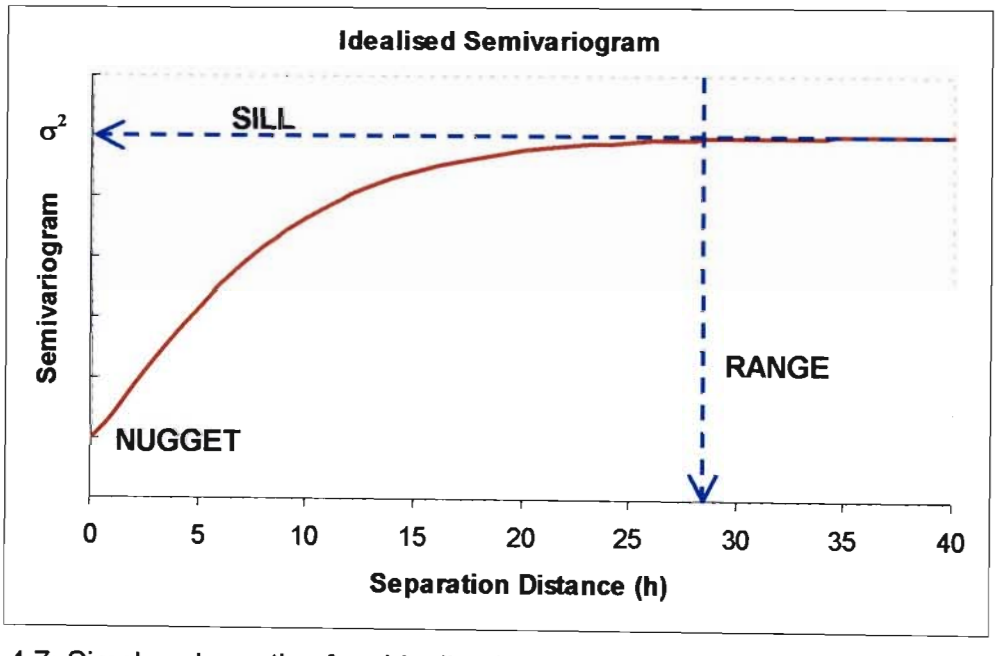


Figure 4.7: Simple schematic of an idealised semivariogram of exponential type with the positions of the sill, range and nugget indicated on the figure.

4.4.1 Empirical Computation of the Semivariogram

The semivariogram is most commonly computed by the Classical Variogram estimator proposed by Matherhon (1962) and is given by Equation (4.2):

$$2 \cdot \hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left[z(s_i) - z(s_j) \right]^2 \qquad (4.2)$$

where $\hat{\gamma}(h)$ is the sample semivariogram at lag h (the specified lag distance), $z(s_i)$ and $z(s_j)$ are the values of the variables at the specified locations, s_i and s_j which are h apart and $N(h)$ is the number of pairs separated by lag h. Unfortunately the classical estimator of Equation (4.2) is badly affected by non-typical observations whose influence is magnified by the $[\cdot]^2$ term in the summand (Cressie, 1993: 40).

This can be demonstrated by examining the distribution of the $[z(s_i) - z(s_j)]^2$ points at each lag distance h. As indicated in Figure 4.8, the computed mean and median value for each lag distance h computed from a selected portion of an instantaneous reflectivity field are distinctly different, indicating that the distribution of points is highly skewed and not normally distributed.

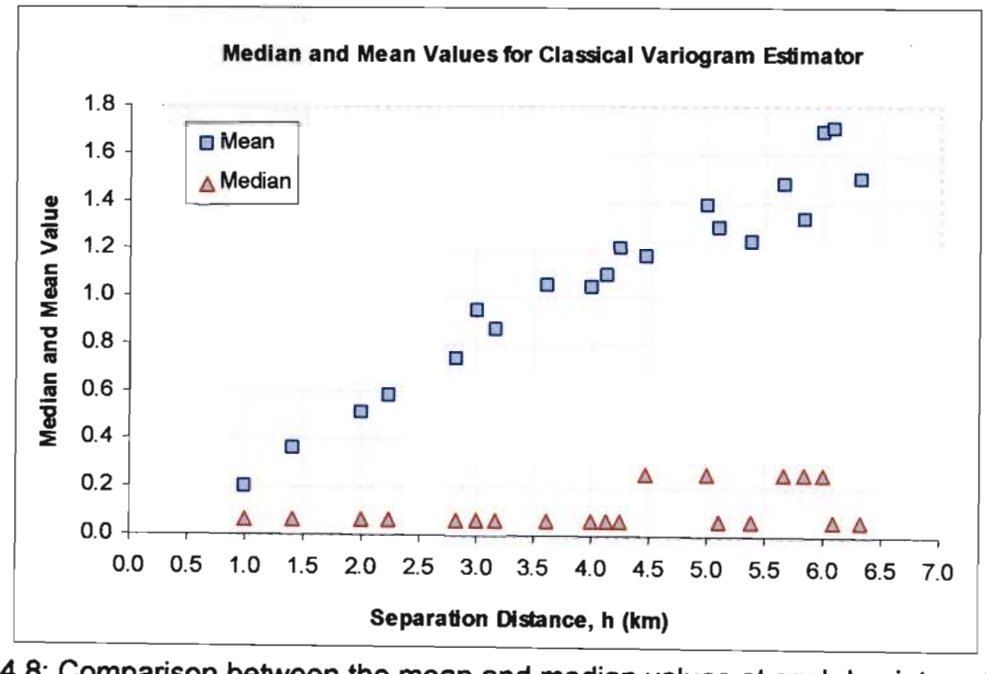


Figure 4.8: Comparison between the mean and median values at each lag interval for the Classical Variogram estimator. The results indicate that the distribution of points is heavily skewed and might follow a chi-square distribution as suggested by Cressie (1993: 75).

It was thus decided to see if the Robust Variogram, proposed by Cressie and Hawkins (1980) was the more appropriate model to use; it is given by Equation (4.3):

$$2 \cdot \bar{\gamma}(h) = \left\{ \frac{1}{|N(h)|} \sum_{N(h)} |z(s_i) - z(s_j)|^{1/2} \right\}^4 \Bigg/ \left(0.457 + \frac{0.494}{|N(h)|} \right) \qquad (4.3)$$

By computing the sum of the absolute differences of the pairs of data points, $z(s_i)$ and $z(s_j)$ in the square root domain and then raising the result to the power of four dramatically reduces the effect of uncharacteristic observations.

When computing the mean and median values from the same data set as used in Figure 4.8 but now using Equation (4.3) it was found that the skewness of the points at each lag (h) was greatly reduced and also that the mean and the median value of the sets of points for each lag interval more closely coincided, as shown in Figure 4.9, both indicating that the distribution of points at each lag (h) more closely approximates a normal distribution, yielding an unbiased estimate of the semivariogram.

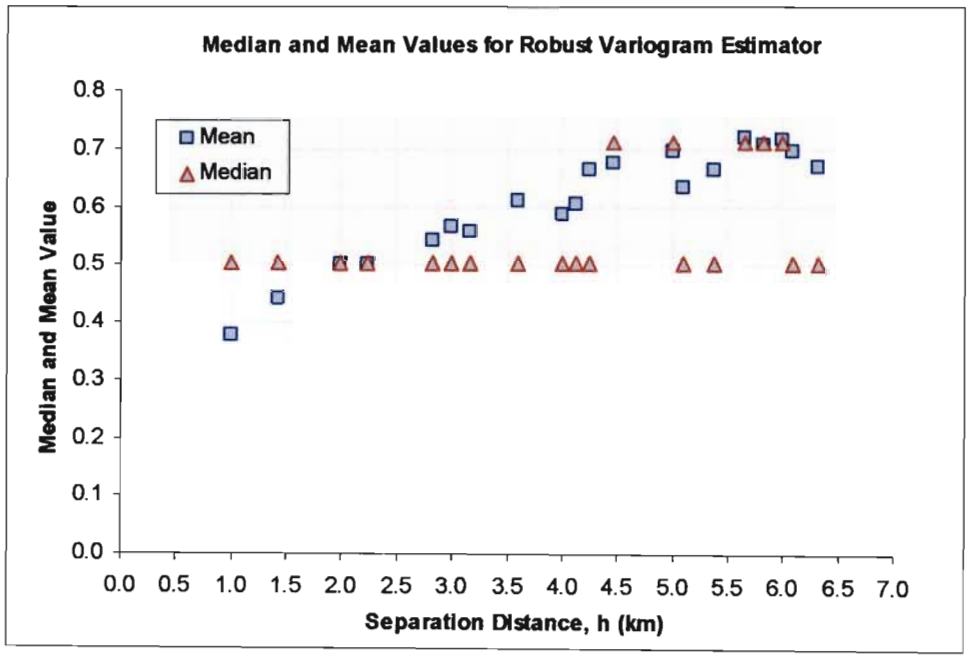


Figure 4.9: Comparison between the mean and median values at each lag interval for the Robust Variogram estimator. The results indicate that the distribution of points are now symmetrical and approximate to a normal distribution.

4.4.2 Semivariogram Model Fitting

After examining many sample semivariograms, as suggested by Seed and Pegram (2001) and introduced as Equation (3.3), the two-parameter isotropic exponential model was chosen as the model to fit to the empirical semivariogram values, as given by Equation (4.4):

$$g(h) = 1 - \exp[-(h/L)^{\alpha}] \qquad (4.4)$$

where h is the Euclidian distance between data points, L the correlation length and α the exponent parameter which lies in the range $0<\alpha\leq 2$. Equation (4.4) is applied to standardised data as discussed in Section 4.7 (pages 131 to 132) therefore the σ^2 term, which is the field variance reduces to unity and is not shown in Equation (4.4). The following was taken into consideration as recommended by Journel and Huijbrechts (1978: 194) when fitting the model. The model was fitted up to only half the maximum possible lags and only to the bins at each h that contained thirty or more points. The Modified Downhill Simplex technique was then used to solve for the α and L parameter values. These were found by minimising an objective function U defined in Equation (4.5) by a weighted least squares approach:

$$U = \sum_{j=1} N(h_j) \cdot \left[\frac{\hat{\gamma}(h_j)}{\gamma(h_j, b)} - 1 \right]^2 \qquad (4.5)$$

where $N(h_j)$ is the number of points in bin j, $\hat{\gamma}(h_j)$ are the empirical semivariogram values given by Equation (4.3) and $\gamma(h_j, b)$ are the model values (Cressie, 1993: 97). The weighted fitting criterion ensures that the bins with the greater number of observations are fitted more accurately; the bins closer to the origin are also more likely to contain the greater number of observations ensuring a good fit near the origin. This is particularly important in the context of nearest neighbour Kriging which is employed operationally.

The Modified Downhill Simplex technique works as follows, as described by Press *et al* (1992: 408-410). Two parameters, α and L, need to be solved for which results in a 3-sided simplex, i.e. a triangle. The algorithm works by determining which vertex of the triangle has the highest objective function value, this vertex is then reflected through the opposite face to a lower valued point. The simplex then steadily converges towards the minimum point in the space of the objective function. This continues until the decrease in the objective function values at each vertex of the triangle is smaller in magnitude than a specified amount. Because the parameters must be bounded and the simplex method is an unbounded optimizer, the first step involves the logit transform where Equation (4.6) is used to transform the parameter space into the unbounded domain:

Logit transform:

$$y = \ln\left[\frac{x-a}{b-x}\right] \qquad (a \le x \le b) \qquad (4.6)$$

The Modified Downhill Simplex search is then carried out in the unbounded domain where the parameter values are determined by the reverse logit transform, Equation (4.7), each time the objective function is evaluated (Sinclair, 2001).

Reverse logit transform:

$$x = \left[\frac{b \cdot e^{y} + a}{1 + e^{y}}\right] \qquad (a \le x \le b) \qquad (4.7)$$

To ensure that the minimum value of the objective function *U*, defined in Equation (4.5), is found, care was taken to start the calculation with the vertices of the simplex closely positioned to one another. The simplex was also started at 15 different locations inside the 2D bounded parameter space, as indicated in Figure 4.10, to increase the probability that the minimum value of the objective function *U* was located.

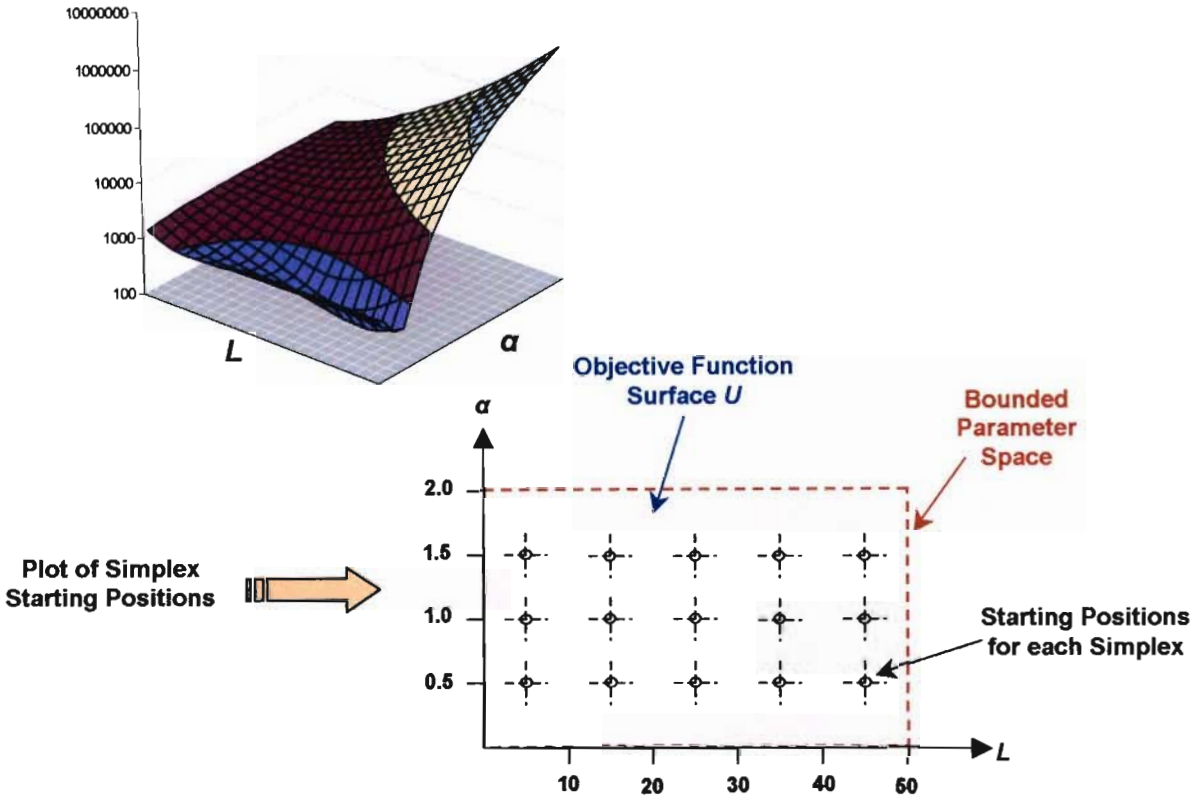


Figure 4.10: Simple schematic of an objective function plot (U) and the allocated starting position for each simplex inside the bounded parameter space to find the minimum value of the objective function U. The upper bound of L was selected from examining typical correlation functions fitted to radar rainfields.

4.4.3 Example of Semivariogram Computation

An example of the computation of a standardised climatological semivariogram of a selected segment from a reflectivity field is demonstrated and detailed in Wesson and Pegram (2005) on pages 131 to 132, which is contained in Section 4.7 of this chapter.

4.4.4 Parameter Fitting to Rainfall Types

Computing the empirical semivariogram values and then applying a model fitting routine to solve for the α and L parameters in Equation (4.4) is a computationally burdensome and time-consuming task, especially if this is done each time an unknown reflectivity value is

estimated. Reducing the computational burden of computing the semivariogram by the use of a climatological semivariogram was explored in the context of radar rainfields.

As detailed in Section 4.7, pages 133 to 135, instantaneous radar reflectivity images were selected from the Durban and Bethlehem weather radars over a period of several years. Initially the rainfall classification proposed by Mittermaier (1999) was applied as described in Section 4.2 to the selected images, which separated the rainfall into three zones stratiform, intermediate and convective rainfall zones. These regions were approximately 1200 km^2 (40 km by 30 km) in area or larger. From the selected regions the Robust semivariogram was computed and the two parameter exponential model fitted. This was done in the horizontal direction and the two parameters, α and L estimated from each selected region, were then plotted against one another in a scatter plot as illustrated in Figure 4.11.

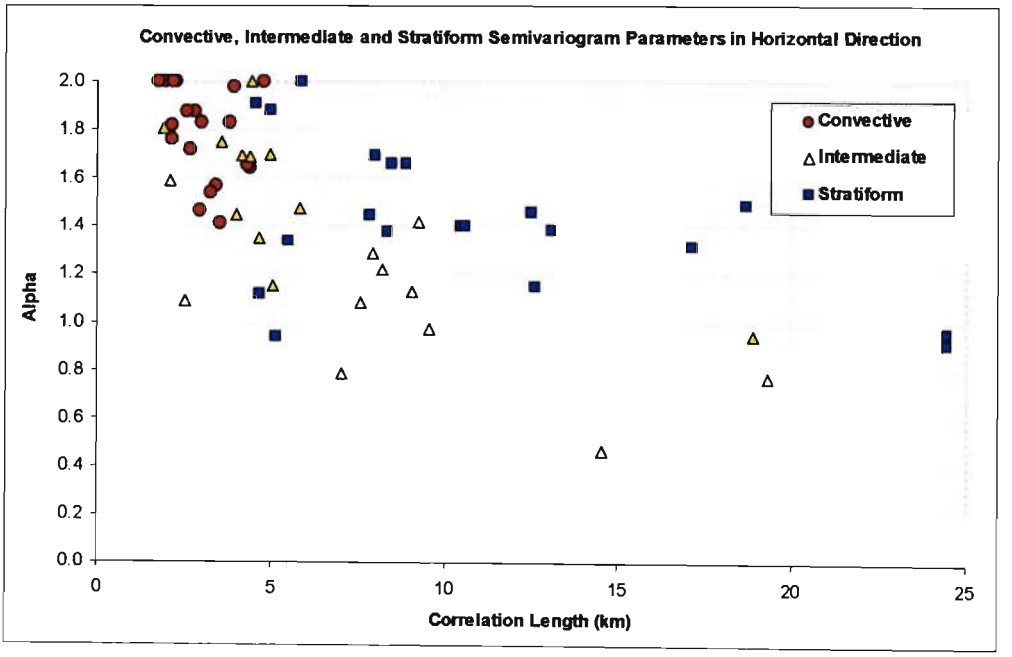


Figure 4.11: Scatter plot of alpha (α) and correlation length (L) parameters for stratiform, intermediate and convective rainfall in a horizontal direction, using Mittermaiers (1999) algorithm.

As indicated by Figure 4.11 there is no clear distinction in this case between the convective, intermediate and stratiform zones. The intermediate zone also overlaps into both the convective and stratiform zone with no clear clustering of the variables into any particular region or zone on the scatter plot. Based on the result in Figure 4.11 it was

decided that Mittermaiers rain classification algorithm was unable to distinguish the classes clearly enough and needed to be simplified and adjusted so as to effectively classify the rainfall into two zones – convective and stratiform rainfall.

The new scatter plots based on the simplified rainfall classification and contained in Wesson and Pegram (2005), Section 4.7, page 134 illustrate that the semivariogram parameters tend to separate into distinct regions according to the type of rainfall.

4.4.5 Cluster and Sensitivity Analysis

To test if a natural clustering of the semivariogram parameters occurred in the vertical and horizontal directions, a Fuzzy C-means clustering algorithm was run on the data. Fuzzy C-means is a clustering technique (Gordon, 1981: 58-60) where each data point is assigned to a cluster by a memebership grade. Initially a guess is made for the centroid of each cluster, with each data point assigned a membership grade for each cluster. The algorithm then iteratively updates the membership grades and cluster centroids until the minimum value of an objective function is attained.

In this instance the specification set in the cluster algorithm was to separate the scatter plot of semivariogram parameters (α and L) into two separate groups for both the plots in the horizontal and vertical directions. The results indicated a natural separation of the two groupings into convective and stratiform rainfall parameters, the details of which are outlined in Wesson and Pegram (2005) contained in Section 4.7, pages 135 to 136.

After it was determined that a natural clustering and separation of the semivariogram parameters for convective and stratiform rainfall, an analysis was undertaken to determine how sensitive the final Kriged solutions are to the range of expected semivariogram parameters within each rainfall type. The testing procedure involved removing portions of instantaneous reflectivity images and then infilling the removed data using Ordinary Kriging. This was done in two ways; firstly with horizontal information and secondly with vertical information.

Five different sets of parameters were used: firstly the centroid value of α and L for convective and stratiform rain and then the α and L values one standard deviation away from the centroid values, as computed by Equations (4.8) and (4.9) and illustrated in Section 4.7, (page 137) for the horizontal direction. Table 4.3 contains the standard

deviation values obtained for each of the semivariogram parameters in each rainfall type for the horizontal and vertical direction.

$$\alpha \pm \sigma_\alpha \quad , \quad 0 < \alpha \leq 2 \tag{4.8}$$

$$L \pm \sigma_L \quad , \quad 0 < L \tag{4.9}$$

Table 4.3: Standard deviation of horizontal and vertical semivariogram parameters

	HORIZONTAL		VERTICAL	
	σ_α	σ_L	σ_α	σ_L
STRATIFORM	0.36	3.71	0.32	1.01
CONVECTIVE	0.18	1.06	0.21	1.15

Figure 4.12 illustrates the testing procedure, where a selected portion of a convective reflectivity field has a segment of data removed and then infilled with Ordinary Kriging using five sets of different semivariogram parameters. The five sets of infilled data can then be compared to see if a significant difference exists between the infilled data sets.

An overview of the sensitivity analysis tests for the semivariogram parameters for convective and stratiform rainfall in the horizontal and vertical directions follows. The statistical results returned for each of the sensitivity tests undertaken are also discussed.

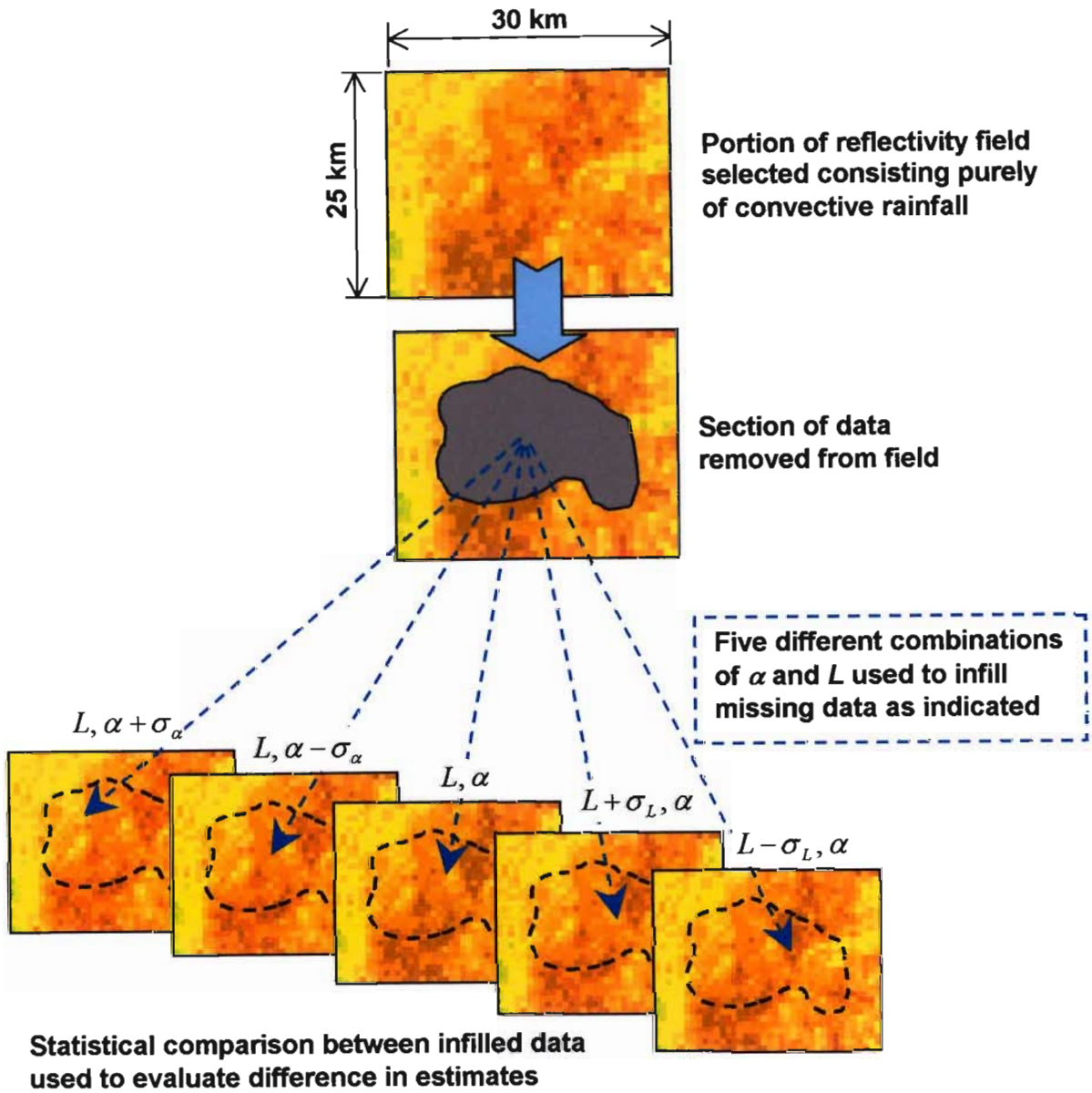


Figure 4.12: Schematic illustrating the testing procedure to determine the range of final Kriging solutions.

HORIZONTAL SENSITIVTY ANALYSIS: CONVECTIVE RAINFALL

The tests were carried out in a convective rainfall segments selected from instantaneous reflectivity images. The first test data set was taken from the Bethlehem weather radar (24 January 2002) with 76 targets being estimated. The second data set was also from the Bethlehem weather radar (19 December 2001) with 101 targets being estimated. The

targets for both data sets were infilled using the 25 nearest controls. A summary of the statistics for both are given in Figures 4.13 to 4.16 and Tables 4.4 and 4.5.

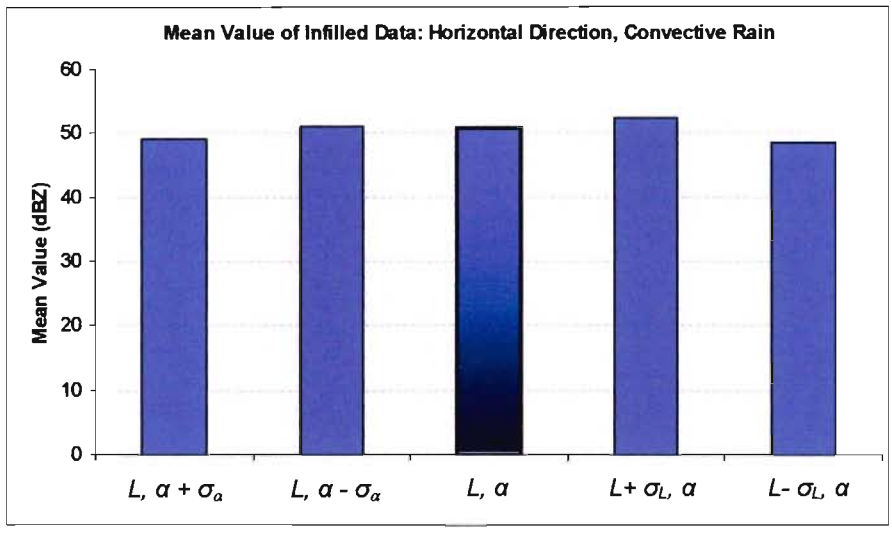


Figure 4.13: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (24 January 2002) with 76 targets being infilled.

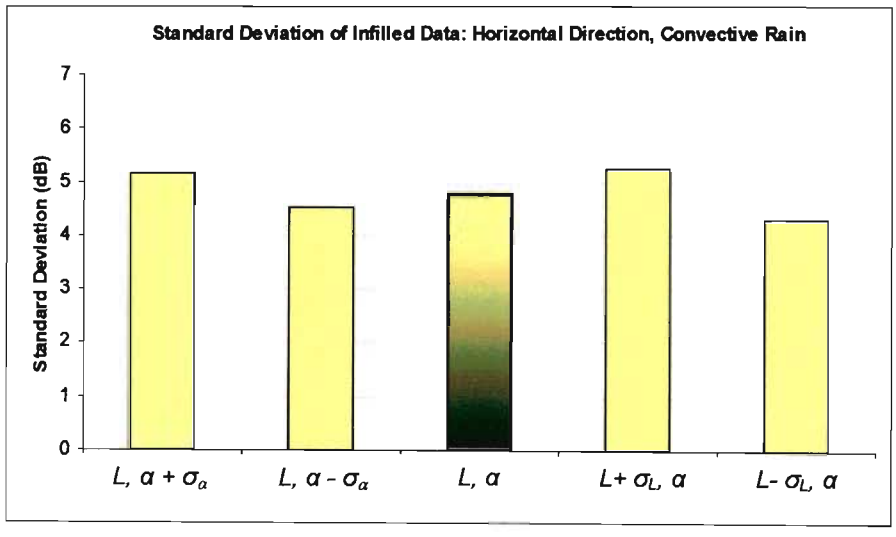


Figure 4.14: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (24 January 2002) with 76 targets being infilled.

Table 4.4: Summary of statistics returned for sensitivity analysis test for convective rainfall in the horizontal direction. Test data was from the Bethlehem weather radar (24 January 2002) and consisted of 76 targets being infilled.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	49.2	51.0	50.9	52.4	48.4
Median	48.9	49.3	49.4	50.5	47.5
Standard Deviation	5.2	4.5	4.8	5.2	4.3
Range	21.0	16.8	18.2	20.1	15.8
Minimum	38.9	44.2	43.8	44.1	42.8
Maximum	59.9	61.0	61.9	64.2	58.6

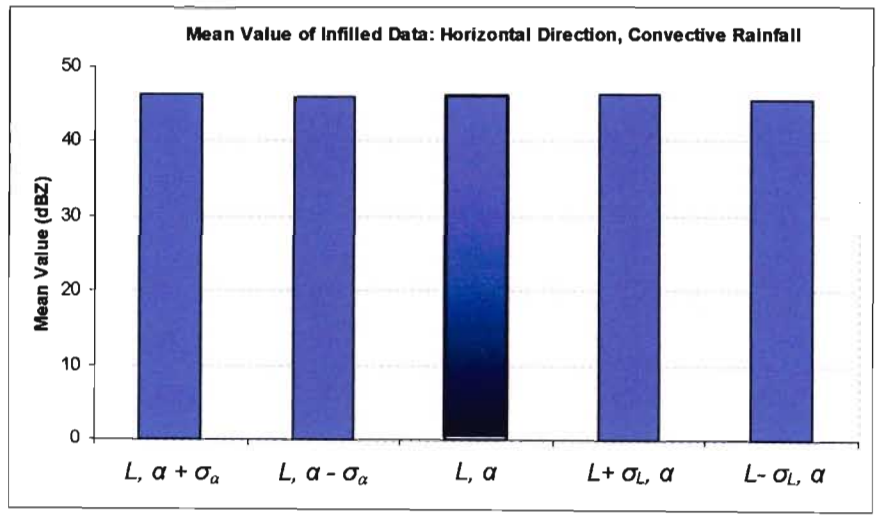


Figure 4.15: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (19 December 2001) with 101 targets being infilled.

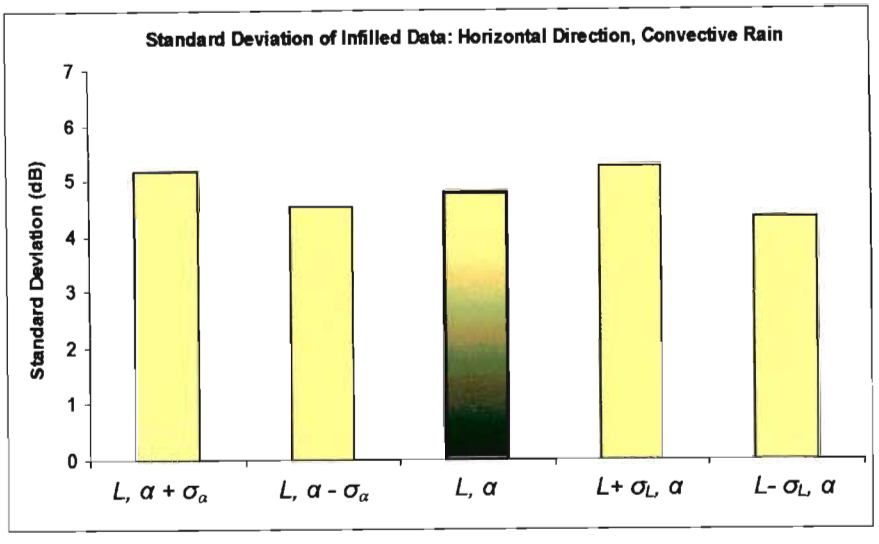


Figure 4.16: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (19 December 2001) with 101 targets being infilled.

Table 4.5: Summary of statistics returned for sensitivity analysis test for convective rainfall in the horizontal direction. Test data were from the Bethlehem weather radar (19 December 2001) and consists of 101 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	46.2	45.9	46.0	46.2	45.4
Median	46.0	45.3	45.5	45.7	44.8
Standard Deviation	4.8	3.3	3.5	3.4	3.3
Range	20.9	13.3	14.1	13.6	13.6
Minimum	36.8	38.7	38.2	38.5	38.1
Maximum	57.7	52.0	52.3	52.2	51.7

The results for the two sensitivity analysis tests for convective rainfall in the horizontal direction indicate there is no significant difference between the statistics returned for the 5 different combinations of semivariogram parameters tested. A greater correspondence existed between the semivariogram parameters tested for the 19 December 2001 data set than the 24 January 2002. Although both returned results indicating no significant difference in the statistics retuned for each set of semivariogram parameters.

HORIZONTAL SENSITIVTY ANALYSIS: STRATIFORM RAINFALL

The tests were carried out in a stratiform rainfall segments selected from instantaneous reflectivity images. The first test data set was taken from the Bethlehem weather radar (24 January 2002) with 149 targets being estimated. The second data set was also from the Bethlehem weather radar (11 December 2000) with 145 targets being estimated. The targets for both data sets were infilled using the 25 nearest controls. A summary of the statistics for both are given in Figures 4.17 to 4.20 and Tables 4.6 and 4.7.

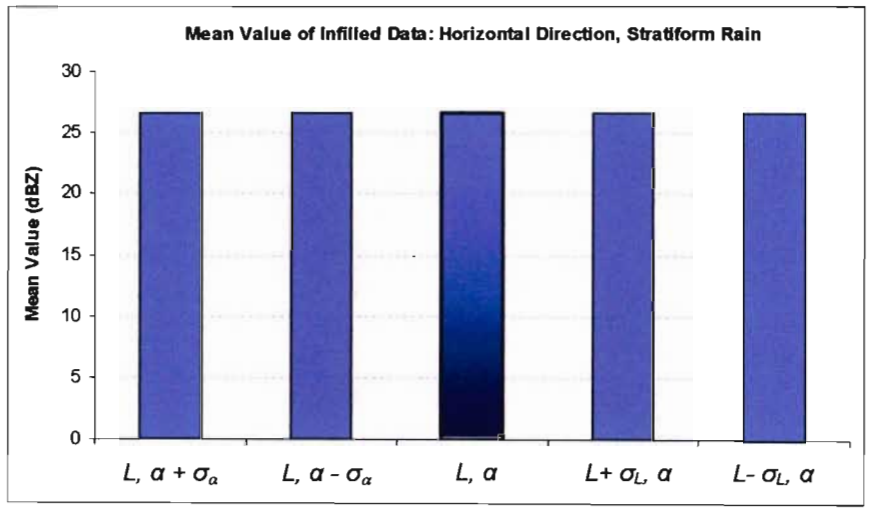


Figure 4.17: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (24 January 2002) with 149 targets being infilled.

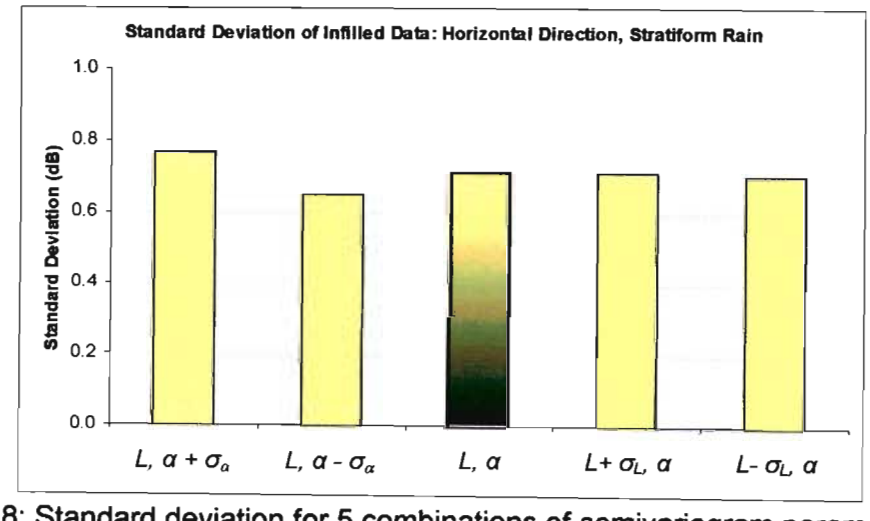


Figure 4.18: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (24 January 2002) with 149 targets being infilled.

Table 4.6: Summary of statistics returned for sensitivity analysis test for stratiform rainfall in the horizontal direction. Test data were from the Bethlehem weather radar (24 January 2002) and consists of 149 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	28.5	29.0	28.8	28.8	28.8
Median	28.7	29.3	29.0	29.0	29.0
Standard Deviation	2.0	1.6	1.8	1.8	1.7
Range	7.3	6.2	6.8	6.8	6.7
Minimum	24.7	25.3	25.0	25.0	25.0
Maximum	32.0	31.5	31.8	31.8	31.7

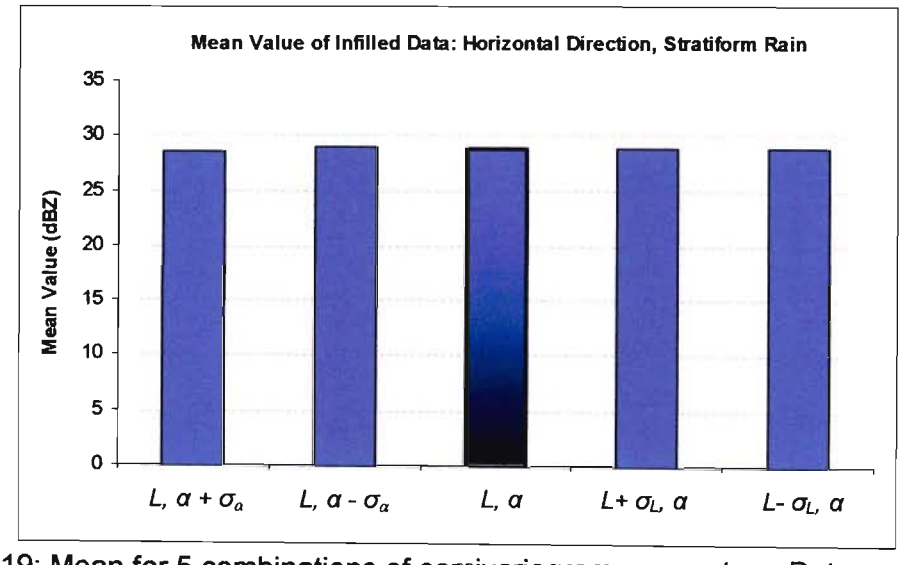


Figure 4.19: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (11 December 2000) with 145 targets being infilled.

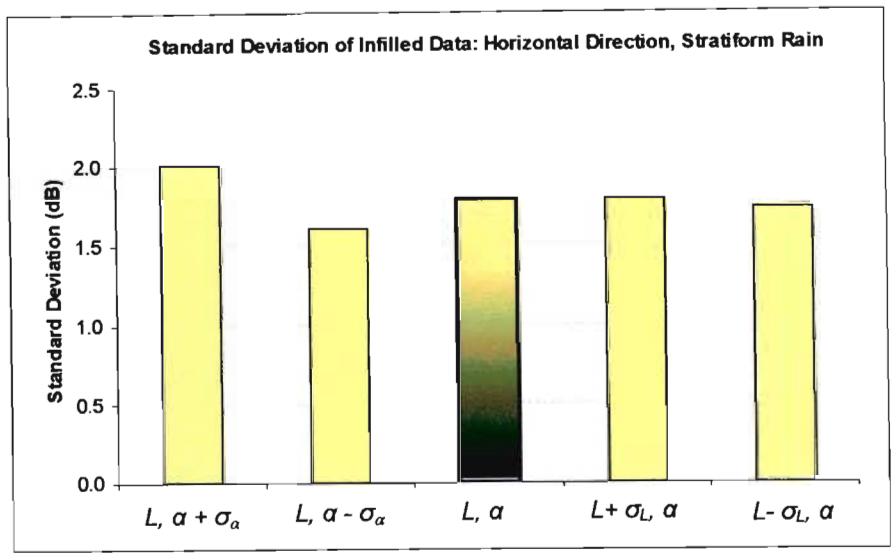


Figure 4.20: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (11 December 2000) with 145 targets being infilled.

Table 4.7: Summary of statistics returned for sensitivity analysis test for stratiform rainfall in the horizontal direction. Test data were from the Bethlehem weather radar (11 December 2000) and consists of 145 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	26.6	26.7	26.6	26.6	26.7
Median	26.5	26.6	26.6	26.5	26.6
Standard Deviation	0.77	0.65	0.71	0.71	0.70
Range	3.75	3.38	3.57	3.57	3.55
Minimum	25.3	25.4	25.4	25.4	25.4
Maximum	29.0	28.8	28.9	28.9	28.9

The summary statistics of the two tests for sensitivity in the horizontal direction for stratiform rainfall indicate little difference of the infilled data for the semivariogram parameters tested. The results returned for the stratiform rainfall parameters in the horizontal direction indicate the infilled values for stratiform rainfall are less sensitive to the range of expected semivariogram parameters than the parameters for convective rainfall.

VERTICAL SENSITIVTY ANALYSIS: CONVECTIVE RAINFALL

The tests were carried out in a convective rainfall segment selected from instantaneous reflectivity images. The first test data set was taken from the Bethlehem weather radar (24 January 2002) with 121 targets being estimated. The second data set was also from the Bethlehem weather radar (19 December 2001) with also 121 targets being estimated. The targets for both data sets were infilled using the 10 nearest controls in the vertical direction. A summary of the statistics for both are given in Figures 4.21 to 4.24 and Tables 4.8 and 4.9.

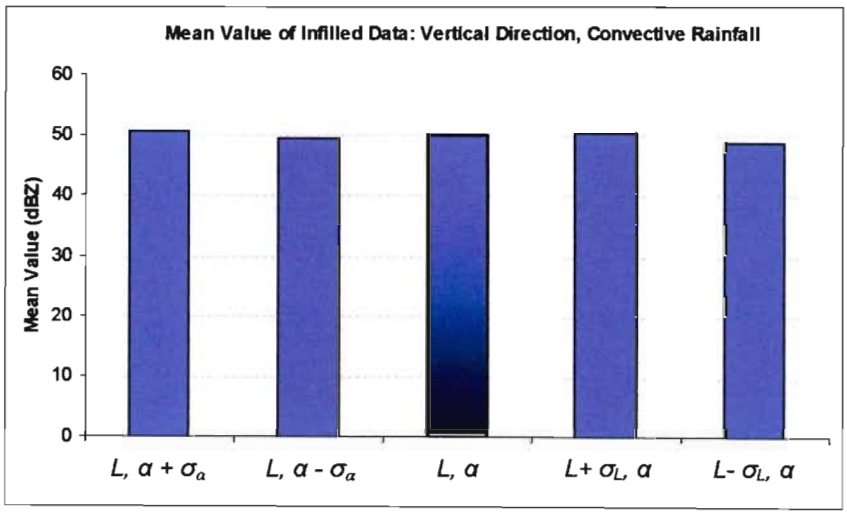


Figure 4.21: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (24 January 2002) with 121 targets being infilled.

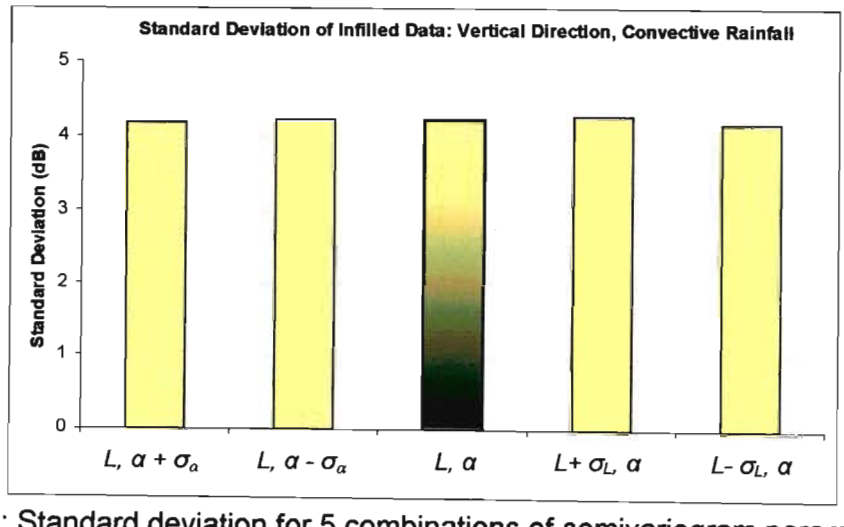


Figure 4.22: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (24 January 2002) with 121 targets being infilled.

Table 4.8: Summary of statistics returned for sensitivity analysis test for convective rainfall in the vertical direction. Test data were from the Bethlehem weather radar (24 January 2002) and consists of 121 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	50.8	49.5	50.0	50.6	48.9
Median	50.5	48.9	49.7	50.4	48.9
Standard Deviation	4.19	4.23	4.21	4.27	4.18
Range	20.5	20.2	20.6	21.4	19.3
Minimum	40.1	38.7	39.1	39.0	38.8
Maximum	60.7	58.9	59.7	60.5	58.1

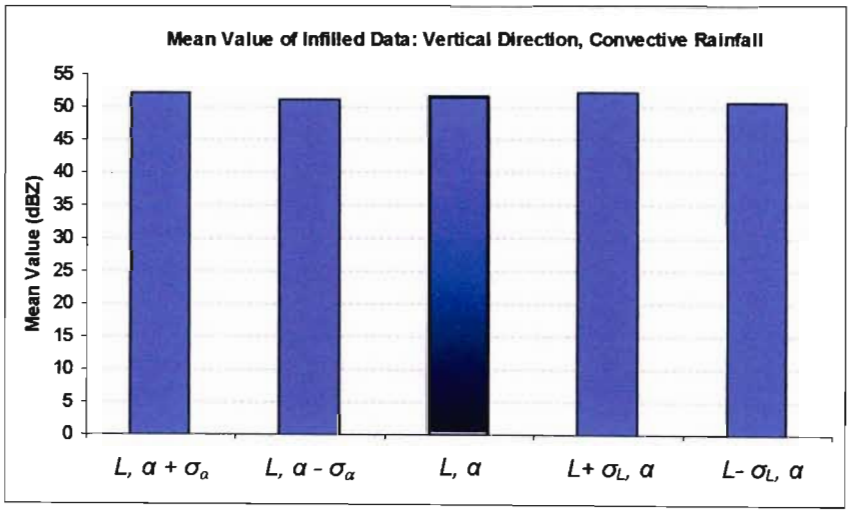


Figure 4.23: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (19 December 2001) with 121 targets being infilled.

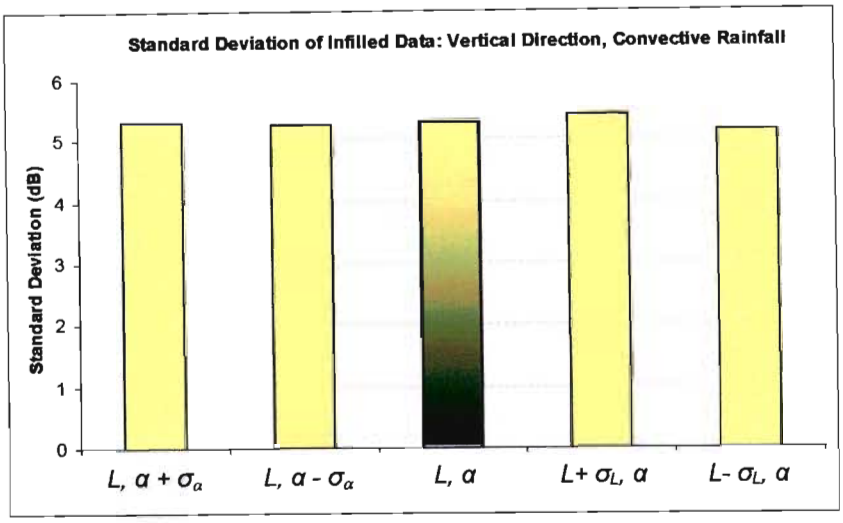


Figure 4.24: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (19 December 2001) with 121 targets being infilled.

Table 4.9: Summary of statistics returned for sensitivity analysis test for convective rainfall in the vertical direction. Test data were from the Bethlehem weather radar (19 December 2001) and consists of 121 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	52.1	51.1	51.6	52.1	50.6
Median	52.9	51.9	52.3	52.8	51.2
Standard Deviation	5.30	5.26	5.30	5.40	5.14
Range	27.9	27.3	27.6	27.8	26.9
Minimum	35.1	34.5	34.8	35.3	34.1
Maximum	63.0	61.8	62.4	63.1	61.0

The results indicate that the semivariogram parameters are once again relatively insensitive to the range tested for the vertical direction. However the test results do indicate a greater sensitivity to the shape parameter (α) than changes in the correlation lengths (L) for the test undertaken.

VERTICAL SENSITIVTY ANALYSIS: STRATIFORM RAINFALL

The tests were carried out in stratiform rainfall segments selected from instantaneous reflectivity images. The first test data set was taken from the Bethlehem weather radar (11 December 2000) with 121 targets being estimated. The second data set was also from the Bethlehem weather radar (19 May 1995) with 121 targets being estimated. The targets for both data sets were infilled using the 10 nearest vertical controls. A summary of the statistics for both are given in Figures 4.25 to 4.28 and Tables 4.10 and 4.11.

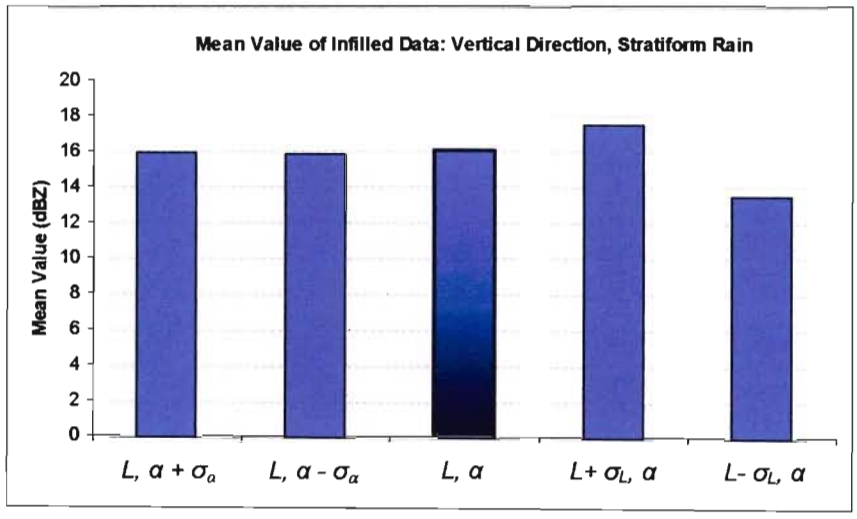


Figure 4.25: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (11 December 2000) with 121 targets being infilled.

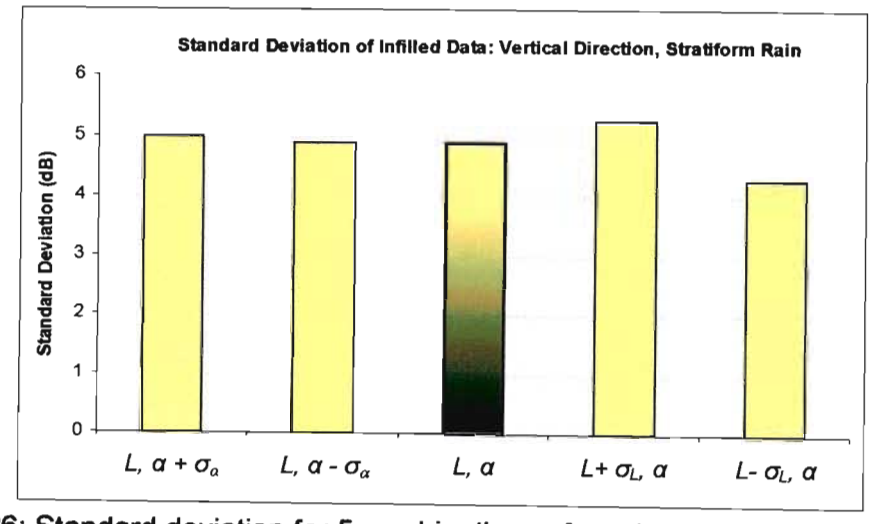


Figure 4.26: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (11 December 2000) with 121 targets being infilled.

Table 4.10: Summary of statistics returned for sensitivity analysis test for stratiform rainfall in the vertical direction. Test data were from the Bethlehem weather radar (11 December 2000) and consists of 121 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	15.9	15.9	16.1	17.5	13.5
Median	16.8	16.6	17.0	18.5	14.2
Standard Deviation	4.99	4.89	4.91	5.26	4.28
Range	24.7	23.2	22.9	24.5	19.9
Minimum	0.0	0.0	0.0	0.0	0.0
Maximum	24.7	23.2	22.9	24.5	19.9

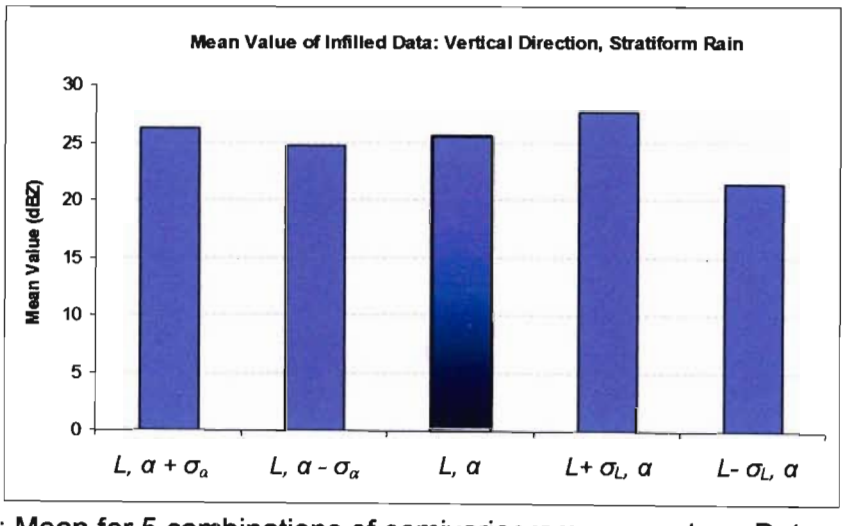


Figure 4.27: Mean for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (19 May 1995) with 121 targets being infilled.

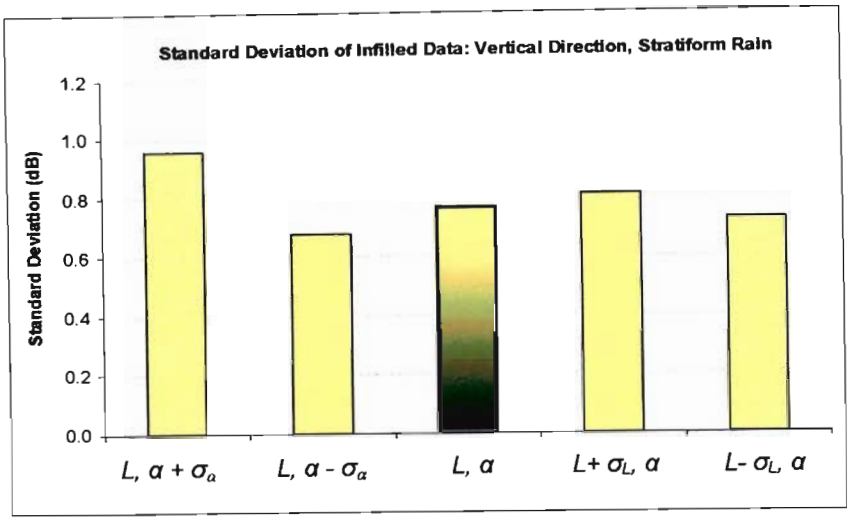


Figure 4.28: Standard deviation for 5 combinations of semivariogram parameters. Data was from the Bethlehem weather radar (19 May 1995) with 121 targets being infilled.

Table 4.11: Summary of statistics returned for sensitivity analysis test for stratiform rainfall in the vertical direction. Test data were from the Bethlehem weather radar (19 May 1995) and consists of 121 targets.

	$L, \alpha+\sigma_\alpha$	$L, \alpha-\sigma_\alpha$	L, α	$L+\sigma_L, \alpha$	$L-\sigma_L, \alpha$
Mean	26.3	24.7	25.5	27.6	21.4
Median	26.3	24.8	25.6	27.6	21.4
Standard Deviation	0.96	0.68	0.76	0.81	0.73
Range	4.6	3.1	3.5	3.9	3.2
Minimum	24.2	23.3	23.8	25.6	19.9
Maximum	28.7	26.4	27.3	29.5	23.0

The results for the stratiform rainfall from that both data sets indicate that the infilled Kriged values are relatively insensitive to the range of semivariogram parameters expected for stratiform rainfall conditions. Overall there is no indication that there any significant difference between the infilled data sets for the five sets of semivariogram parameter values used in each instance, giving a clear indication that the final Kriged solution is reasonably insensitive to the range of the shape parameter, α, and correlation length, L, values that can possibly occur when a sample semivariogram is fitted to a specific rainfall type. Instead of solving for each set of semivariogram parameters for each

neighbourhood on-line, one can simply use the centroid value computed for each rainfall type depending on the observed rainfall.

4.4.6 Application to 3D Reflectivity Data

When selecting controls for computing the estimated value of a target in 3D space, three possible combinations of points can be selected:

- all convective,
- mixed, containing both stratiform and convective points and
- all stratiform

For situations with either all convective or all stratiform, the appropriate subset of parameters in Table 4 (Section 4.7, page 139) is utilised for all estimations.

When it comes to the practical calculation of the semivariogram elements in the Kriging equations there are no numerical or technical difficulties if the controls are all of one type, either convective, or stratiform. A problem arises when the controls are of mixed type. It was found that if the individual semivariogram elements between controls appears in the matrix *G* in the Kriging equation (Equation (3.2)) were set to their common type, with a compromise value for a mixed pair, unpredictable instabilities occurred in the solution of the equation.

The novel methodology adopted to solve this problem was to use a fixed linear weighting of the α and L parameters for all semivariogram computations for a given target, where the weights are in proportion to the type. The equations to compute the targets appropriate correlation length and new shape parameter in the horizontal and vertical directions are given by Equations (4.10), (4.11) and (4.12):

$$L_H = \frac{3.38 \cdot \sum C_{pixels} + 8.40 \cdot \sum S_{pixels}}{\sum Controls} \tag{4.10}$$

$$L_V = \frac{4.11 \cdot \sum C_{pixels} + 2.56 \cdot \sum S_{pixels}}{\sum Controls} \tag{4.11}$$

$$\alpha = \frac{1.78 \cdot \sum C_{pixels} + 1.43 \cdot \sum S_{pixels}}{\sum Controls} \tag{4.12}$$

where L_H and L_V is the target specific correlation length in the horizontal and vertical direction and α the shape parameter. ΣC_{pixels} is the number of convective pixels in the controls, ΣS_{pixels} is the number of stratiform pixels and $\Sigma Controls$ is the total number of controls selected. For example in a 3D situation if a total number of 25 controls are selected of which 15 are of convective pixels and 10 stratiform pixels, the computed semivariogram parameters would be L_H = 5.39 km, L_V = 3.49 km and α = 1.64, and these values would be used for all calculations of $\gamma(h)$ in the Kriging equations.

When Kriging with a 3D data set, the semivariogram model developed by Seed and Pegram (2001) was modified and is given by Equation (4.12):

$$\gamma(h) = \sigma^2 \cdot \left[1 - \exp\left(-h^{\alpha}\right)\right] \tag{4.12}$$

in which σ^2 is the field variance, which reduces to unity when applied to standardised data and α is the scaling exponent. The term h in Equation (4.12) is computed by Equation (4.13):

$$h^2 = \left(\frac{r}{L_H}\right)^2 + \left(\frac{z}{L_V}\right)^2 \tag{4.13}$$

$$r^2 = x^2 + y^2$$

Where h is the scaled distance in spherical co-ordinates, r is the distance in the horizontal direction, L_H the horizontal correlation length, z is the distance in the vertical direction, L_V the vertical correlation length and x and y are the distances in the Cartesian horizontal direction.

4.5 FURTHER DEVELOPMENT OF KRIGING METHODS

An overview of the Kriging techniques developed to infill the missing and contaminated data in the radar volume scan is given in this section. Ordinary and Universal Kriging are discussed in their application to radar reflectivity infilling and how they are utilised in different types of rainfall. The full algorithm to infill missing data and provide an estimate of the rainfall at ground level is presented.

In Ordinary Kriging, the mean is assumed constant and unknown throughout the field whereas in Universal Kriging, the mean is also assumed to be unknown but varying (Chiles and Delfiner, 1999: 151). A detailed explanation of the application Universal Kriging in this context is provided in Wesson and Pegram (2005) contained in Section 4.7 of this chapter, pages 140 to 141. Universal Kriging allows any trends or variations in the mean to be modelled by the addition of external variables to the coefficient matrix, in Equation (3.2). Typical models for the variation of the mean are either linear or polynomial; alternatively, variables associated with physical characteristics of the controls, that influence the mean value of the field may be included.

4.5.1 Computational Stability and Efficiency in 3D Data Sets

As discussed in Chapter 3 there is a high computational burden associated with the Kriging technique. In order to infill all missing data contained in the radar volume scan and provide an estimate of the rainfall at ground level in real time, the data need to be infilled in substantially less than five minutes, which is the interval at which radar volume scans are processed in South Africa.

As mentioned in Chapter 3 it was decided that a nearest neighbours approach would be the most appropriate to use in order to ensure computational efficiency and accuracy. In application, the search strategy was adapted as to work in 3D space. From each individual target a search for controls is initiated working outwards in increasing spherical patterns until the required number of nearest controls are located.

The CAPPI 3 km above ground level was selected as the testing level to determine the optimum number of controls. The data from aloft was used to estimate the reflectivity values on the 3 km CAPPI; the observed reflectivity data were then compared to the estimated reflectivity data. This process was repeated for differing numbers of controls ranging from 5 to 50 over ten different instantaneous reflectivity images from the Bethlehem weather radar, which were selected from the 1995 and 1996 period.

To compare the estimated and observed reflectivity values the Mean Sum of Square of Errors (MSSE), mean value and standard deviation of the reflectivity values were computed for each set of controls used from 5 through to 50. The northern half of the image only was considered in the analysis due to ground clutter contamination in the southern half, which resulted in 15370 targets being estimated for each of the

instantaneous reflectivity images. An example of two of the ten instantaneous reflectivity images is illustrated in Figure 4.29 where the observed and estimated reflectivity values are indicated.

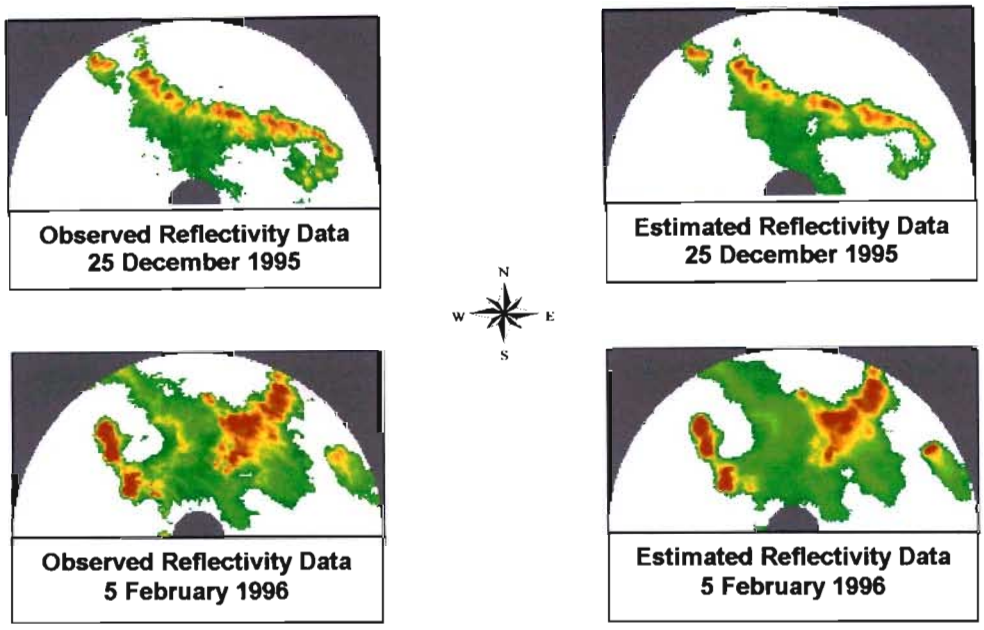


Figure 4.29: Observed and estimated reflectivity values at the CAPPI 3 km above ground level for instantaneous reflectivity images from the Bethlehem weather radar from the 25 December 1995 and 5 February 1996. The sizes of the images are 200 km by 100 km.

Figure 4.30 indicates the average MSSE between observed and estimated reflectivity values computed from 10 different instantaneous images with controls ranging from 5 to 50. As indicated in Figure 4.30 the average MSSE increases dramatically once less than 20 controls are used to infill a target. The minimum MSSE is at 30 controls however for computational efficiency reasons it was decided that less than 30 controls would be utilised to infill a target. Based on Figure 4.30 it was decided that the controls to infill a target would be limited to 25 values as a conservative estimate to ensure computational efficiency and accuracy.

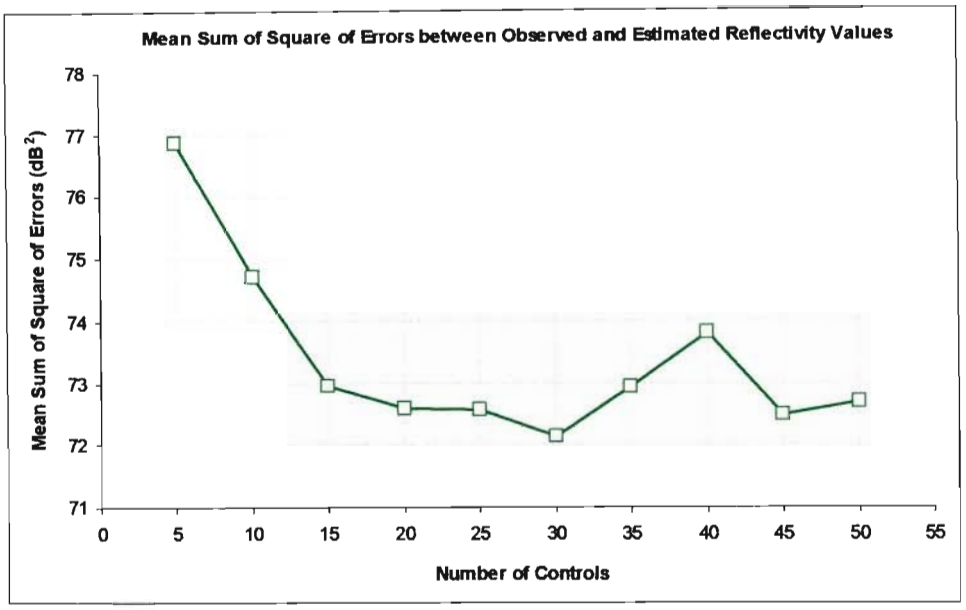


Figure 4.30: Average MSSE (dB^2) between observed and estimated reflectivity values for ten instantaneous images using controls ranging from 5 to 50 to estimate the reflectivity values on the 3 km CAPPI.

The SSE between the estimated and observed mean and standard deviation values for controls ranging from 5 to 50 was also computed to help determine the optimum number of controls. Figure 4.31 indicates the plots of the SSE between the means and standard deviations. The plot of the SSE between estimated and observed mean values indicates a reasonably constant SSE when utilising controls in the range of 15 to 25. The plot of the SSE between observed and estimated standard deviation values indicates that the optimum number of controls to utilise is 15. It was decided that 25 controls would be utilised due to the low MSSE returned as indicated in Figure 4.30 and the low SSE returned between the observed and estimated mean values as indicated in Figure 4.31.

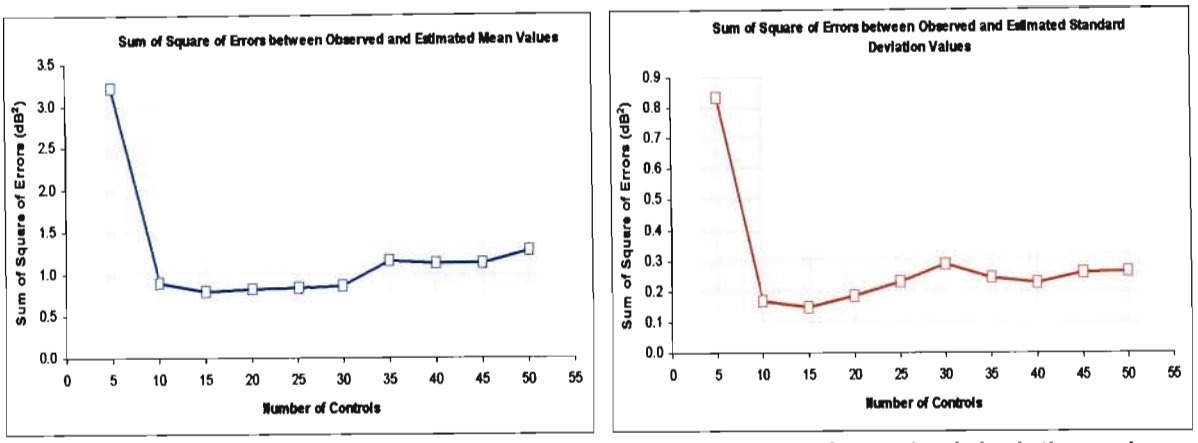


Figure 4.31: SSE between estimated and observed mean and standard deviation values for controls ranging from 5 to 50.

To determine the optimum percentage trimming of the singular values (w_i) in the calculation of the inverse of the coefficient matrix using SVD for 25 controls a similar approach as described in Chapter 3 was employed. In this instance the testing was carried out in on a 3D data set where, as described above, the reflectivity values on the 3 km CAPPI was estimated and the compared to the observed reflectivity values. The number of controls was fixed at 25 and the number of singular values progressively trimmed based on the cumulative sum of the singular values squared (s_j), as computed by Equation (3.5). Five different instantaneous images were tested that contained both stratiform and convective rainfall; in determining the appropriate s_j, the MSSE between observed and estimated reflectivity values was computed at each selected trimming percentage and the sum of the weighting values examined to determine if the weights summed to unity for each target that required infilling.

Figure 4.32 indicates the MSSE determined from the 5 instantaneous images tested at each percentage point trimmed. The instantaneous images were taken from the Bethlehem weather radar over the 1995 and 1996 year periods, where the number of targets infilled via Kriging ranged from 7651 to 14063. The results indicate that the sum of difference squared is reasonably constant until the 99.995% threshold point is reached after which the MSSE increases dramatically illustrating a decrease in accuracy of the Kriging estimate.

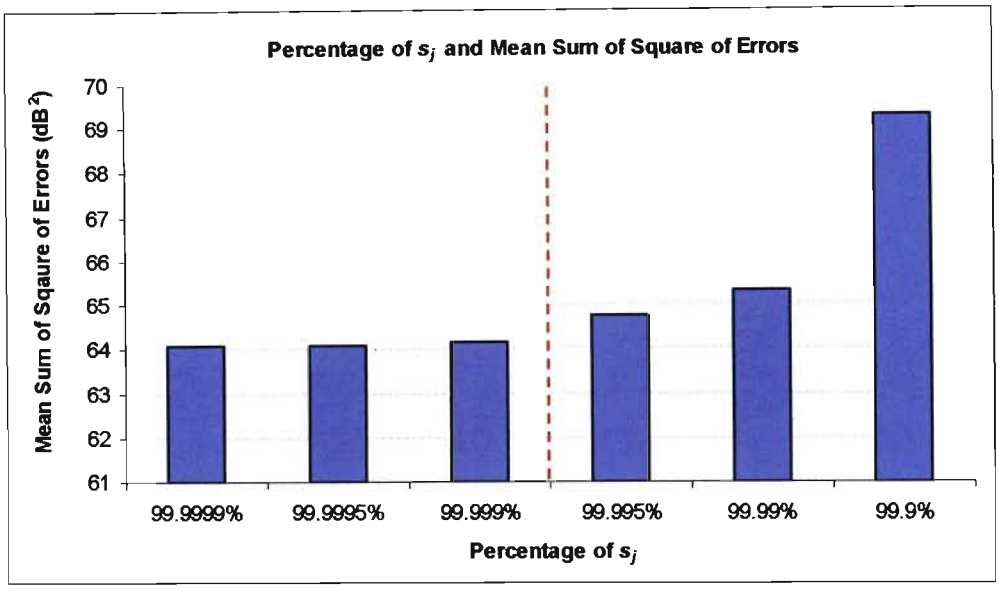


Figure 4.32: MSSE between observed and estimated data for different trimming of singular values from the W matrix at set percentage values of s_j.

The average percentage difference of the weights summing to unity was also computed at various s_j percentage positions. Indicated in Figure 4.33 is the percentage difference of weights summing to unity, as the singular values in the coefficient matrix is progressively trimmed a point is reached where the weights no longer sum to unity and too many of the singular values have been removed, as indicated in Figure 4.33 after the trimming of the singular values reaches or exceeds a percentage sum of 99.995% the weights no longer sum to unity and will return unstable solutions.

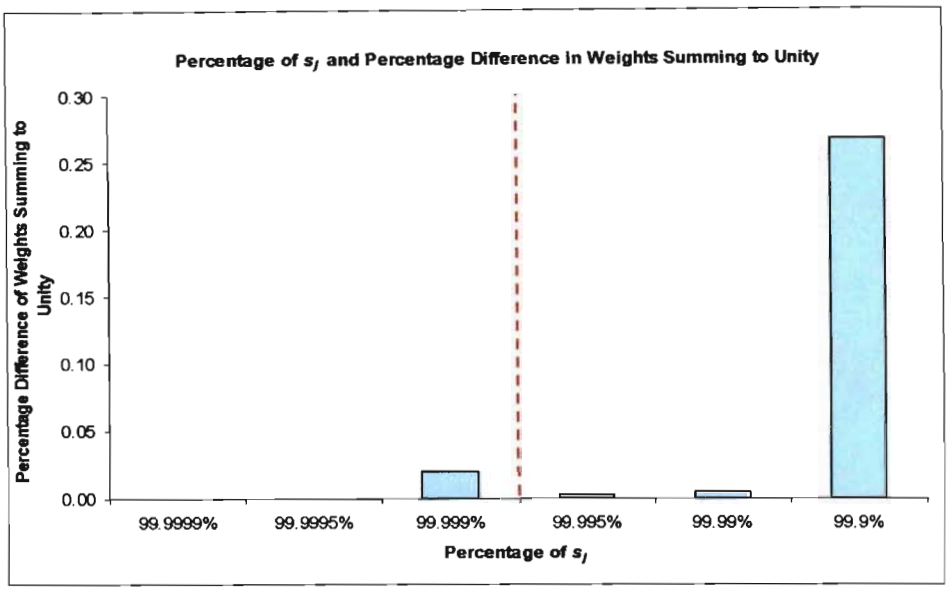


Figure 4.33: Percentage difference of Kriging weights summing to unity and the percentage of s_j that correspond to the point where the singular values are trimmed.

4.5.2 Development of Ordinary and Universal Kriging Techniques

Kriging was chosen as the computational method to interpolate the observations missing internal to the CAPPI stack and to extrapolate the reflectivity values contained aloft to ground level. Kriging is considered to be the optimal technique for the spatial prediction of Gaussian data (Cressie, 1993: 106). In this application two types of Kriging were used, Ordinary and Universal Kriging (also called External Drift Kriging (Hengl *et al*, 2003)).

A description of Ordinary Kriging is contained in Chapter 3; a brief overview of Universal Kriging and its application in the three different rainfall zones will now be elaborated upon and tested in terms of its effectiveness in improving the estimates over those provided by Ordinary Kriging. The rainfall can be divided into three zones in terms of the controls selected:

- Stratiform Rainfall - controls consist entirely of stratiform rain.
- Mixed Rainfall - controls consist of a combination of stratiform and convective rain.
- Convective Rainfall - controls consist entirely of convective rain.

STRATIFORM ZONE

Universal Kriging was investigated by modelling the vertical variation of the mean by use of either a second order polynomial ($y = a \cdot x_1 + b \cdot x_2^2 + c$) or a linear ($y=a \cdot x+b$)

relationship to take advantage of the trend in the vertical reflectivity profile. Ordinary Kriging and Universal Kriging, with linear and second order polynomial trends, were compared on 5 different instantaneous reflectivity images. Four of which were taken from the Bethlehem weather radar from the years 2001 and 2002, and one from the Durban weather radar from the year 2000. The number of stratiform targets estimated over the five images ranged from 3628 to 4533 pixels.

The results are indicated in Figure 4.34 where it is evident that Ordinary Kriging provides a far better estimate in the stratiform rainfall zone than Universal Kriging. In Universal Kriging estimates by adding a linear trend resulted in an overestimate of the rainfall and produced a far poorer estimate than Ordinary Kriging. A second order polynomial relationship was also used to model the trend or variation in the mean value but on the 5 images tested did not provide an improved estimate over Ordinary Kriging in either of the images. Universal Kriging in this configuration does not seem to provide an improvement in the estimates over Ordinary Kriging. Thus it was decided that when all the rainfall values consist of stratiform controls Ordinary Kriging should be used to compute the estimate.

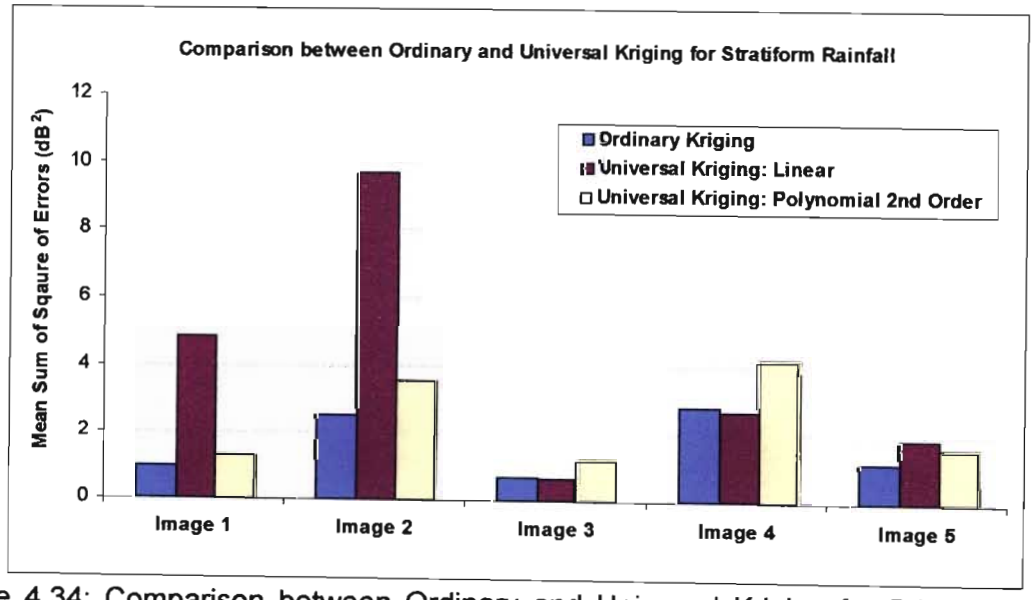


Figure 4.34: Comparison between Ordinary and Universal Kriging for 5 instantaneous reflectivity images for stratiform rainfall. The images are compared in terms of the MSSE.

When computing Kriging estimates from stratiform rainfall SVD does not need to be used to compute the inverse of the coefficient matrix. This is due to the fact that the lower

shape parameter (α) is close to exponential and results in a stable coefficient matrix. This can be demonstrated by computing the condition number for the coefficient matrices corresponding to different rainfall types; the condition number is defined as the ratio of the largest to the smallest singular value (Wilkinson, 1988: 191). This is illustrated in Figure 4.35 where the condition number for coefficient matrices corresponding to different rainfall types are ranked and plotted, the image from which the condition number is computed is from the Bethlehem weather radar (5 February 1996).

The coefficient matrices computed from convective rainfall are the highest and represent the most numerically unstable matrices. The coefficient matrices computed from the mixed rainfall zones vary from high, consisting of predominantly convective rainfall, to low, consisting of predominantly stratiform rainfall. The coefficient matrices for stratiform rainfall are consistently low. Based on the computed condition numbers it was decided that LU decomposition would be used in the stratiform rainfall zones and SVD would be used in the mixed and convective rainfall zones with a concomitant saving in computation time.

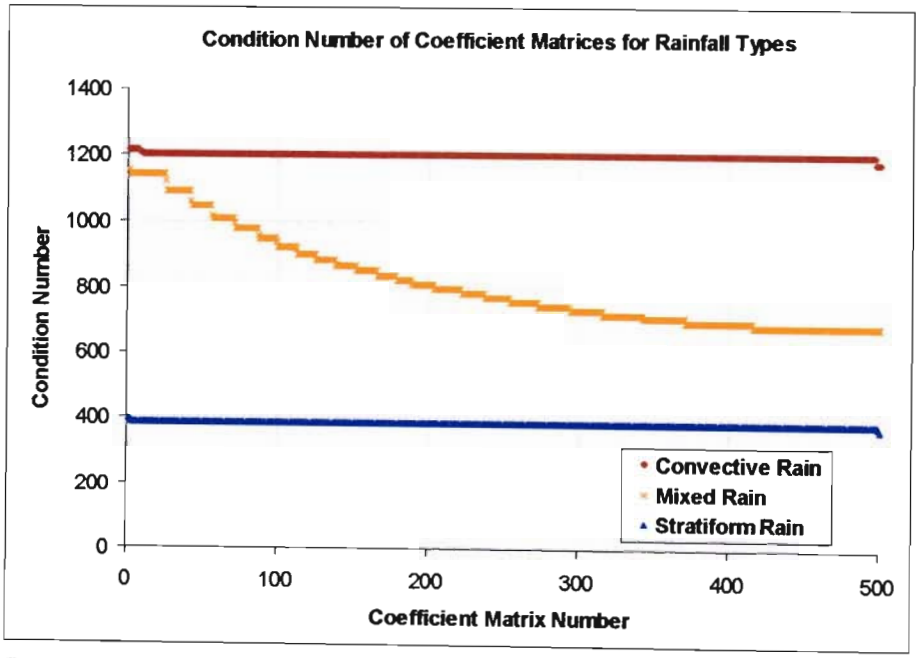


Figure 4.35: Plot of the ranked condition number for 500 coefficient matrices each of size 25 by 25 for convective, mixed and stratiform rainfall from a typical instantaneous reflectivity image from the Bethlehem weather radar.

CONVECTIVE ZONE

In rainfall where the controls consist entirely of convective pixels Universal Kriging was also investigated in computing the Kriging estimate. Linear and second order polynomial trends were investigated to determine improvements in the Kriged estimates. The MSSE between the observed and estimated rainrate values for Ordinary Kriging and Universal Kriging were computed on five different instantaneous images from the Bethlehem weather radar from the years 1999, 2001 and 2002. The number of convective targets over the 5 images ranged from 296 to 1104. The results are indicated in Figure 4.36 where Ordinary Kriging is shown to provide a better estimate than Universal Kriging on all of the five instantaneous images tested. The results indicate that Ordinary Kriging provides the best estimate of convective rainfall. SVD is once again used to compute the inverse of the coefficient matrix.

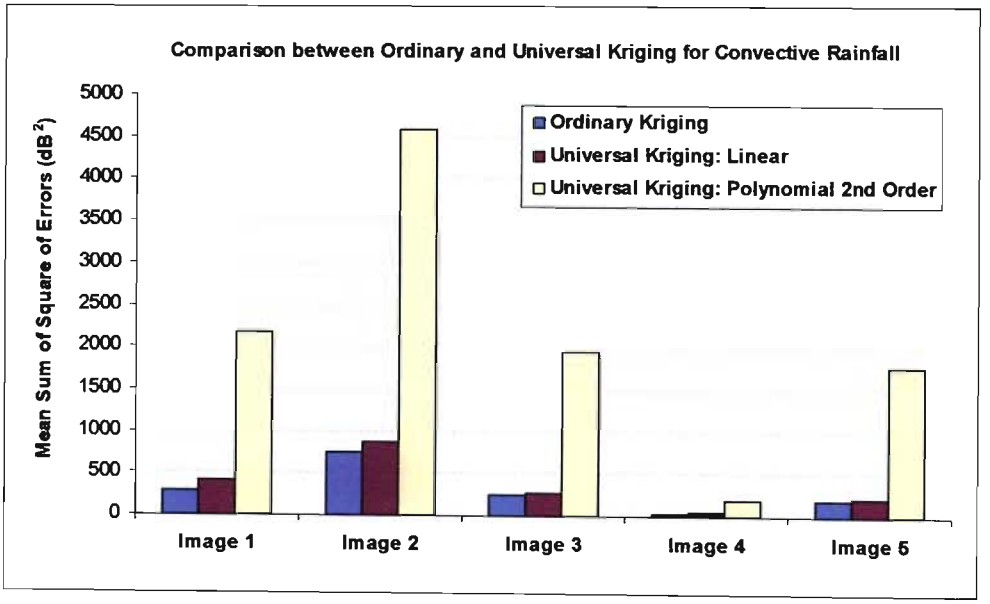


Figure 4.36: Comparison between Ordinary and Universal Kriging for 5 instantaneous reflectivity images for convective rainfall. The images are compared in terms of the MSSE.

MIXED ZONE

The mixed zone comprises a mixture of controls that have been classified as either convective or stratiform rainfall. In this instance Universal Kriging did provide an improvement in the estimates over Ordinary Kriging. Since the controls now consist of both stratiform and convective pixels a stratiform/convective binary switch was used as the two external variables in Universal Kriging. This is indicated in Equation (4.13):

$$Q_c = \begin{bmatrix} C_1 & S_1 \\ \vdots & \vdots \\ C_n & S_n \end{bmatrix} \tag{4.13}$$

where: convective pixel: *[C S] = [1 0]*

stratiform pixel: *[C S] = [0 1]*

Figure 4.37 illustrates the results when applying Universal and Ordinary Kriging to the targets having mixed controls in 10 selected radar reflectivity images consisting of a variety of rainfall types from the Bethlehem weather radar ranging over the years 1995, 1996, 2001 and 2002. The number of targets that required estimating over the 10 images ranged from 953 up to 4204 pixels. Figure 4.37 indicates Universal Kriging provides an improvement in the estimates provided by Ordinary Kriging in 8 of the 10 images tested and in 3 instances the results indicate a dramatic improvement in estimates.

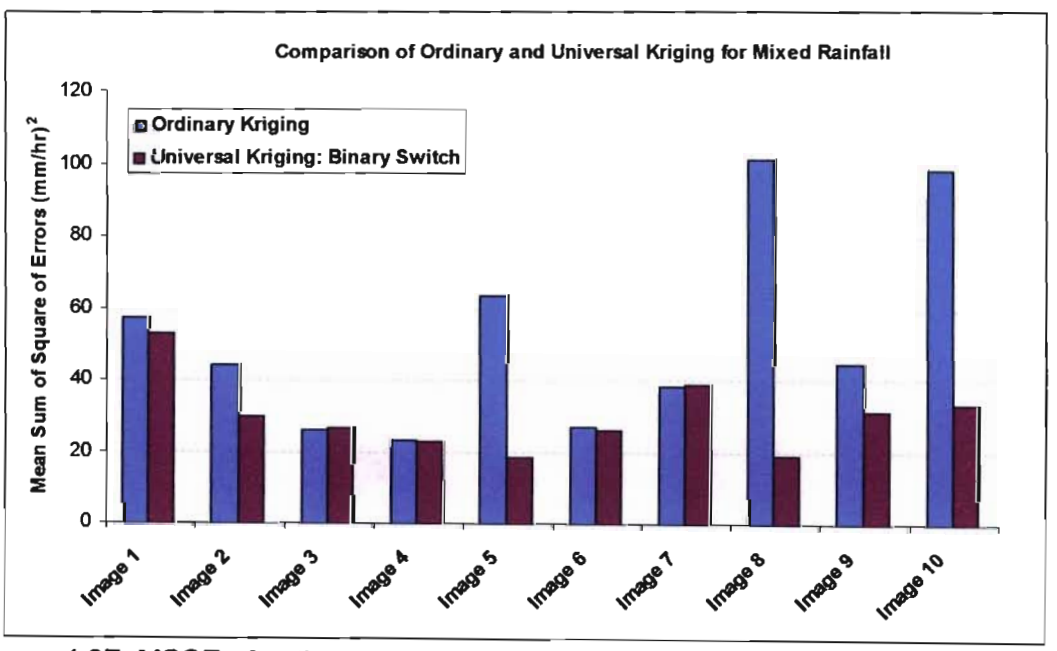


Figure 4.37: MSSE of estimates provided from Ordinary and Universal Kriging for 10 instantaneous reflectivity images for mixed rainfall.

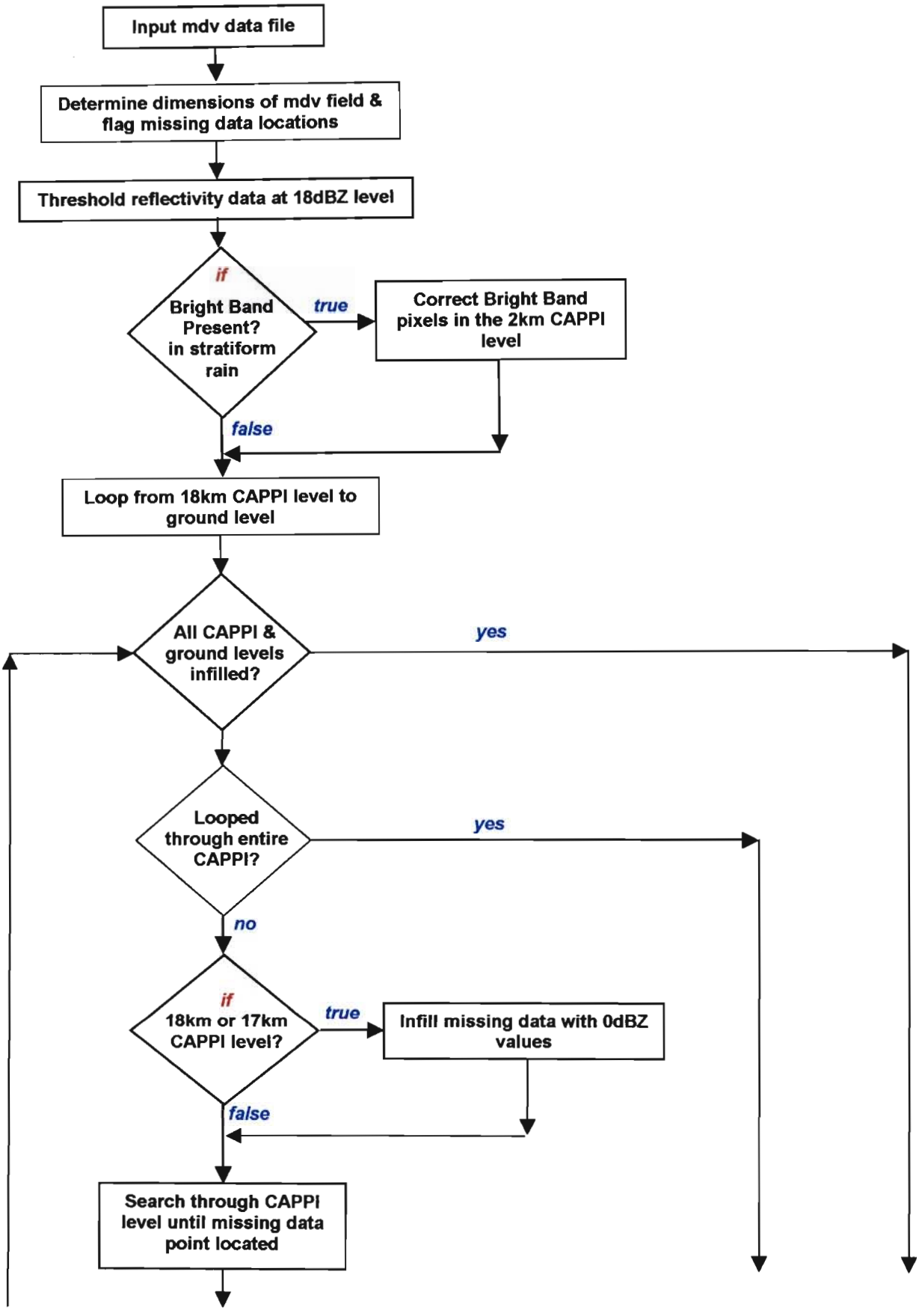
4.5.3 Cascade Kriging

When Kriging directly to ground level using the full set of data in a CAPPI stack, unexpected problems occur. Serious discontinuities occur at the regions located directly under the edges of the CAPPIs and also an inflation of the reflectivity estimates due to the

numerical and geometric distribution of Kriging weights at the CAPPI edge locations. Illustrations and further details of this are contained in Wesson and Pegram (2005) contained in Section 4.7, pages 147 to 148.

To overcome these difficulties and problems encountered with the CAPPI edges, a novel algorithm entitled Cascade Kriging was developed. The algorithm developed starts at the uppermost CAPPI and infills all missing data using neighbourhood Kriging of the 25 controls vertically above and in the horizontal direction. Once all the unknown data on that CAPPI have been estimated and infilled, the CAPPI directly below is examined and any missing data here are estimated in the same manner; once again controls from above (which now can include previously infilled data) and at the same level as the target are used as control data. This is repeated until all the unknown data from in each CAPPI are estimated. The final step involves an estimate of the rainfall at ground level.

Further details and illustrations of the Cascade Kriging algorithm are contained in Wesson and Pegram (2005) in Section 4.7, pages 148 to 152. A flow chart of the complete algorithm is illustrated in Figure 4.38. Contained in the Appendix is the computer code for the operational infilling of all missing data contained in a radar volume scan.


Input mdv data file
Determine dimensions of mdv field & flag missing data locations
Threshold reflectivity data at 18dBZ level
if
Bright Band Present? in stratiform rain
true
Correct Bright Band pixels in the 2km CAPPI level
false
Loop from 18km CAPPI level to ground level
All CAPPI & ground levels infilled?
yes
Looped through entire CAPPI?
yes
no
if
18km or 17km CAPPI level?
true
Infill missing data with 0dBZ values
false
Search through CAPPI level until missing data point located

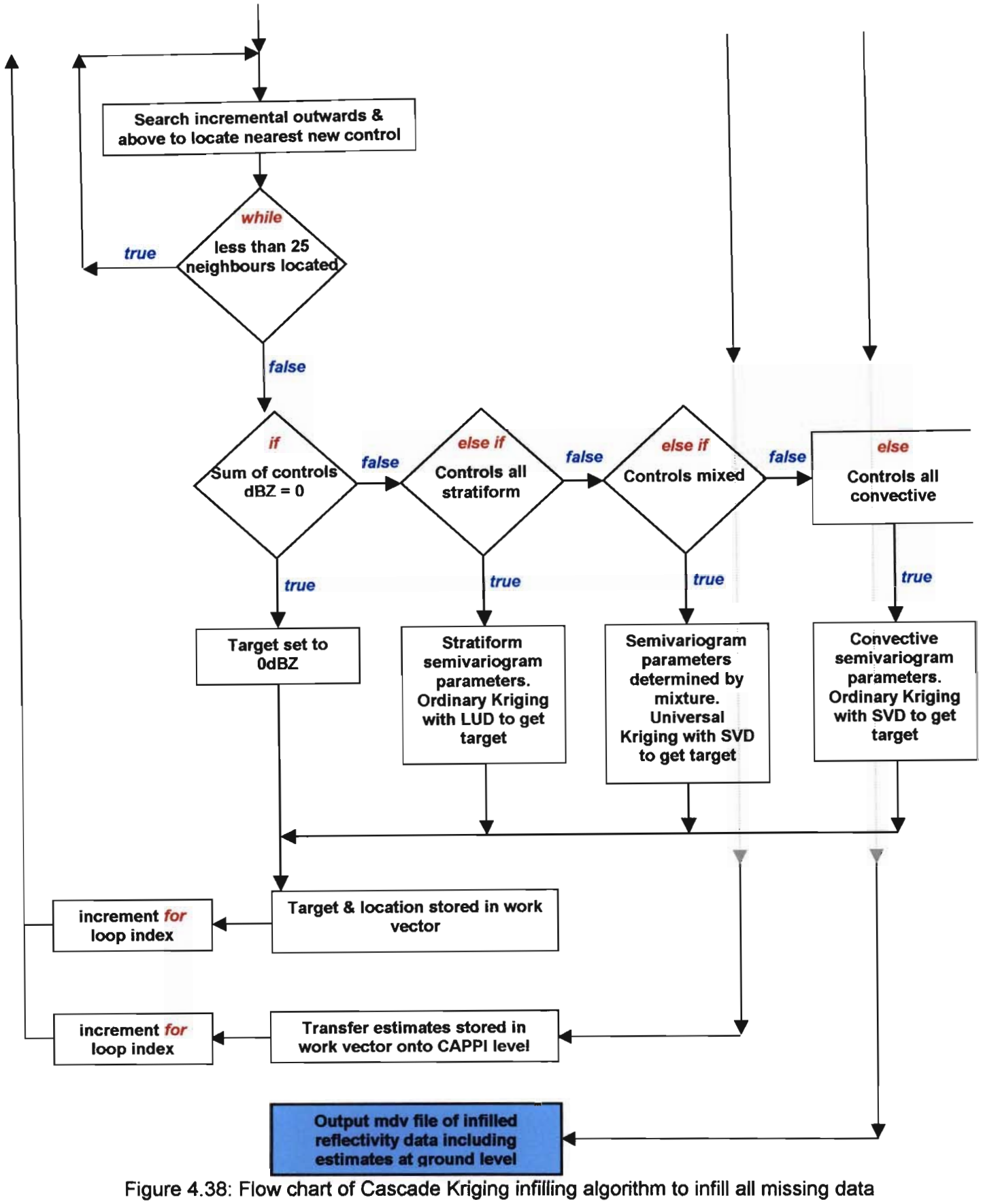


Figure 4.38: Flow chart of Cascade Kriging infilling algorithm to infill all missing data contained in a radar volume scan and provide an estimate of rainfall at ground level.

4.6 COMPARISON AT GROUND LEVEL BETWEEN RAINAGUGE AND RADAR ESTIMATES

As stated in Wesson and Pegram (2005), Section 4.7 (page 113) the Liebenbergsvlei catchment was chosen as the test catchment due to its close proximity to the MRL-5 weather radar and the catchment containing 45 telemetering raingauges. In order to determine an effective means of comparing, on the one hand a 1 km by 1 km radar pixel estimate to, on the other hand, a point raingauge estimate, the following approach was taken. The average value of the 9 radar pixels (a 3 km square area), with the centre square covering a gauge, was taken as the radar estimate of instantaneous mean area rainfall at each gauge location. The raingauge accumulation for that same 9 km^2 area was also computed by means of Block Kriging so that two area estimates are compared instead of comparing the rain on an area with that at a point. A morphing algorithm was utilised in order to determine the radar rainfall accumulations at ground level that takes into account the motion of the rainfield between radar volume scans (Sinclair and Pegram, 2003). The Block Kriging computation methodology will now be outlined.

Universal and Ordinary Kriging both provide an estimation of unknown data at a specific point. In some instances, however, the estimation over an area is required. This can be determined by the use of Block Kriging. Equation (4.14) is used to determine the estimate the average value over a specified area.

$$\hat{z}_D = \frac{1}{D}\int_D Z(x)\,dx \tag{4.14}$$

Equation (4.15) is used to determine the magnitude of the weighting values, $\lambda(s_0)$, where the G matrix would be unchanged. The vector of semivariogram values (g) between the target area and the controls is now a vector of integrals, as illustrated in Equation (4.15):

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda(s_0) \\ \mu(s_0) \end{bmatrix} = \begin{bmatrix} 1/D \cdot \int_D g(s_0) \\ 1 \end{bmatrix} \tag{4.15}$$

Where $\int g(s_0)$ is the vector of the integrals of the semivariogram from the control to the points over the entire target area; the integral can be numerically computed as shown by Equation (4.16):

$$\frac{1}{D}\int g(s_0) = \frac{1}{D}\sum \Delta A \cdot g_{ij} \qquad (4.16)$$

where D refers to the area of the total region for which an estimate is being determined, ΔA is the area of an elemental area in region D and g_{ij} is the semivariogram value from the control (i) to the centre of an individual elemental area (j) (Bras and Rodriguez-Iturbe, 1985: 402-404). A simple illustration of a scenario where Block Kriging is used to provide an estimate over an area is indicated in Figure 4.39.

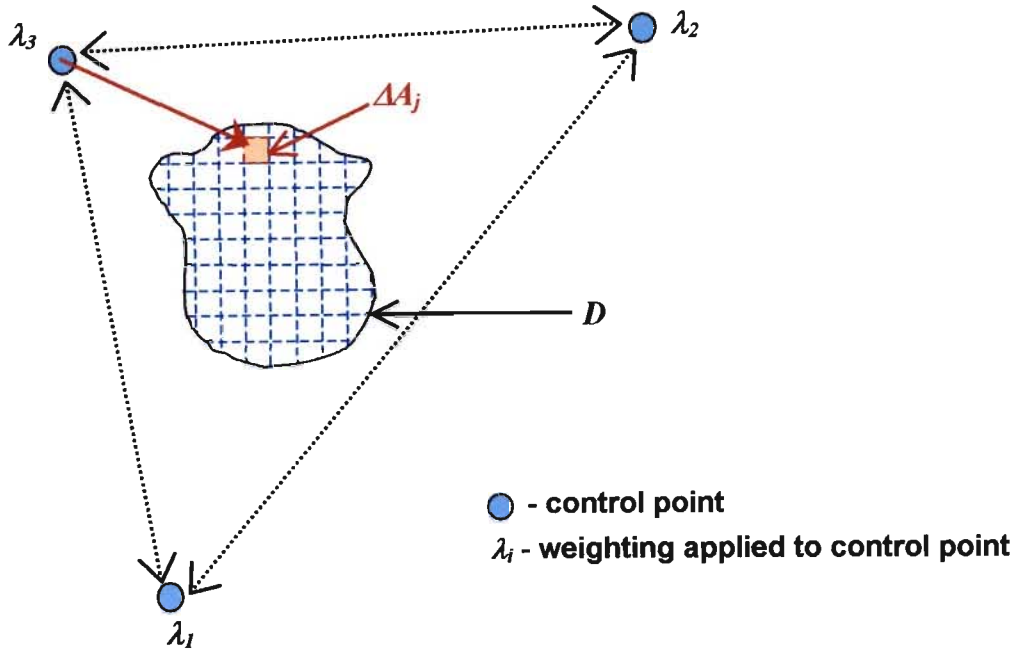


Figure 4.39: Simple illustration of Block Kriging scenario with three controls and an area average to be determined.

In order to determine the appropriate element size ΔA to subdivide the 9 pixel (9 km^2) region into, the following test was undertaken. A point estimate of 25 mm representing a raingauge was placed at the centre of a 3 by 3 pixel region and the Block Kriging estimate computed over that region for decreasing ΔA and for two semivariogram functions, one for stratiform and one for convective rainfall. As indicated in Figure 4.40 the Block Kriged estimate does not alter significantly once the 3 km^2 area has been subdivided into 50 elements, corresponding to a ΔA of size 60 m square.

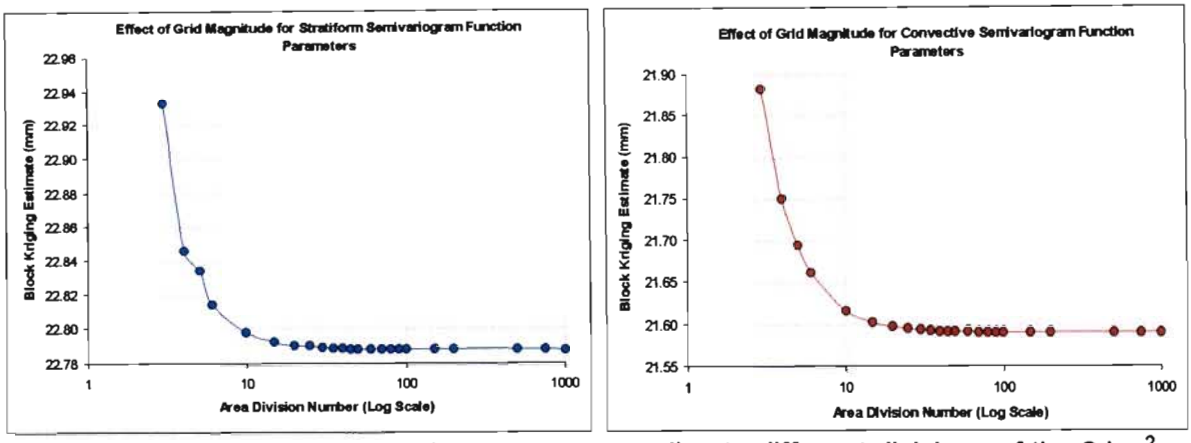


Figure 4.40: Block Kriging estimates corresponding to different divisions of the 9 km^2 region, ranging from 3 divisions (1000 m by 1000 m regions) into 1000 divisions (3 m by 3 m regions) for two different semivariograms: stratiform (left) and convective (right).

When computing the Block Kriged estimate a nearest neighbours approach was adopted where the raingauges situated within 2 correlation lengths of the gauge were used to compute the estimate. A comparison of the Block Kriging rangauge estimates and point raingauge accumulation estimates is indicated in Figure 4.41 where the 24 hour raingauge accumulations for the rain event from the 24 January 1996 from the Liebenbergsvlei raingauge network consisting of 45 raingauges are illustrated.

It is interesting to note the very small differences between the point and Block Kriged estimates, as illustrated in Figure 4.41. The difference in the middle range are mixed in sign, while at the extremes the large gauge regions are diminished by Block Kriging and the zero gauge readings are associated with small positive Block Kriged values as expected from an interpolation scheme. In spite of the small differences, this validates the use of Block Kriging.

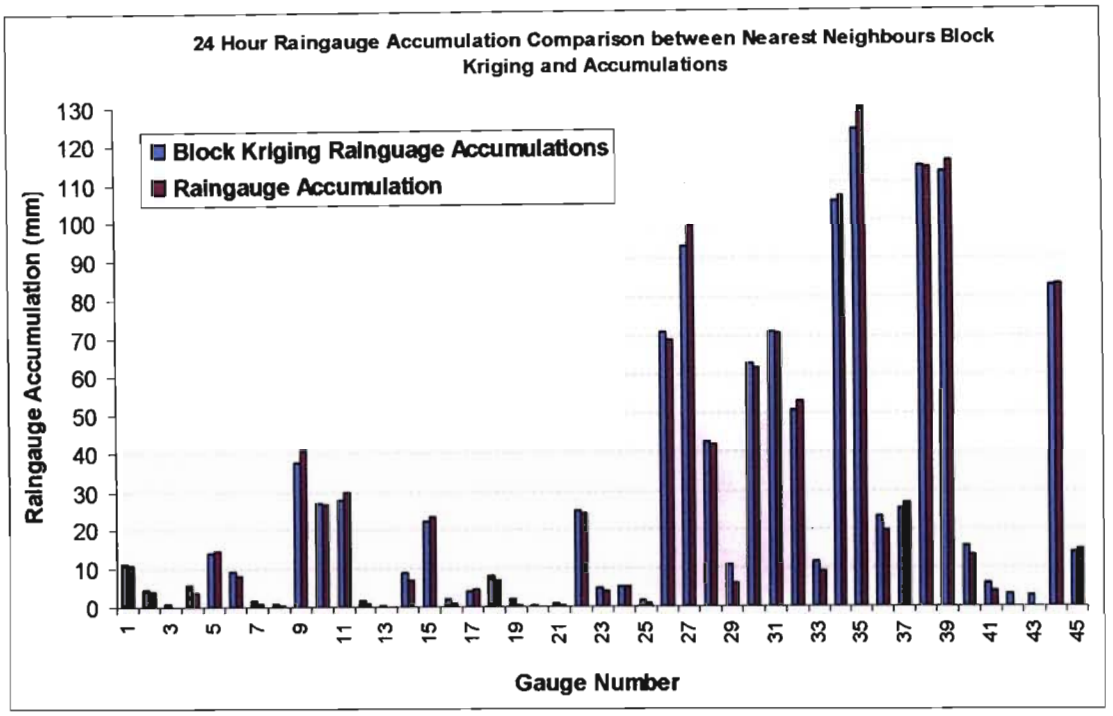


Figure 4.41: Comparison between the point raingauge estimated and the Nearest Neighbours Block Kriged estimates for a 24 hour period is indicated.

The following two rain events were selected for the validation tests: (1) 24 January 1996 - a rain event which consisted of extremely high rainfall intensities over an approximately 12 hour period, (2) 13 February 1996 - a rain event consisting of a combination of convective rain with periods of stratiform rain were selected for validation tests in the paper contained in Section 4.7. A more detailed analysis and results overview is contained in Chapter 5.

4.7 JOURNAL PAPER – *IMPROVED RAINFALL ESTIMATION AT GROUND LEVEL*

Aspects of this work detailed in this chapter were presented at the European Geosciences Union General Assembly in Vienna, Austria in May 2005. Upon invitation a paper was written and submitted in September 2005 for publication in a special edition of Natural Hazards and Earth System Sciences (NHESS), the paper has been reviewed and accepted for publication and is contained in this section.

Improved Radar Rainfall Estimation at Ground Level

Authors: Stephen M. Wesson and Geoffrey G.S. Pegram
Civil Engineering, University of KwaZulu-Natal, Howard College Campus, Durban.
Email: wessons@ukzn.ac.za and pegram@ukzn.ac.za

ABSTRACT

A technique has been developed to provide an estimate of the rainfall reaching the earth's surface by extrapolating radar data contained aloft to ground level, simultaneously estimating unknown data in the radar volume scan. The technique has been developed so as to be computationally fast, to work in real time and comprises the following steps. A rainfall classification algorithm is applied to separate the rainfall into two separate types: convective and stratiform rainfall. Climatological semivariograms based on the rainfall type are then defined, which result in a fast and effective means of determining the semivariogram parameters anywhere in the radar volume scan. Then, extrapolations to ground level are computed by utilising 3D Universal and Ordinary Cascade Kriging; computational efficiency and stability in Kriging are ensured by using a nearest neighbours approach and a Singular Value Decomposition (SVD) matrix rank reduction technique. To validate the proposed technique, a statistical comparison between the temporally accumulated radar estimates and the Block Kriged raingauge estimates is carried out over matching areas, for selected rainfall events, to determine the quality of the rainfall estimates at ground level.

Keywords: ground level rainfall estimation, rainfall classification, climatological semivariograms, 3D Universal and Ordinary Cascade Kriging, radar and raingauge accumulations, statistical comparison.

INTRODUCTION

Raingauges are the traditional tool used for the recording of rainfall and are often regarded as the "true", or reference, rainfall estimates at ground level. Raingauges have the advantage of being relatively cheap and easy to maintain and also of providing a direct estimate of the accumulated rainfall at a particular point. However there are various disadvantages associated with raingauges. They tend to underestimate during heavy rainfall periods (Wilson and Brandes, 1979) and by providing a point estimate they can fail to capture the spatial variability of rainfall. Random errors are also present which can be

demonstrated by the difference in measurement obtained between closely situated raingauges (Ciach, 2002).

Weather radars overcome some of the disadvantages associated with raingauges. Currently in South Africa the weather radar network of eleven C-band radars and one S-band radar provides images of the instantaneous reflectivity values at approximately five-minute intervals and at a one-kilometre horizontal resolution supplied in Cartesian co-ordinates on Constant Altitude Plan Position Indicators (CAPPIs) at 1 km intervals above ground level. This type of data provides a detailed spatial representation of the rainfield in real time and over a large area. Figure 1 illustrates a typical instantaneous radar reflectivity image taken from the Bethlehem (South Africa) S-band weather radar where the white portions indicate no rainfall, grey portions indicate areas out of the weather radar's range (and also where there are no data available), and the black portions mark ground clutter locations.

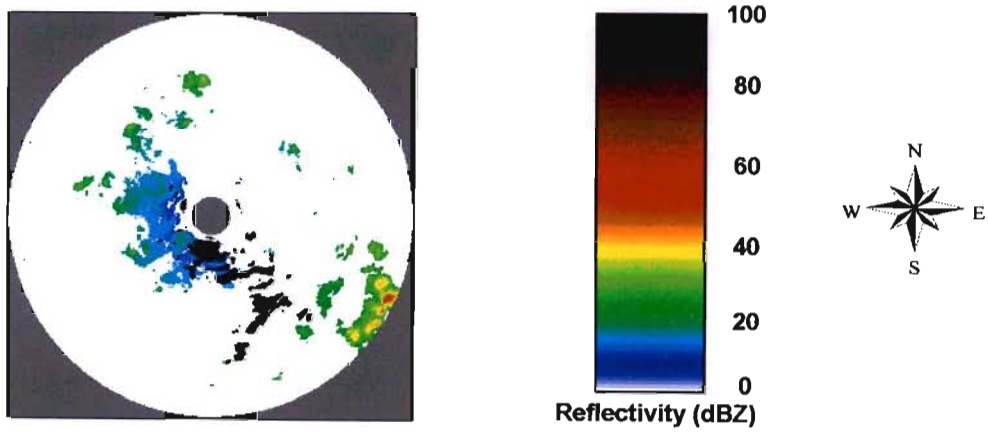


Figure 1: Typical radar reflectivity image taken from the Bethlehem (South Africa) weather radar on the 25 February 2003. The grey portions of the image indicate where no data are available and the black portions indicate ground clutter. The image is 300 km square.

Rainfields estimated from weather radars experience various data quality problems such as ground clutter, anomalous propagation and beam blocking, to name a few (Terblanche *et al*, 2001). Another disadvantage of weather radars is that they provide an indirect measurement of precipitation intensity, so the returned power of measured reflectivity values has be converted to rainrate by an appropriate transformation, such as the Marshall Palmer formula (Marshall and Palmer, 1948) given by Eqn. (1), where Z is the reflectivity (dBZ) and R is the rainrate (mm/hr).

$$Z = 200 \cdot R^{1.6} \quad (1)$$

The data in the CAPPIs are only available from one-kilometre above ground level and there are regions within the volume scan where the rainrate is unknown. In applications, such as disaster management, hydrology and agriculture, the rainfall estimates at ground level are of more interest and importance than the measurements aloft, which are unlikely to be an accurate indication of the rainfall at ground level (Jordan *et al*, 2000) since precipitation tends to be affected by a variety of atmospheric phenomena before reaching the earth's surface.

Attempts have been made to provide an improved rainfall estimate at ground level by taking into account atmospheric factors such as wind drift (Lack and Fox, 2004). A different approach, taken by Franco *et al* (2002), was to compute a Mean Apparent Vertical Profile of Reflectivity and by extending the profile shape obtain an estimate of the rainfall at ground level. Currently in South Africa the real-time radar rainfall estimates at ground level are simply taken as the average vertical reflectivity in a column (Visser, 2003).

Seed and Pegram (2001) proposed using 3D Kriging to estimate the rainfall at ground level and that paper is the source of some of the ideas for this paper; which consists of the following sections. Firstly the details of the rainfall classification algorithm are outlined and the characteristics of the classified rainfall described. A bright band correction algorithm based on local gradients is then introduced and tested. The next part describes the empirical computation of the robust semivariogram from radar reflectivity data and outlines the use and validation of climatalogical semivariograms based on rainfall type. Kriging theory is then briefly discussed, where the methods designed to ensure computational efficiency and stability are outlined. 3D Universal and Ordinary Kriging follow in their application to Cascade Kriging, a novel idea for extrapolation. Finally a statistical comparison between the temporally accumulated radar estimates and the Block Kriged raingauge estimates is carried out over matching areas, for selected rainfall events, to determine the quality of the rainfall estimates at ground level.

The data from 1996 from the Liebenbergsvlei catchment near Bethlehem were selected as the test-set for this study. This was because there was a dense network of 45 tipping

bucket raingauges placed in a grid, evenly spaced at approximately 10 km throughout the catchment at that time (since decommissioned). The catchment is approximately 3600 km^2 in area, which gives a raingauge density of 1 raingauge per 80 km^2. The topography is one where the elevation is approximately 2400 m above sea level in the Sourthern reaches of the catchment and slopes downwards to the Northern region of the catchment to an elevation of approximately 1500 m. The Bethelehem S-band radar is also situated in close proximity to the catchment. Both of these factors are important in ensuring accurate estimates of rainfall accumulation values. The geographic position of the area and layout of the tipping bucket raingauges are illustrated in Figure 2 as well as a Digital Elevation Map depicting the topography of the catchment.

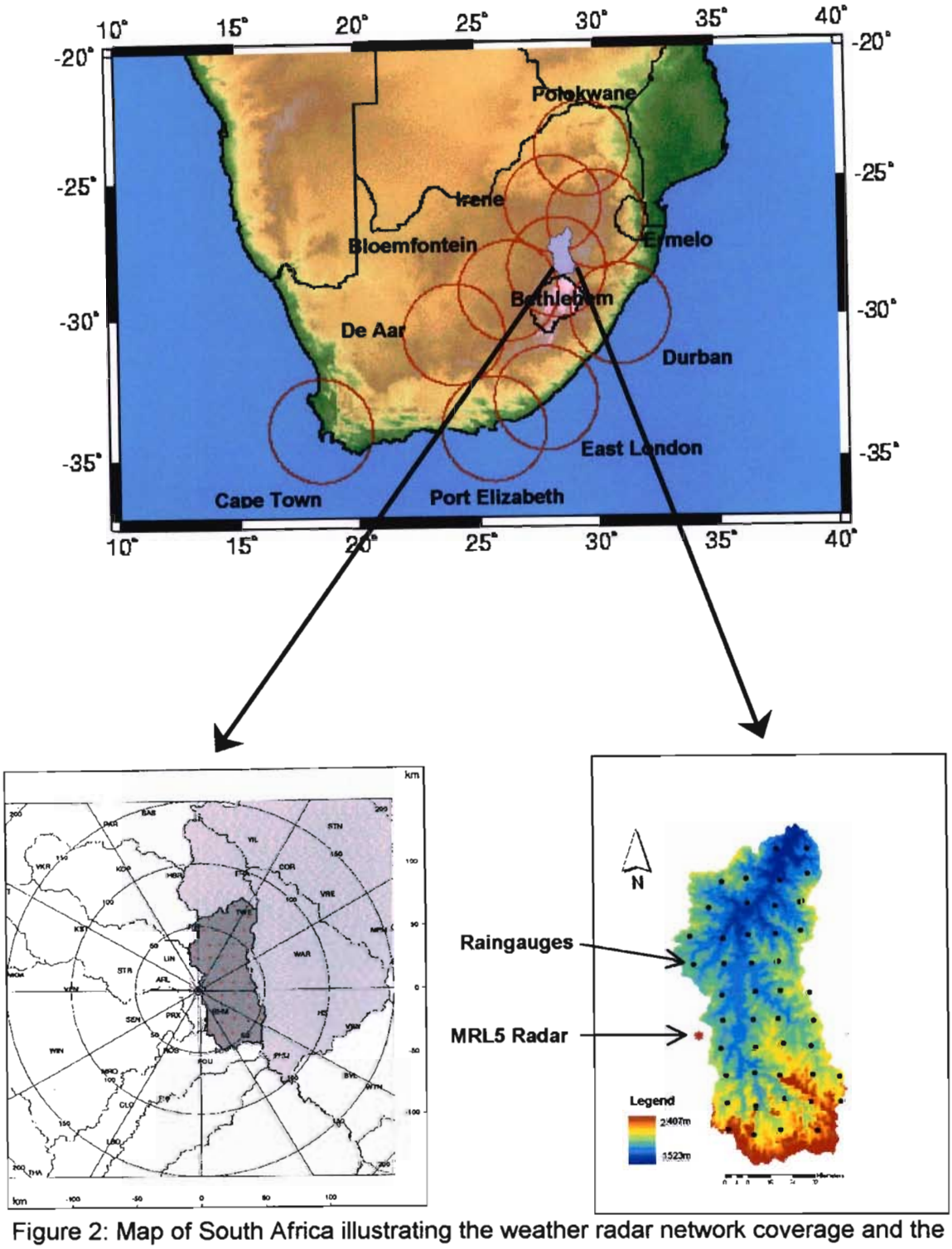


Figure 2: Map of South Africa illustrating the weather radar network coverage and the geographic location of the Liebenbergsvlei catchment. An enlarged image of the catchment shows the approximate positions of the raingauges and the MRL5 weather radar.

RAINFALL CLASSIFICATION

The initial step, of the operational algorithm for infilling the missing or inaccurate data on a CAPPI, requires a simple classification and thresholding of the instantaneous radar volume scan data, pixel by pixel. A variety of rainfall classification algorithms have been suggested that involve subdividing the rainfall into three zones – convective, intermediate and stratiform rain. An example of a current rainfall classification algorithm employed by The Group of Applied Research on Hydrometeorology (GRAHI) (Sempere-Torres *et al*, 2000) also subdivides the rainfield into the same three separate groupings; the convective rain is identified by applying a threshold value of 40 dBZ; the stratiform rain is identified when a bright band is present; the rest of the rainfield classified as transitional. Another rainfall classification algorithm developed by Mittermaier (1999) was also designed along similar principles.

The above-mentioned algorithms are not suitable for the purpose of rainfall estimation at ground level over large areas. This is due to the fact that the data at the 2 km CAPPI (which with a base scan at 1.5° above horizontal has a range limited to 73 km) plays a pivotal role in classifying the rainfall as stratiform or intermediate. This means that only a small proportion of the total volume scan data can be classified. Another drawback of the existing algorithms in real time applications is that they are computationally expensive, especially for large data sets.

The algorithm developed here, with assistance of meteorologists from the South African Weather Services Innovation and Research Division, classifies the rainfall into only two groups: convective and stratiform and works as follows. The classification is rather like Steiner's (1995) algorithm and is done pixel by pixel throughout the volume scan data. The initial step of the algorithm works on a simple threshold where all reflectivity values less than or equal to 18 dBZ are set to zero due to the fact that 18 dBZ approximates to 0.50 mm/hr, a rate that can be considered negligible. The remainder of the reflectivity values above zero are then classified by the simple criterion: if the reflectivity value is 35 dBZ or above, the rainfall is classified as convective; the reflectivity values in the range 18 dBZ to 35 dBZ are classified as stratiform rainfall.

A summary of the classification criteria is given in Table 1 and two examples of classified instantaneous images both 4 km above ground level are illustrated in Figure 3. Two

examples of vertical cross sections through the volume scan data are illustrated in Figure 4.

Table 1: Rainfall classification and threshold criteria to separate radar volume scan data into stratiform and convective rainfall domains.

CRITERIA	CLASSIFICATION
Pixel ≤ 18 dBZ	No Rainfall: Pixel set to 0 dBZ
18 dBZ < Pixel < 35 dBZ	Stratiform Rain
Pixel ≥ 35 dBZ	Convective Rain

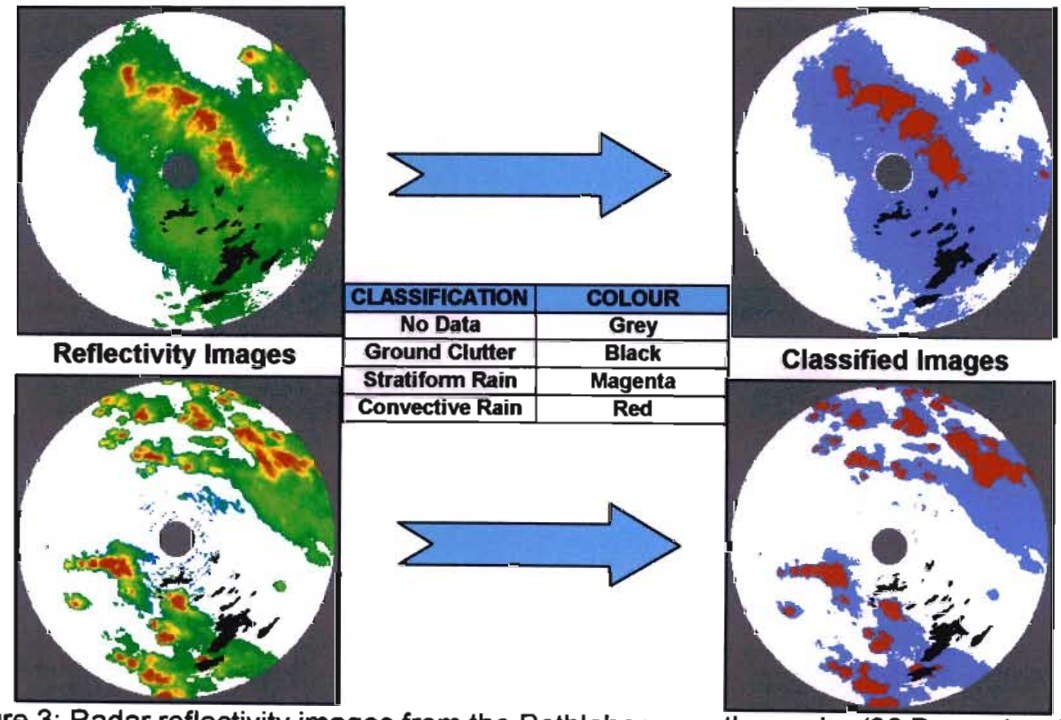


Figure 3: Radar reflectivity images from the Bethlehem weather radar (30 December 2001 and 8 January 2002) and the corresponding classified images. All images are 300 km square.

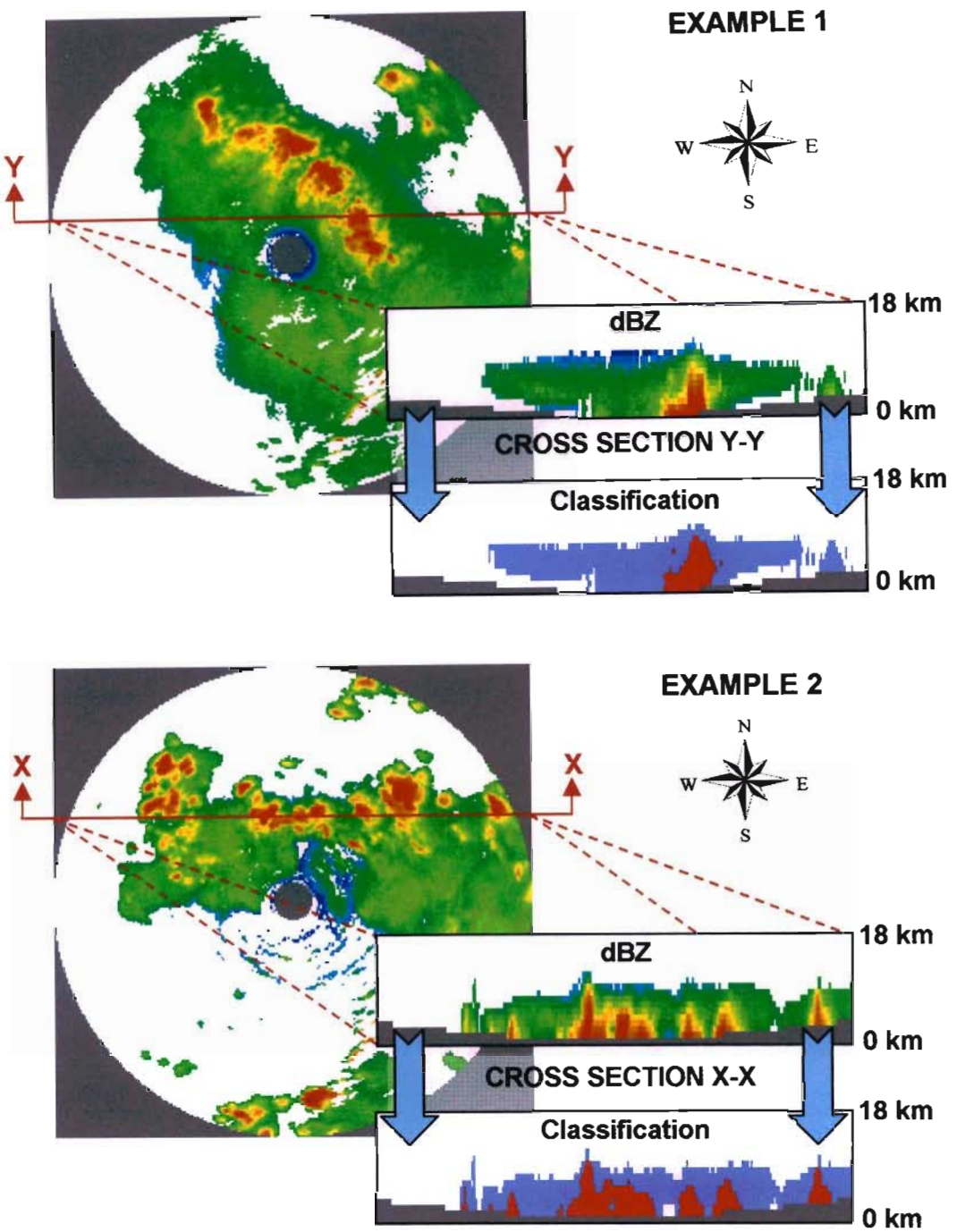


Figure 4: Cross sectional profile through two instantaneous reflectivity images. Indicated in the plan is the CAPPI 4 km above ground level and the position of the vertical cross section in the radar volume scan CAPPI. Example 1 is from the Bethlehem weather radar, 30 December 2001, and Example 2 from the Bethlehem weather radar, 24 January 2002. All images are 300 km across. Note the effectiveness of the classification algorithm.

The classification algorithms proposed by Steiner (1995) and Mittermaier (1999) work on the basis that the classification of convective rainfall on the 4 km CAPPI is applied to all levels below and that the 2 km CAPPI is used to classify stratiform rain. However the rainrates at a large number of pixels outside the range of the 4 km and 2 km CAPPI need to be estimated and large portions of the image on the 4 km and 2 km CAPPI may also be marked as ground clutter and therefore be contaminated. As indicated in Figure 4, where the convective rainfall approaches ground level (below 4 km in altitude) the convective rain zones may increase substantially in the horizontal direction, which would then result in a scenario where rainfall of high intensity (e.g. 54 dBZ ≈ 84 mm/hr) is miss-classified as stratiform rainfall. For the above reasons a pixel by pixel classification approach was chosen throughout the CAPPI volume scan due to its greater suitability for the purpose of rainfall estimation at ground level.

The main advantage of this algorithm is its simplicity, in that for a real time application it has little impact on the overall computation time; this is especially true for large data sets such as those returned from weather radar volume scans. Secondly the algorithm is not restricted to the 2 km CAPPI which is frequently the level of the melting layer in a South African context; thirdly, a far greater spatial range can be classified than that limited to 73 km at the 2 km level.

CHARACTERISITCS OF CLASSIFIED RAINFALL

The rainfall classification algorithm was tested on numerous images ranging over five different years (1995, 1996, 2000, 2001 and 2002) from two weather radars in South Africa: Bethlehem (mixed convective and stratiform rain, 1800 m altitude, sub-tropical savannah) and Durban (mostly orographic warm rain, coastal, bush). The mean and standard deviation of the reflectivity values of selected portions (areas in size of approximately 600 km^2 and greater) of the classified sections, which were respectively dominated by either convective or stratiform precipitation, were then computed and plotted as indicated in Figure 5. As shown, the statistical characteristics of the stratiform and convective rainfall are distinctly different. The stratiform rain tends to have a low and reasonably constant standard deviation of approximately 3 dB and a mean value ranging from 18 dBZ to 29 dBZ. In the convective rainfall the standard deviation showed a far greater variation ranging from 2 dB to 8 dB, exhibiting a strong, steep linear trend where a higher mean value corresponds to higher standard deviation, to be expected from data

sampled from a field which is approximately lognormally distributed (Pegram and Clothier, 2001).

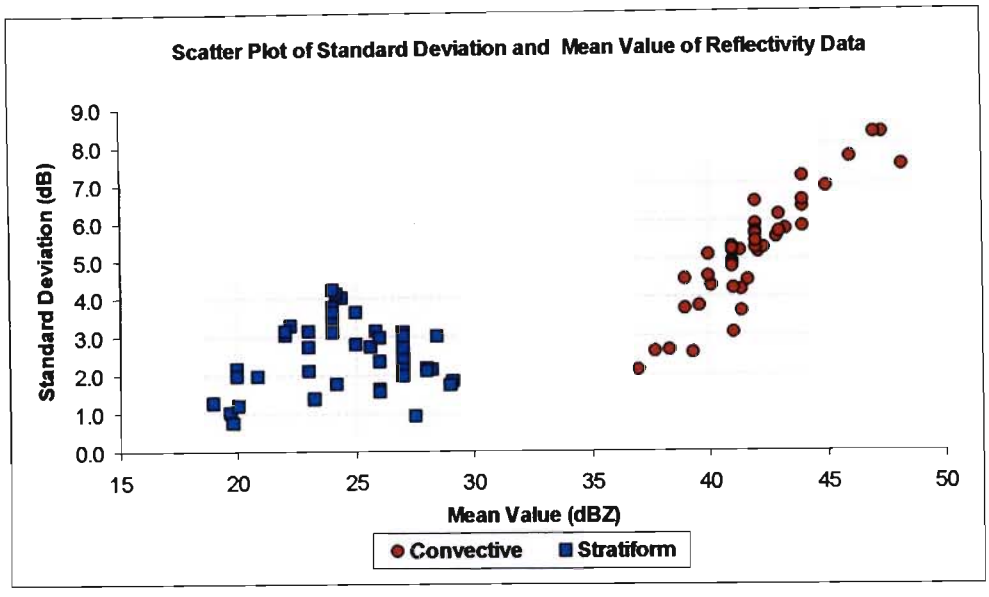


Figure 5: Scatter plot of the mean and standard deviation of rainfall zones classified as convective or stratiform rain indicating the different statistical properties of each rainfall type. The number of 1 km pixels used in defining each point is approximately 600 or greater, from data over a five year period collected from the Bethlehem and Durban weather radars.

The vertical characteristics of the classified stratiform and convective rain were also examined by computing the mean vertical profile of reflectivity from selected images from the Durban and Bethlehem weather radars over the same five years as mentioned above. As shown in Figure 6 the stratiform rain generally has a low average vertical height, limited to 8 km above ground level. The convective rainfall has considerable vertical extent ranging up to 14 km above ground level and a bigger range in mean reflectivity values. Both the stratiform and convective profiles show a constant increase in rainfall intensity as ground level is approached, with the exception that in the stratiform profile, evidence of a bright band at the 2 km level indicates that the stratiform rain has been classified correctly. The variability of the reflectivity in both are characterised by a standard deviation of 3.5 dB or less at any level.

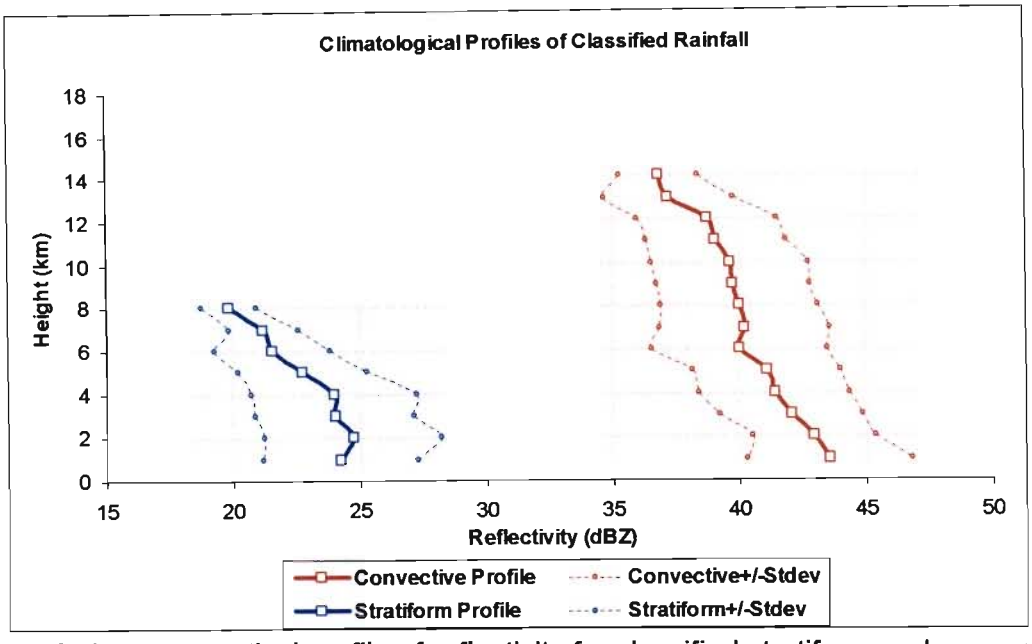


Figure 6: Average vertical profile of reflectivity for classified stratiform and convective rainfall. Profiles are computed from 20 different images ranging over a 5 year period from the Bethlehem and Durban weather radars.

BRIGHT BAND CORRECTION

As snow and ice crystals drop through the 0°C isotherm they begin to melt and are surrounded by a thin layer of water. The result is that, from the radar's point of view, the melting snow and ice crystals resemble large blobs of water and the reflectivity values for the melting level are enhanced greatly, resulting in an overestimation of rain (Sanchez-Diezma *et al*, 2000). This level is called the bright band. At Bethlehem, South Africa the bright band generally occurs 2 km above ground level (Mittermaier, 1999) as can be seen by the stratiform climatological profile of reflectivity in Figure 6. This bright band effect results in a vertical profile of reflectivity that resembles Figure 7.

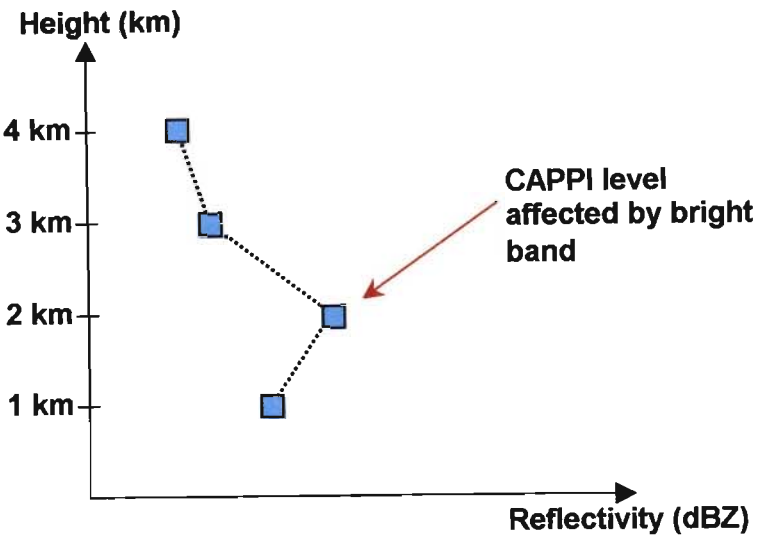


Figure 7: Idealised stratiform climatological profile of reflectivity with bright band affecting the reflectivity values situated on the 2 km CAPPI, indicated by the sharp increase in reflectivity at the 2 km level.

Predominantly two different approaches have been proposed, as described by Mittermaier (2003), in order to correct the problem of bright band: (1) A vertical profile is derived based on radar volume scan data and is used to correct the values and provide an estimate of the rainfall rate at ground level (e.g. Sanchez-Diezma *et al*, 2000); (2) A standard high resolution profile is used for correction with the height of the top of the bright band set by the height of the 0^0C wet bulb temperature derived from a model.

EFFECT OF BRIGHT BAND ON GROUND LEVEL ESTIMATES

The presence of a bright band has an adverse effect on the estimates at ground level when applying an extrapolation algorithm to volume scan radar reflectivity data. The reflectivity values at the 2 km CAPPI, which are on average higher than the values contained on the 1 km CAPPI, create a trend of decreasing reflectivity values as ground level is approached. When extrapolating to ground level and pixels are selected from the 1 km and 2 km CAPPIs, the trend of decreasing reflectivity is extrapolated to ground level resulting in lower estimates of stratiform rainfall at ground level than should be expected. There is therefore a need to adjust the reflectivity values at the 2 km CAPPI to provide an improved estimate of the rainfall at ground level.

BRIGHT BAND CORRECTION ALGORITHM

The first step of the algorithm is to compute the wetted area ratio (WAR) of stratiform rainfall at the 3 km, 2 km and 1 km CAPPIs. If the WAR of stratiform rainfall is less than 10% at any of the CAPPIs, then no correction takes place due to there being insufficient rainfall to determine if a bright band is present; if the WAR of stratiform rainfall is greater than 10% the 2 km CAPPI is then examined to detect evidence of a bright band.

The existence of the bright band is determined by comparing the temporally weighted mean values of reflectivity ($\widetilde{x}_1$, $\widetilde{x}_2$ and $\widetilde{x}_3$) at the 1 km, 2 km and 3 km CAPPIs; the calculation is limited to the reflectivity values classified as stratiform rainfall on the image. If the computed temporally weighted mean values of the stratiform rainfall at each level are $\widetilde{x}_1 > \widetilde{x}_3$ and $\widetilde{x}_2 \geq \widetilde{x}_1$ then a bright band is considered to be present. Some memory of the mean value from previous time steps is retained by temporal weighting as indicated by Eqn. (2):

$$\widetilde{x}_T = \lambda \cdot \bar{x}_T + (1-\lambda) \cdot \widetilde{x}_{T-1} \qquad (\lambda = 0.50) \qquad (2)$$

where $\widetilde{x}_T$ is the temporally weighted mean, λ represents the weighting value which is set to 0.50 (a value chosen to smooth the means, but not excessively), and $\bar{x}_T$ is the spatial mean value of reflectivity at a give level computed from the stratiform data at time *T*.

Throughout the rain event a second temporally weighted mean is also computed, as indicated by Eqn. (3), where the weighting value δ is set to 0.05; this has the result of producing a slowly adapting mean value w_T. If the WAR of the stratiform rainfall is below 10% then a break occurs and the $\widetilde{x}_T$ value is now computed by Eqn. (4) where the $\widetilde{x}_T$ is based on the final w_T value computed before the break which had the value w_{TB}:

$$w_T = \delta \cdot \bar{x}_T + (1-\delta) \cdot w_{T-1} \qquad (\delta = 0.05) \qquad (3)$$

$$\widetilde{x}_T = \delta \cdot w_{TB} + (1-\delta) \cdot \widetilde{x}_{T-1} \qquad (\delta = 0.05) \qquad (4)$$

an example of this is illustrated in Figure 8 where Eqn. (2), (3) and (4) were used to compute the mean values of reflectivity ($\widetilde{x}_T$) by geometric weighting. The reflectivity data are from the Bethlehem weather radar, 25 January 1996. As indicated in Figure 8, in three periods, the WAR of stratiform rainfall is less than 10%; in these periods $\widetilde{x}_T$ is computed from Eqn. (4), not (2), so that toward the end of the period, it tends to w_{TB}, which can be thought of as a locally adaptive historical mean value of reflectivity.

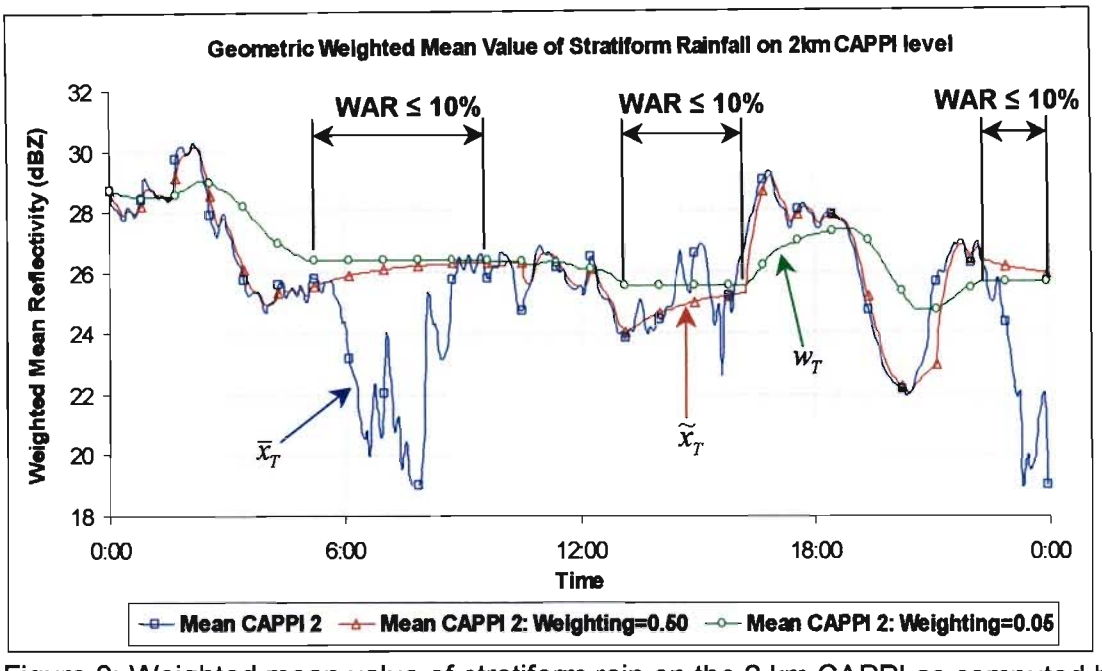


Figure 8: Weighted mean value of stratiform rain on the 2 km CAPPI as computed by geometric weighting, Eqns. (2), (3) and (4). The data was taken from the Bethlehem weather radar (25 January 1996) over a 24-hour period.

To correct the bright band it is also necessary to compute the standard deviation (σ) of the stratiform rainfall on the 1 km CAPPI. Some memory of the σ value from the previous time step is retained by weighting as in Eqn. (2); to compute the weighted standard deviation, Eqn. (5) is used.

$$\widetilde{\sigma}_T = \lambda \cdot \sigma_T + (1-\lambda) \cdot \widetilde{\sigma}_{T-1} \qquad (\lambda = 0.50) \qquad (5)$$

where $\widetilde{\sigma}_T$ is the weighted standard deviation, λ represents the weighting value which is set to 0.50 and σ_T is the mean value of reflectivity at time *T*. Once again if the WAR of

the stratiform rainfall is less than 10% a break occurs and the $\widetilde{\sigma}_T$ value is now computed by Eqn. (7) where the $\widetilde{\sigma}_T$ is based on final s_T value computed by Eqn. (6) before the break which has the value s_{TB}:

$$s_T = \delta \cdot \sigma_T + (1-\delta) \cdot \sigma_{T-1} \qquad (\delta = 0.05) \qquad (6)$$

$$\widetilde{\sigma}_T = \delta \cdot s_{TB} + (1-\delta) \cdot \sigma_{T-1} \qquad (\delta = 0.05) \qquad (7)$$

If a bright band is identified, all the values classified as stratiform rainfall are then evaluated on an individual pixel by pixel basis within a vertical column, to decide if the pixel at the 2 km CAPPI is affected by the bright band. For each stratiform pixel location on the 2 km CAPPI the gradient ($\Delta_{critical}$ = *ΔdBZ/Δheight*) is calculated between the pixel value (directly above it) on the 3 km CAPPI and the value equal to $\widetilde{x}_1 + \widetilde{\sigma}_1$ on the 1 km CAPPI, as illustrated in Figure 9; the gradient is computed from Eqn. (8). The gradient ($\Delta_{observed}$) is also calculated between the pixel on the 3 km and 2 km CAPPI; as indicated by Eqn (9):

$$\Delta_{critical} = \frac{\Delta dBZ}{\Delta Height} = \frac{dBZ_3 - (\widetilde{x}_1 + \widetilde{\sigma}_1)}{2} \qquad (8)$$

$$\Delta_{observed} = \frac{\Delta dBZ}{\Delta Height} = \frac{dBZ_3 - dBZ_2}{1} \qquad (9)$$

If $|\Delta_{observed}| > |\Delta_{critical}|$ then the pixel on the 2 km CAPPI is classified as a bright band pixel and adjusted by altering the value of that pixel as shown in Eqn. (10):

$$New\ dBZ_2\ Pixel = \frac{dBZ_3 + \widetilde{x}_1}{2} \qquad (10)$$

Figure 9 provides a schematic of a scenario where a pixel on the 2 km CAPPI is identified as being affected by bright band. The correction of the pixel is then also indicated as the new value is set to the midpoint of the pixel on the 3 km CAPPI and the mean value ($\widetilde{x}_1$) of the reflectivity data on the 1 km CAPPI.

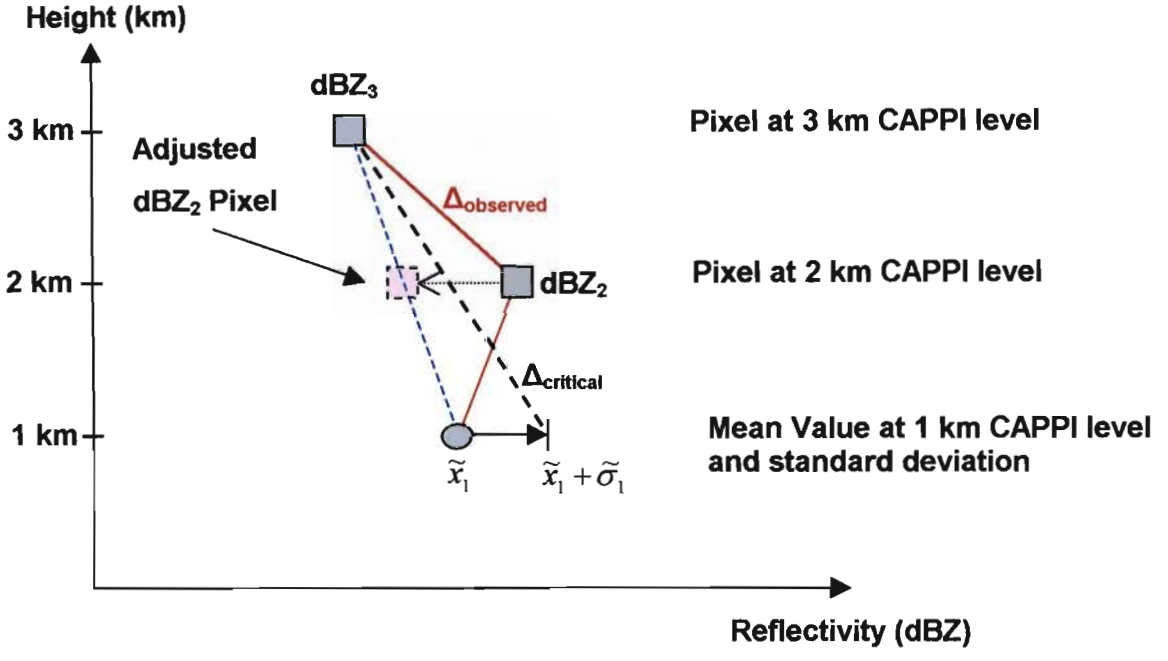


Figure 9: Adjustment of a stratiform pixel value on the 2 km CAPPI that is identified as being affected by bright band. The bright band pixel is adjusted so as to sit at the midpoint of the reflectivity value situated on the 3 km CAPPI and the mean value of the 1 km CAPPI.

TESTING THE BRIGHT BAND CORRECTION PROCEDURE

To test the effectiveness of the proposed bright band correction procedure a period of rainfall from the Bethlehem weather radar was selected (17 December 1995) that consisted predominantly of stratiform rainfall and showed clear evidence of a bright band. Figure 10 illustrates the computed mean values $\tilde{x}_T$ at CAPPIs 1, 2 and 3 throughout the day. The mean value of 2 km CAPPI (green line with triangle markers) is consistently higher throughout the day than the mean values on the 1 km and 3 km CAPPIs, indicating a definite bright band.

The bright band correction procedure described in the previous sub-section was applied to each 5-minute instantaneous image throughout the 24-hour period. The mean value of the adjusted reflectivity data on the 2 km CAPPI was then recomputed. Figure 10 illustrates the adjusted mean value for the 2 km CAPPI, which now lies at the approximate midpoint of the mean value for the 1 km and 3 km CAPPI.

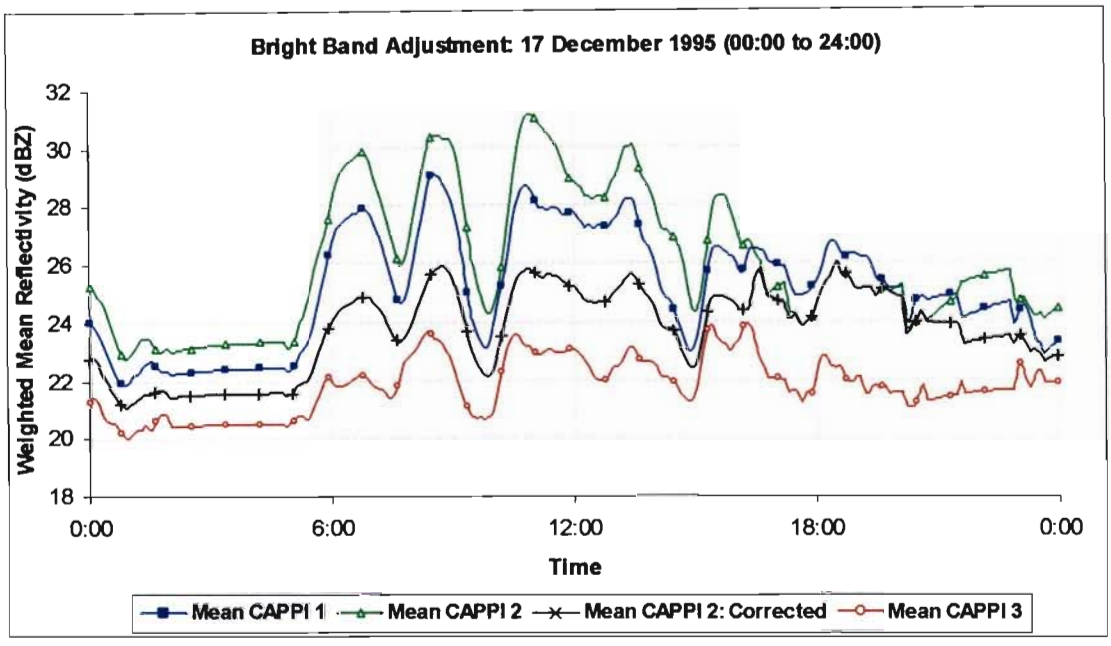
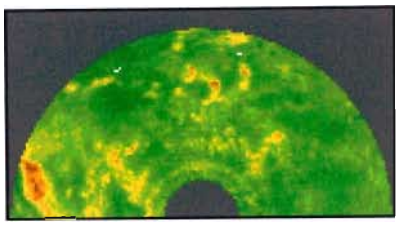
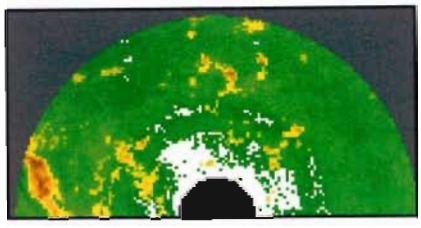


Figure 10: Bright band adjustment of the 2 km CAPPI for the 17 December 1995 (00:00 to 24:00). The green line with triangle markers indicates the 2 km CAPPI affected by bright band and the black line with plus sign markers indicates the mean value of the corrected data on the 2 km CAPPI.

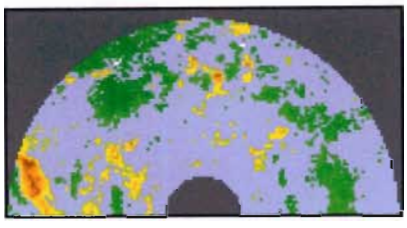
Figure 11 provides an example of an instantaneous reflectivity image that is affected by bright band. The northern half is only shown due to ground clutter contamination in the south of the image. The top image of Figure 11 is contaminated by bright band with the right hand side image indicating marked pixels that are affected by bright band. The bottom image of Figure 11 displays the corrected image.



2 km CAPPI before bright band correction



2 km CAPPI after bright band correction



2 km CAPPI pixels marked which are affected by bright band

Figure 11: An example of bright band adjustment of an instantaneous image for the 2 km CAPPI for the 17 December 1995. The images have the dimensions of width = 150 km and height = 75 km.

To test the effectiveness of the bright band adjustment algorithm, a comparison was carried out between Block Kriged raingauge estimates and the radar accumulation estimates before and after the bright band correction procedure. The raingauge network utilised was from the Liebenbergsvlei catchment as described in the Introduction. The comparison was carried out on a 24-hour rain event from the Bethlehem weather radar from the 17 December 1995 with a strong bright band evident, as indicated in Figure 11. The sum of square of errors (SSE) between the raingauge and radar accumulations for a 24-hour period was computed and the r^2 values also estimated before and after the bright band correction. There was a notable improvement in the radar estimates as the radar accumulation values more closely matched the raingauge estimates after the bright band was corrected. This is demonstrated by the lower SSE and higher r^2 value returned for radar estimates where the bright band correction was implemented as summarised and shown in Table 2.

Table 2: Sum of square of errors (SSE) and r^2 value between radar and raingauge estimates before and after bright band correction.

	Before Bright Band Correction	After Bright Band Correction
SSE	16 601	13 161
r^2	0.34	0.43

The proposed bright band correction procedure provides an effective means of adjusting the 2 km CAPPI reflectivity values when affected by bright band so as to provide an improved estimate of the rainfall at ground level. The bright band correction algorithm is computationally inexpensive and therefore highly suitable for real time applications.

SEMIVARIOGRAM ESTIMATION

One of the advantages of Kriging compared to other interpolation/extrapolation techniques, is that the basis function is determined by the actual data set and takes into account the spatial structure of the data. The basis function in Kriging is determined by the computation and fitting of an appropriate model to the computed semivariogram. The variogram is defined as the expectation of the square of the differences of the field variables separated by a specified distance, as given by Eqn. (11):

$$2 \cdot \gamma(x,h) = E\left\{\left[Z(x) - Z(x+h)\right]^2\right\} \tag{11}$$

where the semivariogram is defined as $\gamma(x, h)$ (Journel and Huigbregts, 1978: 11). The semivariogram provides a way of measuring the spatial dependence that exists amongst the variables in a stationary random field. The Kriging computation can be carried out using either a covariance or a semivariogram function. However in this application the semivariogram is used in preference due to its more robust properties as outlined by Cressie (1993: 70-73).

EMPIRICAL COMPUTATION OF THE SEMIVARIOGRAM

The semivariogram is most commonly computed by the Classical Variogram estimator as was proposed by Matherhon (1962) and is given by Eqn. (12):

$$2 \cdot \hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left[Z(s_i) - Z(s_j)\right]^2 \tag{12}$$

where $\hat{\gamma}(h)$ is the sample semivariogram at lag h (the specified lag distance), $Z(s_i)$ and $Z(s_j)$ are the values of the variables at the specified locations, s_i and s_j which are h apart and $N(h)$ is the number of pairs separated by lag h.

Unfortunately the classical estimator of Eqn. (12) is badly affected by non-typical observations which can be attributed to the $(\cdot)^2$ term in the summand (Cressie, 1993: 40). The effect of non-typical data points can have a dramatic effect on the semivariogram since they are used numerous times in the calculation at different lag intervals. This can result in peaks or shifting of the entire semivariogram upwards (Sabyasachi *et al*, 1997). It was also noted that the distribution of points at each lag distance (h) computed from radar rain fields was highly skewed and did not approximate a normal distribution.

It was thus decided that the Robust Variogram, proposed by Cressie and Hawkins (1980) was the more appropriate model to use and is given by Eqn. (13):

$$2\cdot\bar{\gamma}(h)=\left\{\frac{1}{|N(h)|}\sum_{N(h)}\left|Z(s_i)-Z(s_j)\right|^{1/2}\right\}^4 \Bigg/ \left(0.457+\frac{0.494}{|N(h)|}\right) \qquad (13)$$

By computing the sum of the absolute difference of the pairs of data points, $Z(s_i)$ and $Z(s_j)$ in the square root domain and then raising the result to the power of four dramatically reduces the effect of uncharacteristic observations. It was found that the skewness of the points at each lag (h) was greatly reduced and also that the mean and the median value of the sets of points for each lag interval more closely coincided, both indicating that the distribution of points at each lag (h) more closely approximates a normalized distribution.

An example of the difference between the robust and classical estimator is given in Figure 12. These two semivarigorams were computed from a segment (30 km square in size) of an instantaneous reflectivity image from the Bethlehem weather radar (24 January 2002) in the horizontal direction 4 km above ground level, previously classified as stratiform rainfall. As can be seen, there are significant differences between the two estimates, especially (and importantly) near the origin. Computing the empirical semivariogram using the classical estimator would result in an underestimation of the "actual" correlation length and overestimation of the "actual" alpha parameter described in Eqn. (14) below.

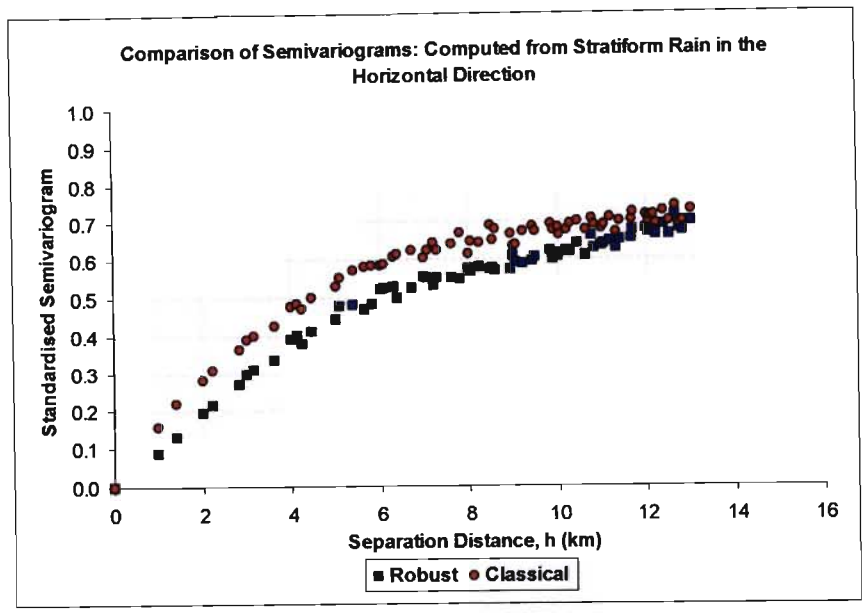


Figure 12: Comparison of robust (Eqn. (4)) and classical (Eqn. (3)) semivariograms computed from a segment (900 km^2) of stratiform rainfall in the horizontal direction. Differences between the two semivariogram estimates are evident, especially at close range (h).

SEMIVARIOGRAM MODEL FITTING

The two-parameter isotropic exponential model was chosen as the model to fit to the empirical semivariogram values, as defined by Eqn. (14):

$$g(h) = 1 - \exp[-(h/L)^{\alpha}] \qquad (14)$$

where h is the Euclidian distance between data points, L the correlation length and α the shape parameter which lies in the range $0<\alpha\leq 2$. As recommended by Journel and Huijbrechts (1978: 194), the model was fitted using up to only half the maximum possible lag and only to the bins at each lag distance that contained thirty or more points. The α and L parameter values were found by minimising an objective function U defined in Eqn. (15) by a weighted least squares approach:

$$U = \sum_{j=1} N(h_j) \cdot \left[\frac{\hat{\gamma}(h_j)}{\gamma(h_j, b)} - 1 \right]^2 \qquad (15)$$

where $N(h_j)$ is the number of points in bin j, $\hat{\gamma}(h_j)$ are the empirical semivariogram values and $\gamma(h_j, b)$ are the model values (Cressie, 1993: 97).

EXAMPLE OF SEMIVARIOGRAM COMPUTATION

An example of computing a semivariogram is illustrated in Figure 13. A 30 km square section of radar reflectivity data classified as stratiform rainfall was selected from the Durban weather radar (11 December 2000). Initially the selected reflectivity values were standardized by use of Eqn. (16):

$$Z = \frac{x - \mu}{\sigma} \tag{16}$$

where Z is the standardised variable, μ the sample mean value of the selected data, x the selected variable and σ the sample standard deviation of the selected data. Standardising the field has the effect of producing a scaled and dimensionless field to ease the fitting of the climatological semivariograms. If the field were not to be scaled by σ, this factor would in any case cancel from both sides of the Kriging matrix equations, Eqn. (21). As can be seen in Figure 13 the standardized semivariogram increases smoothly with lag h until the sill is reached, whose value is equal to 1.0, the variance of the selected data. The range is also indicated, which is the distance beyond which the variables are considered to be uncorrelated. The correlation length in this case is L = 8.31 km, the shape parameter is α = 1.38 and the nugget is zero.

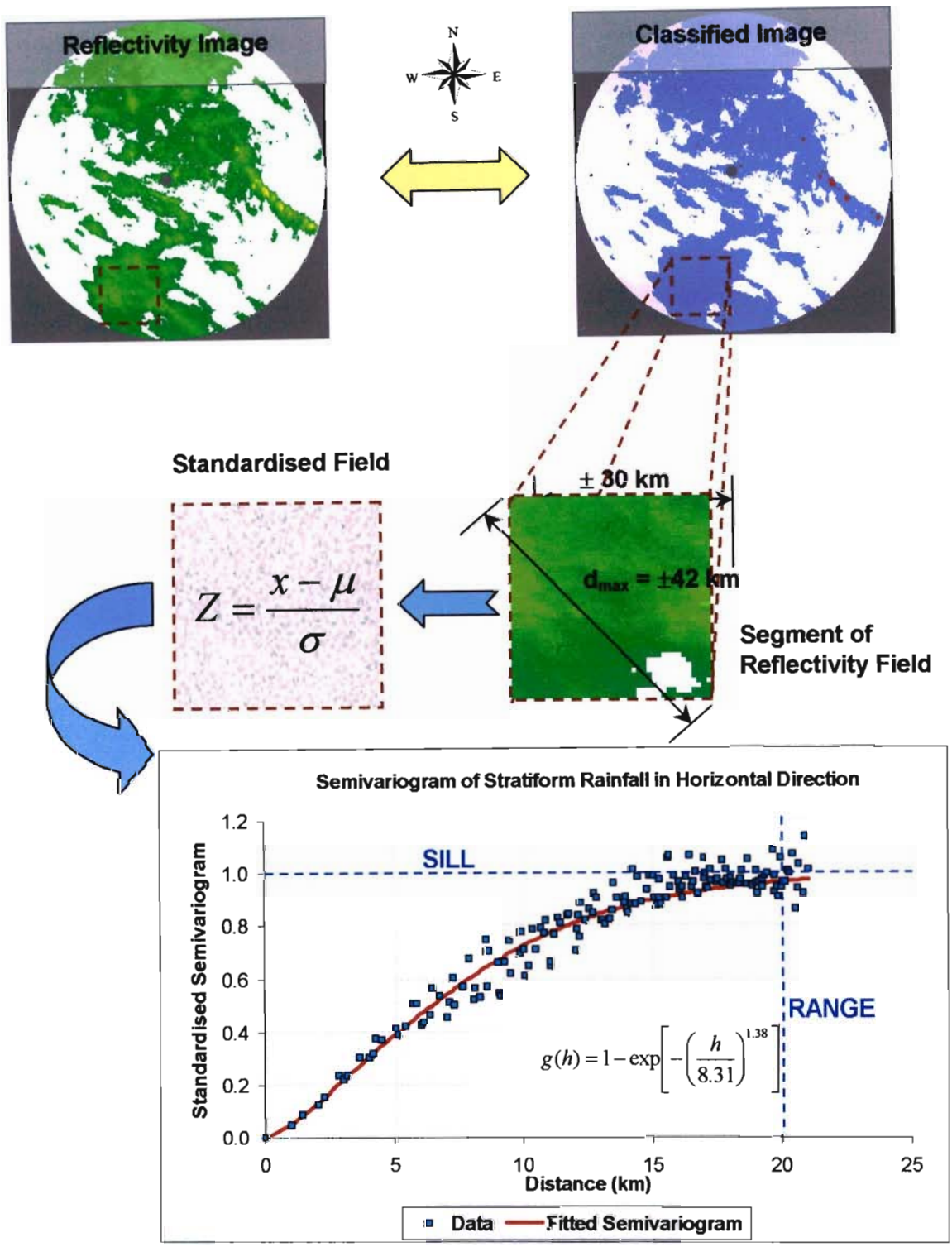


Figure 13: Typical semivariogram computed from a selected portion of stratiform rain with the sill and approximate range indicated. The maximum disatnce (d_{max}) is 42 km, h is limited to 21 km as recommended by Journel and Huijbrechts (1978: 194).

PARAMETER FITTING TO RAINFALL TYPES

Computing the empirical semivariogram values and then applying a model fitting routine to solve for the α and L parameters in Eqn. (14) is a computationally burdensome and time-consuming task, especially if this is done each time an unknown reflectivity value is estimated. This problem is exacerbated in the context of real time calculation, when the images are collected every 5 minutes. An alternative is therefore needed that is computational efficient yet does not compromise the accuracy of the estimates. Reducing the computational burden of computing the semivariogram by the use of a climatological semivariogram was employed for raingauge interpolation by Lebel *et al.* (1987) in order to improve computational efficiency. This idea was explored in the context of radar rainfields.

A set of instantaneous rainfall images was selected from the Bethlehem and Durban weather radars ranging over a period of several years (1995, 2000, 2001 and 2002). The images consisted of a variety of rainfall types ranging from images containing solely stratiform rain to images containing a combination of both stratiform and convective rainfall.

The rainfall classification algorithm of Table 1 was applied to the CAPPI 4 km above ground level. From the images various rectangular portions were selected that contained either predominantly stratiform or convective rainfall. These regions were approximately 1200 km^2 (40 km by 30 km) in area or larger. From the selected regions the Robust semivariogram was computed and the two parameter exponential model (Eqn. (14)) fitted. This was done separately in both the horizontal and vertical direction. The two parameters, α and L estimated from each field, were then plotted against one another in a scatter plot. This is shown in Figure 14 for the horizontal direction and Figure 15 for the vertical direction.

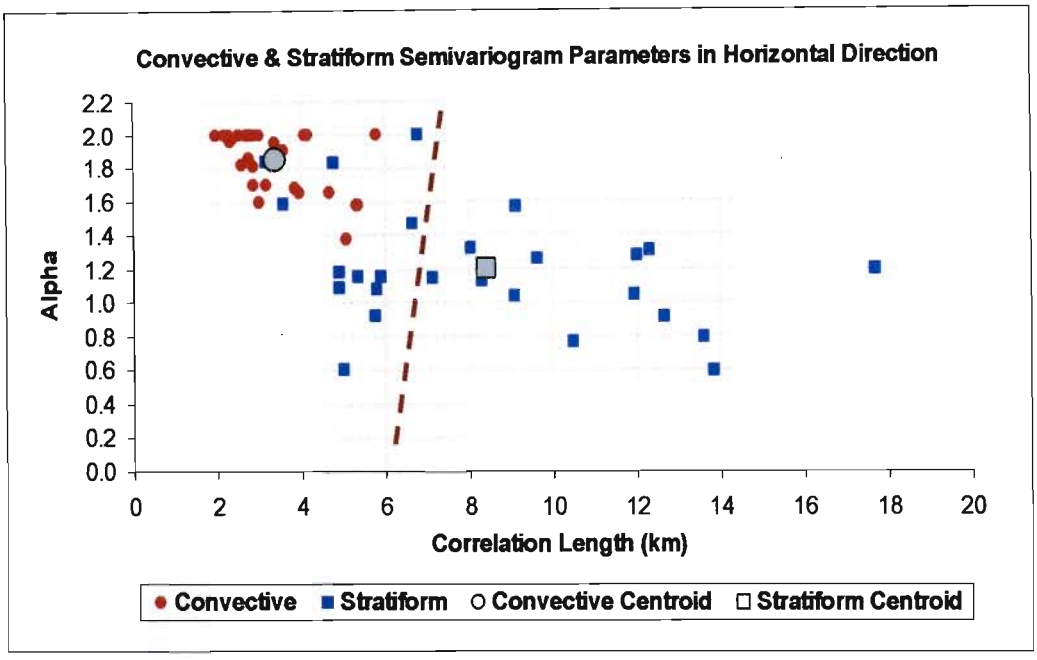


Figure 14: Scatter plot of alpha (α) and correlation length (L) parameters for stratiform and convective rainfall in a horizontal direction. The dashed line indicates the separation boundary of the two clusters as identified by a Fuzzy C-Means cluster algorithm.

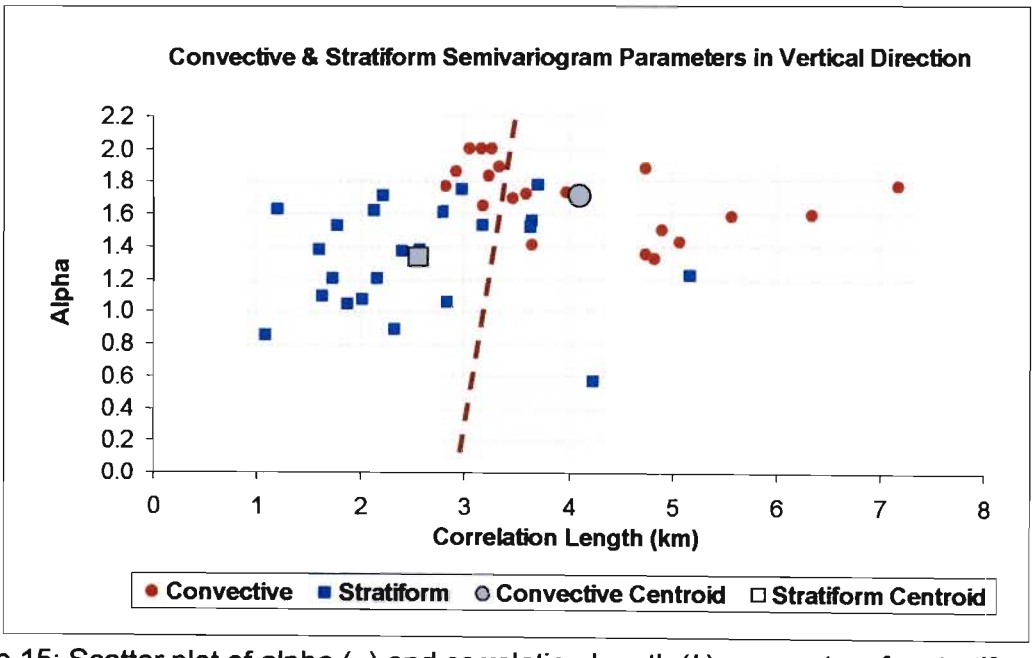


Figure 15: Scatter plot of alpha (α) and correlation length (L) parameters for stratiform and convective rainfall in a vertical direction. The dashed line indicates the separation boundary of the two clusters as identified by a Fuzzy C-Means cluster algorithm.

The scatter plots in Figure 14 and 15 illustrate that the semivariogram parameters tend to cluster in a particular region according to the type of rainfall, indicating that a fixed set of parameters can comfortably be used depending on the rainfall type.

CLUSTER ANALYSIS

To test if a natural clustering of the variables occurs, and also to ascertain the degree of clustering in the vertical and horizontal directions, two cluster analysis techniques were utilised. Initially a Fuzzy C-Means clustering algorithm (Gordon, 1981: 58-60) was run on the scatter plot of the data, where the specification set in the algorithm was that the data should be divided into two groups. In the horizontal direction the Fuzzy C-means cluster algorithm split the groups roughly along the 6.5-kilometre correlation length. In this instance 24% of the points were misclassified and associated with the incorrect rain type. When the clustering algorithm was run on data in the vertical direction it resulted in a division running approximately along the 3.5-kilometre correlation length, in this instance 30% of the points were misclassified. The divisions for the clusters in the vertical and horizontal direction determined by the Fuzzy C-means cluster algorithm are indicated on Figures 14 and 15 by the dashed line.

The cophenetic correlation coefficient (c) (Gordon, 1981: 125) was also used to determine the degree of clustering of each group. If the value of the cophenetic correlation coefficient is close to one it indicates a high degree of clustering, while a value close to zero indicates a low degree of clustering (Gordon, 1981: 125). As indicated in Table 3 the cophenetic correlation coefficient shows a high degree of clustering as all the values are close to one. Also indicated by the values in Table 2 is that the convective clusters show a higher degree of clustering than the stratiform clusters in both the horizontal and vertical directions.

Table 3: Cophenetic correlation coefficient (c) (Gordon, 1981: 125) for convective and stratiform semivariogram parameter clusters in the horizontal and vertical direction.

	HORIZONTAL		**VERTICAL**	
	Convective	**Stratiform**	**Convective**	**Stratiform**
c	0.842	0.724	0.849	0.752

Because the semivariogram parameters naturally cluster around a centroid depending on the rainfall type, it seems to be a safe assumption to use fixed climatological

semivariogram parameters depending on the type of rainfall classified; the adopted values are given in Table 3. This idea is tested in the next sub-section.

SENSITIVTY ANALYSIS OF CLIMATOLOGICAL SEMIVARIOGRAM PARAMETERS

To determine how sensitive final Kriged solutions are to the use of a fixed set of semivariogram parameters, a sensitivity analysis was undertaken. In the testing procedure, various 2D data sets of radar reflectivity values were selected that contained solely convective or stratiform rainfall. Random portions of the selected reflectivity data were then removed and Ordinary Kriging was used to estimate the missing data, but in this instance the Ordinary Kriging was carried out several times using different combinations of semivariogram parameters (a brief overview of Kriging methods will be given in the next section). Five different sets of parameters were used: firstly the centroid value of α and L for convective and stratiform rain and then the α and L values one standard deviation away from the centroid values, as computed by Eqns. (17) and (18) and illustrated in Figure 16 for the horizontal direction.

$$\alpha \pm \sigma_\alpha \quad , \quad 0 < \alpha \leq 2 \tag{17}$$

$$L \pm \sigma_L \quad , \quad 0 < L \tag{18}$$

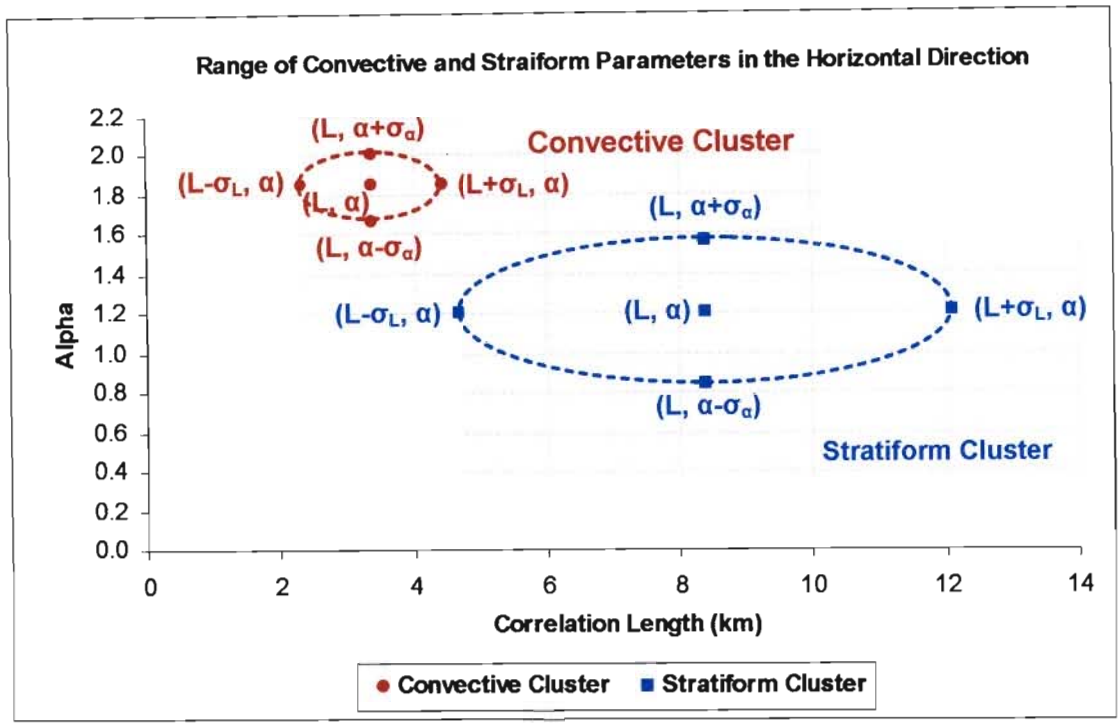


Figure 16: Schematic of the sensitivity analysis for the stratiform and convective clusters in the horizontal direction. The range of expected parameters for each cluster is indicated by the ellipses.

The infilled data were then compared to the original observed reflectivity data by computing the SSE, the means and standard deviations of the observed and estimated data. As indicated by Figures 17 and 18 there is no significant difference between the estimated values returned for the expected range of parameter values. The estimated Kriging values appear to be more sensitive to the shape parameter, α, than the correlation length, L.

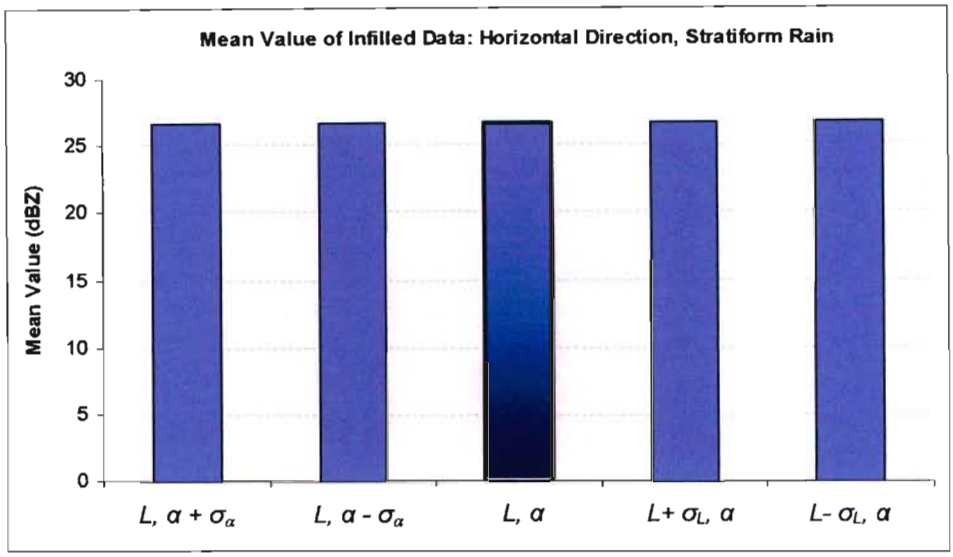


Figure 17: Comparison of the mean value for an infilled region with five different combinations of alpha (α) and correlation length (L) parameters as indicated on the horizontal axis on the graph.

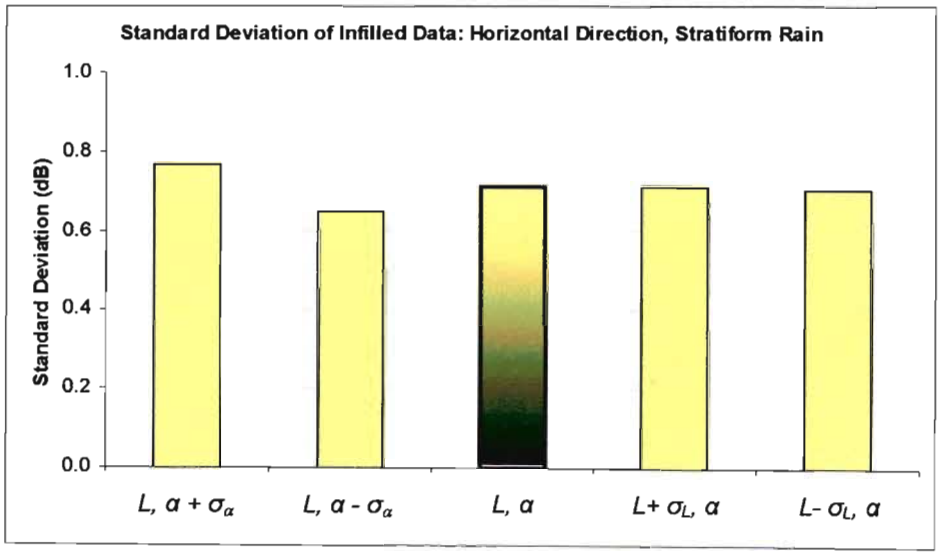


Figure 18: Comparison of the standard deviation for an infilled region with five different combinations of alpha (α) and correlation length (L) parameters as indicated on the horizontal axis on the graph.

The results indicate that there is no significant difference between the five sets of parameter values used in each instance, giving a clear indication that the final Kriged solution is reasonably insensitive to the range of the shape parameter, α, and correlation length, L, values that can possibly occur when a sample semivariogram is fitted to a specific rainfall type. Instead of solving for each set of semivariogram parameters for each neighbourhood one can simply use the centroid value computed for each rainfall type depending on the observed rainfall.

APPLICATION TO 3D REFLECTIVITY DATA

When selecting controls for computing the estimated value of a target in 3D space, three possible combinations of points can be selected: (1) all convective, (2) mixed, containing both stratiform and convective points and (3) all stratiform. For situations with either all convective or all stratiform, the appropriate subset of parameters in Table 4 are utilised for all estimations.

Table 4: Semivariogram parameter values in the horizontal and vertical directions for stratiform and convective rainfall.

	HORIZONTAL		**VERTICAL**	
	α_H	L_H **(km)**	α_V	L_V **(km)**
STRATIFORM	1.53	8.40	1.33	2.56
CONVECTIVE	1.85	3.38	1.71	4.11

When it comes to the practical calculation of the semivariogram elements in the Kriging equations (Eqn. (21)) there are no numerical or technical difficulties if the controls are all of one type, either convective, or stratiform. A problem arises when the controls are of mixed type. It was found that if the individual semivariogram elements in the equation were set to their type, with a compromise value for a mixed pair, unpredictable instabilities occurred in the solution of the equation. The methodology adopted was to use a linear weighting of the α and L parameters, where the weights are the proportion of type. For example in a 2D situation in the horizontal direction if, 25 controls are selected of which 15 are of convective pixels and 10 stratiform pixels, the computed semivariogram parameters would be L_H = 5.39 km and α_H = 1.72, and these values would be used for all calculations of $\gamma(h)$ in the Kriging equations (Eqn. (21)).

When Kriging with a 3D data set, the semivariogram model developed by Seed and Pegram (2001) was modified and is given by Eqn. (19):

$$\gamma(h) = \sigma^2 \cdot \left[1 - \exp\left(-h^{\alpha}\right)\right] \tag{19}$$

where: $h^2 = \left(\frac{r}{L_H}\right)^2 + \left(\frac{z}{L_V}\right)^2$

and $r^2 = x^2 + y^2$

here σ^2 is the field variance, h is the scaled distance in spherical co-ordinates, α is the scaling exponent, r is the distance in the horizontal plane, L_H the horizontal correlation length, z is the distance in the vertical direction, L_V the vertical correlation length and x and y are the distances in the Cartesian horizontal direction.

KRIGING METHODS

Kriging was chosen as the computational method to interpolate the observations missing internal to the CAPPI stack and to extrapolate the reflectivity values contained aloft to ground level. Kriging is considered to be the optimal technique for the spatial prediction of Gaussian data (Cressie, 1993: 106). In this application two types of Kriging were used, Ordinary and Universal Kriging (also called External Drift Kriging (Hengl *et al*, 2003)).

ORDINARY AND UNIVERSAL KRIGING

In Ordinary Kriging, the mean is assumed constant and unknown throughout the field whereas in Universal Kriging, the mean is also assumed to be unknown but varying (Chiles and Delfiner, 1999: 151). The Ordinary Kriging equations are given by Eqn. (20):

$$z(s_0) = \lambda^T(s_0) \cdot z \tag{20}$$

where $z(s_0)$ is the value to be estimated at the target s_o, z is the vector of known reflectivity values, or controls, and the row vector $\lambda^T(s_0)$ contains the calculated weighting values which depend on s_0. The vector of weighting values is computed by the matrix Eqn. (21):

$$\begin{bmatrix} G & u \\ u^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda(s_0) \\ \mu(s_0) \end{bmatrix} = \begin{bmatrix} g(s_0) \\ 1 \end{bmatrix} \tag{21}$$

where G is a matrix of semivariogram values between the controls, u a unit vector of ones, $\lambda(s_0)$ the vector of weighting values, $\mu(s_0)$ a Lagrange multiplier and $g(s_0)$ a vector of semivariogram values between the target at s_o and the controls locations; note the dependence of the unknown and right hand side section on the position of the target s_0. In Universal Kriging variables can be added to Eqn. (21) in order to model the variation of the mean throughout the field. This modification is indicated in Eqn. (22.1), by the addition of Q_c to the coefficient matrix and $q_t(s_0)$ to the right hand side and β to the vector of weighting values:

$$\begin{bmatrix} G & u & Q_c \\ u^T & 0 & 0 \\ Q^T{}_c & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda(s_0) \\ \mu(s_0) \\ \beta \end{bmatrix} = \begin{bmatrix} g(s_0) \\ 1 \\ q_t(s_0) \end{bmatrix} \tag{22.1}$$

where the additional terms: $$Q_c = \begin{bmatrix} q_1(s_1) & \cdots & q_p(s_1) \\ \vdots & & \vdots \\ q_1(s_n) & \cdots & q_p(s_n) \end{bmatrix} \tag{22.2}$$

The selected variable $q_j(s_i)$ is associated with the point s and could be altitude, mean annual precipitation or some other classification variable. They range from $q_1(s_i)$ to $q_p(s_i)$ and are associated with the selected control. The vector $q_t(s_0)$ contains the corresponding variables associated with the target. Typical models for the variation of the mean are either linear or polynomial; alternatively variables associated with physical characteristics of the controls may be included, that influence the mean value of the field.

COMPUTATIONAL STABILITY AND EFFICIENCY

In order to supply an estimate of the rainfall at ground level in a real time basis, computational efficiency is of the utmost importance. In South Africa, as volume scan data are to be processed at approximately 5 minute intervals, an estimated 120 000 targets need to be computed in this time. One of the disadvantages of the Kriging technique is that it relies on the solution of a linear system of equations whose size is proportional to the number of selected controls. For large systems of equations this can be time

consuming, computational burdensome and potentially unstable. Previous authors have also made reference to the computational burden associated with Kriging (e.g. Creutin and Obled, 1982).

There are several approaches that can be adopted for Kriging target data that can be adopted for Kriging target data, some of which are:

- compute all target values together.
- separate the targets into contiguous sets and identify their boundary controls.
- use neighbourhood kriging to estimate each target value individually.

It was ascertained that the last of the three options is the most accurate, most stable and fastest of the three by an order of magnitude (Wesson and Pegram, 2004) and is the method adopted here. One of the properties of Kriging that can be exploited to overcome the problem of dimension is the 'screening effect' (Chiles and Delfiner, 1999: 202-206). The screening effect refers to the observation that the significant weighting values are concentrated around the target, with the weighting values rapidly decreasing with distance from the target. For each target only a cluster of the nearest controls needs to be selected. In this application the computed optimum number of controls based on an exhaustive comparison, is twenty-five. This has the effect of drastically reducing the computation time with little or no loss in accuracy of the final results.

Another unexpected problem associated with the Kriging technique, which is not so well known, is the ill-conditioning of the coefficient matrix. It was demonstrated by Wesson and Pegram (2004) that the coefficient matrix can be highly ill-conditioned dependent on a number of factors, but most notably the chosen parameterization of the semivariogram function. This is especially evident as the α value tends to become Gaussian ($\alpha \rightarrow 2$). The method of Singular Value Decomposition (SVD) is therefore used in conjunction with a trimming of the singular values to provide a computational stable method to find the solution to the Kriging equations. Due to the fact the singular values are trimmed, matrix rank reduction techniques can then be exploited to dramatically decrease the time to find the inverse solution of the coefficient matrix with little or no loss in accuracy of the final Kriged solution (Wesson and Pegram, 2004).

DEVELOPMENT OF ORDINARY AND UNIVERSAL KRIGING TECHNIQUES

In the application of the Kriging technique the control data selected divides into three categories which are determined by the 25 controls in the selected neighbourhood and are defined as follows:

- Stratiform - controls consist entirely of stratiform rain.
- Mixed / Intermediate - controls consist of a combination of stratiform and convective rain.
- Convective - controls consist entirely of convective rain.

Although the rain is classified as purely stratiform or convective rainfall an intermediate rainfall zone is implied for computation purposes. Ordinary Kriging was used in the "pure" convective and stratiform zones while Universal Kriging was used in the mixed zone. The problem to be solved was how to get the best combination in the mixed cases.

In order to determine the most effective form of Krigng to use and the most appropriate variables to use in Universal Kriging a validation technique was devised. The CAPPII 2 km above ground level was selected as the target level for a variety of weather types. The data from aloft was then used to estimate all the reflectivity data on the 2 km CAPPI. The estimated data and observed data were then converted from reflectivity to rainrate values and the rain separated into the three categories listed above. These were compared in terms of SSE, means and standard deviations. In this exercise only the northern half of the image was considered in order to ensure there were no ground clutter reflectivity values in the data. Example of this procedure is illustrated in Figure 19. In this way the most appropriate form of Kriging and variables for Universal Kriging can be selected for each particular rainfall zone.

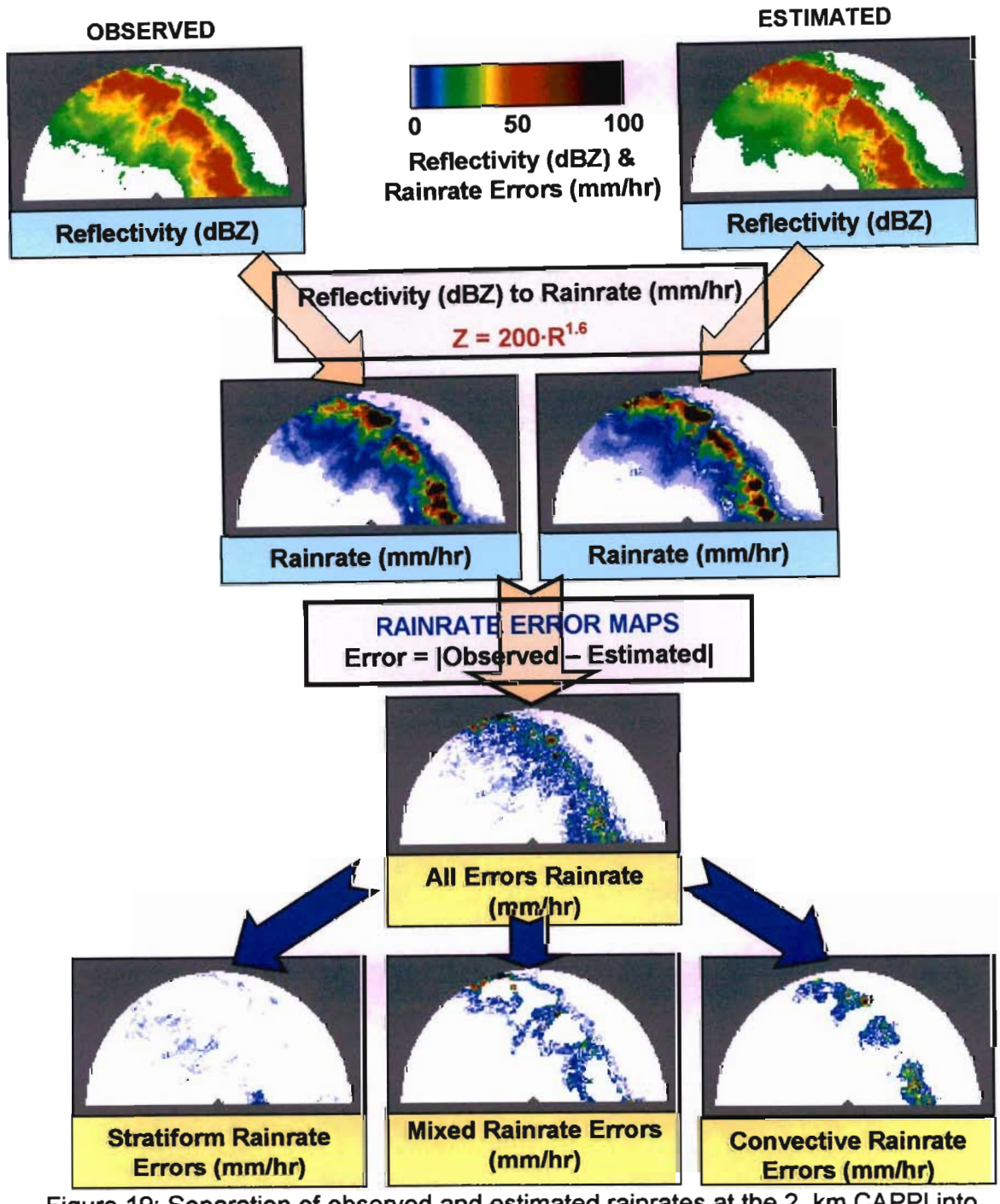


Figure 19: Separation of observed and estimated rainrates at the 2 km CAPPI into stratiform, mixed and convective rainrate zones and the corresponding errors for each zone in order to determine effectiveness of Ordinary and Universal Kriging in each rainfall type. The estimated image at level 2 was completely filled in from the information above it.

STRATIFORM ZONE

In the stratiform zone Ordinary Kriging is used to compute the estimate of the target at the 2 km level. Universal Kriging was investigated by modelling the vertical variation of the

mean by use of either a second order polynomial or a linear relationship to take advantage of the trend in the vertical reflectivity profile. This did not result in any significant improvement in terms of the sum of the difference squared between the estimated and observed rainfall when the trend was ignored (Ordinary Kriging). A computational advantage of Kriging in the stratiform zone is that SVD does not need to be used in order to compute the inverse of the coefficient matrix. This is due to the fact that the shape parameter, α, for stratiform rain is close to unity (exponential) which results in a stable coefficient matrix. By computing the condition number, which for a symmetric matrix is defined as the ratio of the largest to the smallest singular value (Wilkinson, 1988: 191), it was shown that the coefficient matrices for stratiform rain were much more stable than the coefficient matrices for mixed and convective rain. In this instance the quicker LU decomposition algorithm can be employed to find the solution to the Kriging equations.

CONVECTIVE ZONE

The controls in this zone consist entirely of convective rainfall, so SVD is used to compute the inverse of the coefficient matrix because the shape parameter, α, is close to Gaussian. Universal Kriging was not found to provide a significant improvement in the estimates over Ordinary Kriging. As shown in Figure 6, as the convective rain falls towards ground level there is a continuous and significant increase in rain intensity. Based on this, it was decided to use the vertical distance from the control to the targets as the external variables in Universal Kriging. Initially a linear and second order polynomial relationship were investigated. However the linear relationship tended to overestimate the convective rainfall and the second order polynomial provided an underestimate at the target. Based on the above, Ordinary Kriging is now used when the controls selected consist entirely of convective pixels.

MIXED / INTERMEDIATE ZONE

In the mixed zone, SVD is used to compute the inverse of the coefficient matrix because, for convective rain, the shape parameter α is close to Gaussian as seen in Table 4. This results in a numerical unstable coefficient matrix (Wesson and Pegram, 2004). In addition Universal Kriging in this instance provides an improved estimate over Ordinary Kriging. Testing the method on 10 different instantaneous images showed an improvement in estimates (in terms of SSE) on 8 of the images. Since the controls now consist of both stratiform and convective pixels a stratiform/convective binary switch was used as the two

external variables in Universal Kriging. This is indicated as a specialization of Q_c in Eqn. (22.2) and is given as Eqn. (23):

$$Q_c = \begin{bmatrix} C_1 & S_1 \\ \vdots & \vdots \\ C_n & S_n \end{bmatrix} \tag{23}$$

Where *C* represents a convective and *S* a stratiform control; so that a convective control is assigned the values *[C S] = [1 0]*; and a stratiform control is assigned the values *[C S] = [0 1]*.

A decision needs to be made as to which semivariogram is to be used in the coefficient matrix and which switch is to be set in Q_c (Eqn. (23)). This also affects the choice of q_t in the right hand side of Eq. (22.1). It was decided to let this be determined by examining the controls; if the majority of the controls are convective then the target is assumed to be convective, if not then it is assumed to be stratiform. Including these external variables resulted in an improvement in the estimates. An example of this comparison is indicated by Figure 20 where first Ordinary then Universal Kriging were used to estimate the mixed rainfall pixels on an instantaneous image (24 January 2002) from the Bethlehem weather radar; the number of pixels to be estimated were 2628. As indicated in Figure 20, Universal Kriging provides an improved estimate over Ordinary Kriging.

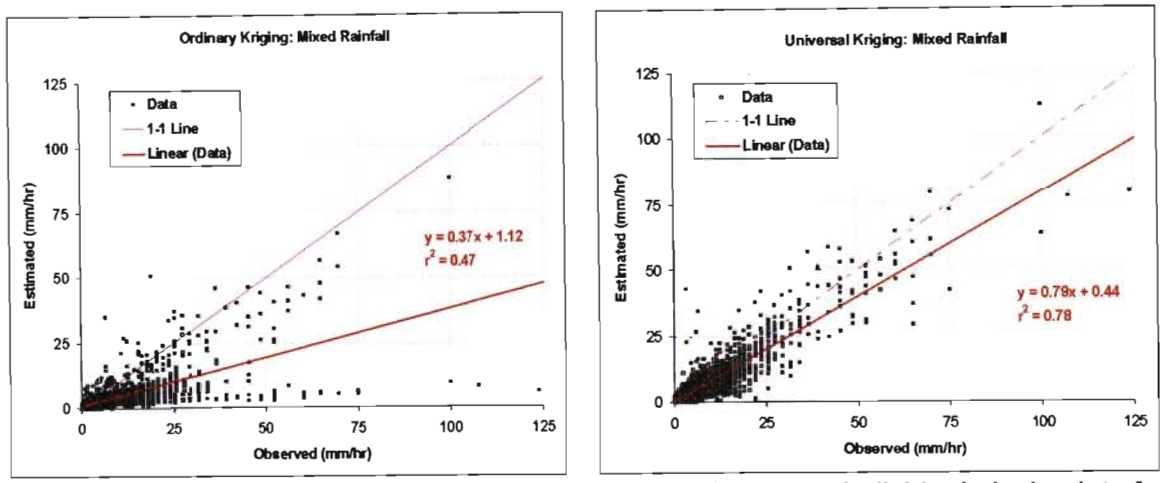


Figure 20: Scatter plot of observed and estimated rainrates at individual pixel points for mixed rainfall for a instantaneous image from the Bethlehem weather radar, 24 January 2002, the number of pixels estimated number 2628. Estimates are computed with both Ordinary and Universal Kriging with the Universal Kriging scatter plot indicating a substantial improvement in estimates.

CASCADE KRIGING

When Kriging directly to ground level using the full set of CAPPI data in a stack, there are unexpected problems which occur. A 24-hour accumulation of rainfall calculated from 5-minute images Kriged from the CAPPI volume down to ground level is shown in Figure 21. The data are from the Bethlehem weather radar (25 January 1996); the image is 200 km square. As can be seen, Figure 21 shows serious discontinuities at the regions located directly under the edges of the CAPPIs and also shows an inflation of the reflectivity estimates due to the numerical and geometric distribution of Kriging weights at the CAPPI edge locations.

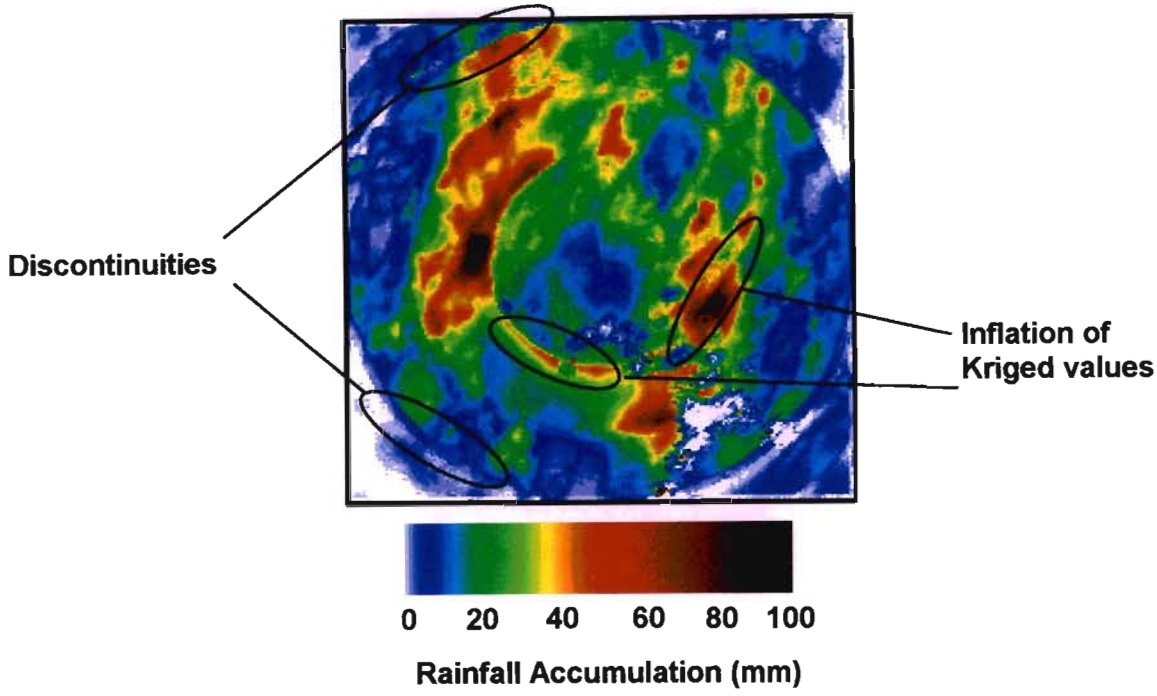


Figure 21: 24-Hour rainrate accumulation for the Bethlehem weather radar (25 January 1996). Image dimensions are 200 km square. Indicated on the image are the serious discontinuities associated with the CAPPI edges and also the inflation of the Kriged estimates at CAPPI edges.

A solution to the above problem is to use a Cascade Kriging approach that works as illustrated in Figure 22. The first step is infill the missing data in the CAPPI at 4 km above ground level, pixel by pixel, using neighbourhood Kriging of the 25 control data vertically above and in the horizontal direction. Once all the unknown data on that CAPPI have been estimated and infilled, the CAPPI directly below is examined and any missing data here are estimated in the same manner; once again estimates from above (which now can include previously infilled data) and at the same level are used as control data. This is repeated until all the unknown data from in each CAPPI are estimated. The final step involves an estimate of the rainfall at ground level.

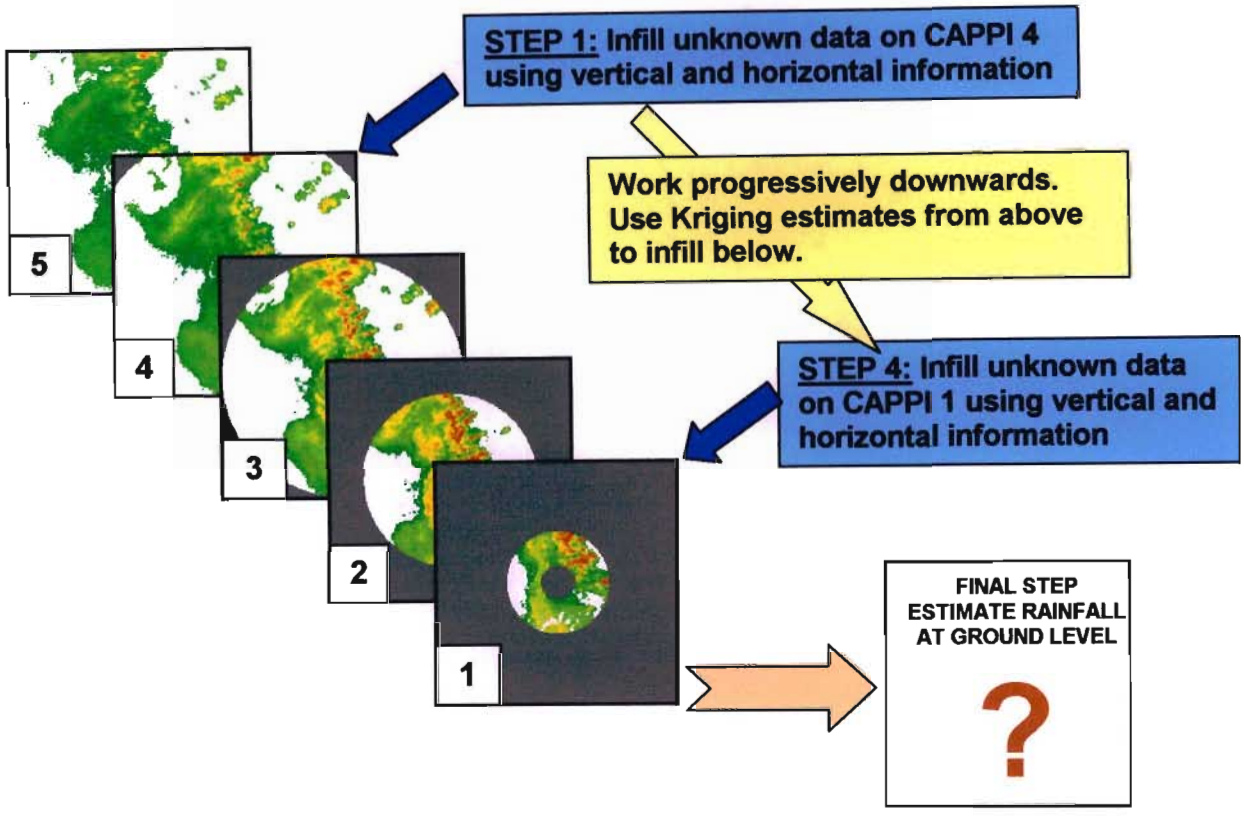


Figure 22: Cascade Kriging algorithm steps. Image taken from the Bethlehem weather radar (24 January 1996) where the images are 200 km square and the grey portions indicate regions where the data are missing.

Figure 23 gives a 3D illustration of radar volume scan data before and after Cascade Kriging. The data are from the Bethlehem weather radar, 14 February 1996, where the image is 200 km square, and the vertical extent is 3 km. The levels above 3 km have been ignored for clarity, but were used to estimate from level 4 downwards as indicated in Figure 22. The image on the left hand side indicates the volume scan data before Cascade Kriging and the right hand image represents the same volume scan data after Cascade Kriging has been used to estimate as much unknown data as possible and provide an estimate at ground level.

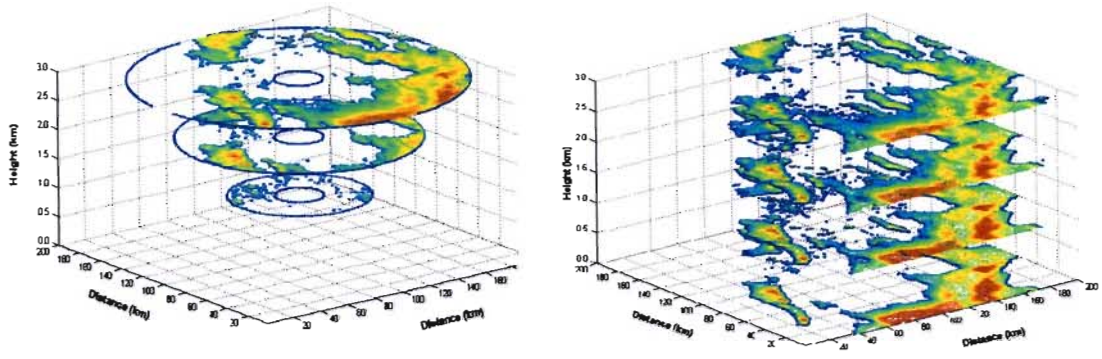


Figure 23: 3D Graphical illustration of radar volume scan data up to 3 km above ground level, before and after Cascade Kriging. Data are from the Bethlehem weather radar (14 February 1996). Data at levels 4 km and higher are not shown, but were used to successively fill the stack from above.

Cascade Kriging has the following influences on reflectivity data. There is increased smoothing of the reflectivity data as ground level is approached. There is also an increase in the reflectivity values as ground level is approached due to the pattern of Kriging Weights as illustrated in Figure 24. This is because, in the case where the reflectivity values increase with proximity to ground level, the negative weights on the upper level and the positive weights on the lower level create a gradient that results in an increase in the magnitude of the Kriged values, which corresponds with the climatological profiles shown in Figure 6.

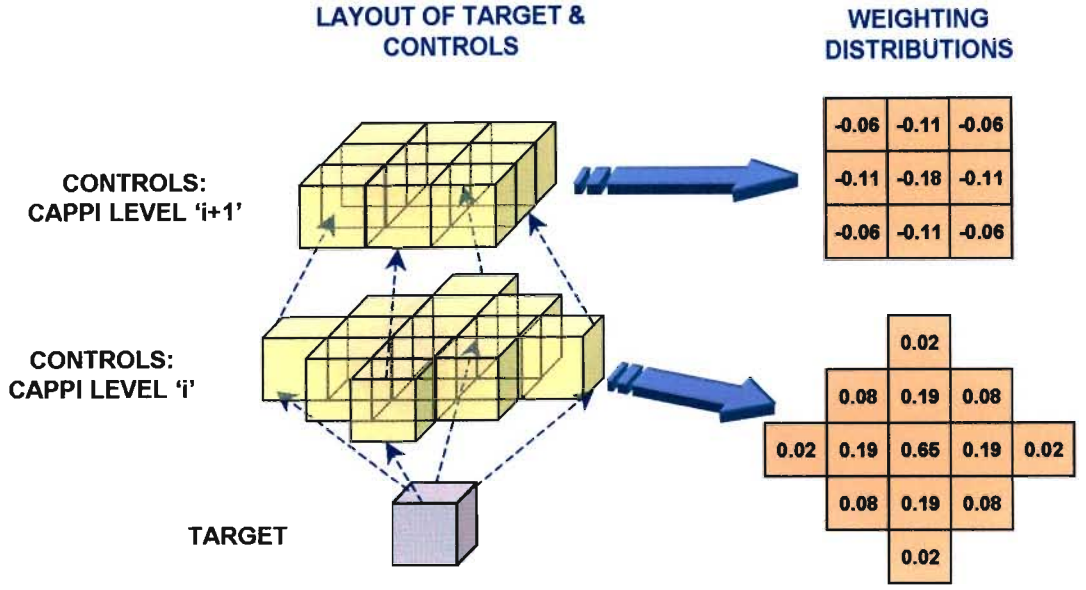


Figure 24: Layout of Kriging weightings, computed with convective semivariogram parameters, for 22 controls centred directly above a target. The negative weights are situated on the upper CAPPI and the positive weights on the lower CAPPIs and maintain a gradient in the field if there is one above the target.

Figure 25 shows an example of the smoothing and the increase in magnitude of the reflectivity values as ground level is approached. The image on the left hand side is at the 4 km CAPPI and the right hand side image indicates the Cascade Kriging estimate at ground level. As can be seen the estimates are smoother and have increased in magnitude.

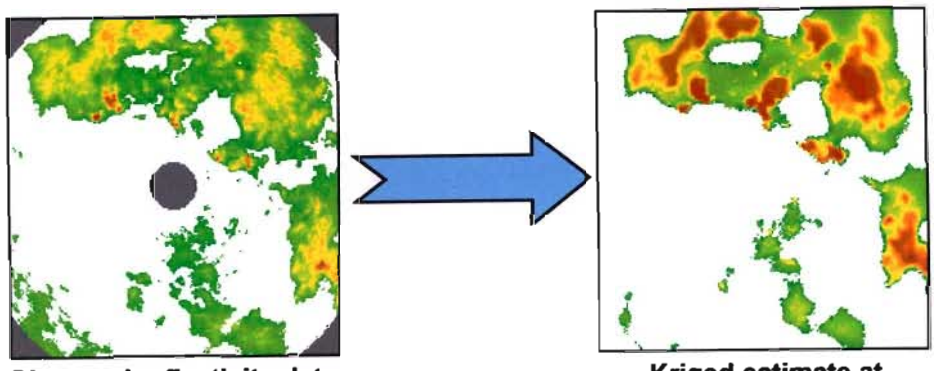


Figure 25: Example of smoothing and increase in magnitude of the reflectivity values as ground level is approached. The image is from the Bethlehem weather radar from the 25 January 1996, dimensions of the images are 200 km square.

The next section details the validation testing procedure adopted to determine the efficiency and accuracy of the algorithm.

TESTING AND RESULTS

In order to determine the efficacy of the proposed technique in providing accurate rainfall accumulation values at ground level, a comparison was carried out between raingauge and radar accumulation values for selected rain events; care was taken to compute values which were comparable at ground level. The Liebenbergsvlei catchment described in the Introduction and illustrated in Figure 2 was used as the test area for the comparison between radar and raingauge accumulation estimates.

The radar rainfall estimates at ground level are estimated in 1 km by 1 km cells and the raingauge provides a point estimate at ground level. In order to provide an appropriate way of comparing the two estimates, the following approach was taken. The average value of the 9 radar pixels (a 3 km square area), with the centre square covering a gauge, was taken as the radar estimate of average rainfall at each gauge location. In computing the radar accumulation values one needs to take into account that the images are sampled at five-minute intervals, so a simple linear accumulation of the images may not provide an accurate accumulation, especially for fast moving rain events. The accumulations are therefore computed by a Morphing Algorithm that takes into account the motion of the rainfield between instantaneous images (Sinclair and Pegram, 2003).

Block Kriging was used to determine the average rainfall over the same 9 km^2 area by using all raingauges within a range of 2 correlation lengths from the centre of the 9 km^2 area, at times corresponding to the radar estimates. The Block Kriging estimates are computed by Eqn. (24):

$$\hat{z}_D = \frac{1}{D}\int_D Z(x)\cdot dx \qquad (24)$$

where $\hat{z}_D$ is the average value over a selected region and D defines the region (Bras and Rodriguez-Iturbe, 1985: 402-404). Due to the wide spacing of the raingauges compared to the 3×3 km area, the semivariogram parameters used were the same as those determined for the radar reflectivity data in the horizontal plane. However a nugget effect is included, which for stratiform rain has a value of 0.90 and convective rainfall 0.97 as was determined by Habib and Krajewski (2002). This allows for the random errors and uncertainties associated with raingauge estimates. Illustrated in Figure 26 is the method of selection of the nearest raingauges to be used in the estimation of the average rainfall over the 9-pixel region, where the raingauges selected are within a range of 2 correlation lengths away from the centre of the 9-pixel region.

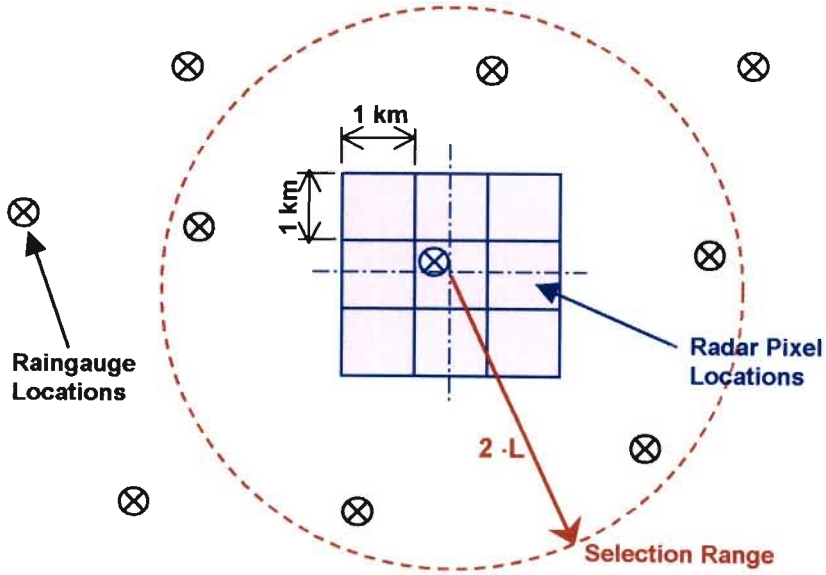


Figure 26: Method of selection of raingauges to compute Block Kriging estimate over the 9 pixels centred at a raingauge location. Raingauges within a range of 2 correlation lengths are selected to compute the Block Kriging estimate over the 9 pixel region.

The following two rain events were selected for the validation tests: (1) 24 January 1996 - a rain event which consisted of extremely high rainfall intensities over an approximately 12-hour period, (2) 13 February 1996 - a rain event consisting of a combination of convective rain with periods of stratiform rain.

STATISTICAL COMPARISON OF RAINGAUGE AND RADAR ACCUMULATIONS

In order to determine the quality of the radar estimates at ground level a statistical comparison of the raingauge and radar accumulations over different periods of 24 hours, 12 hours and 6 hours was conducted. A scatter plot of the radar and raingauge accumulations was done and the correlation coefficient computed. The Kolmogrov-Smirnov test was also used to determine if the cumulative distributions of the radar and raingauge accumulations were significantly different at a significance level of 5%, where the hypothesis being tested is H_0: Raingauge Distribution = Radar Distribution against H_1: Raingauge Distribution $\neq$ Radar Distribution. The means and standard deviations (σ) of the accumulated values were also tested at a significance level of 5%. In all the statistical tests undertaken a Box-Cox Transform was applied to the data in order to ensure the data sets approximated to a normal distribution.

RAINFALL EVENTS

RAINFALL EVENT 1: 24 JANUARY 1996

Severe convective rainfall was recorded in the late afternoon and evening of the 24 January 1996 (Mather *et al*, 1997). The analysis for this event was conducted over a 12-hour period from 12:00 to 24:00. For this rain event 29% of the images in the 12-hour accumulation period were identified as containing bright band and corrected. Figure 27 illustrates a 12-hour accumulation gives an indication of the extreme rainfall intensity during the 12-hour period. Indicated on Figure 27 is an outline of the catchment area as well as the raingauge locations. The region directly above the radar, a circle of 20 km radius, has also not been infilled due to poor data quality close to the radar during this time period; two raingauges within this region were excluded from the study.

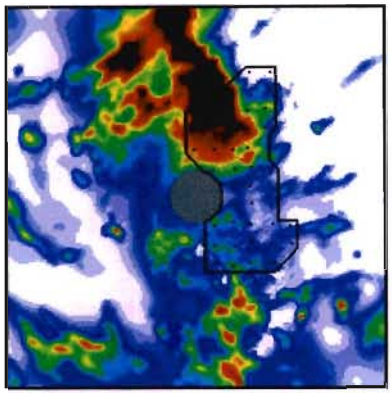


Figure 27: 12-hour rainfall accumulation from the Bethlehem radar on the 24 January 1996. An outline of the test catchment is indicated on the image as well as the raingauge locations. Dimensions of the image are 200 km square.

A scatter plot of the raingauge and radar accumulations after a 12 hour period at each of the 43 gauge sites computed by the appropriate Kriging method over each 9 km^2 region, is shown in Figure 28. There is a high correspondence for this rain event between radar and raingauge accumulations with an r^2 value of 0.86 being returned. The highest pair of radar observations bias what is otherwise a reasonable fit, and also affects Figures 29 and 31. This is probably due to under estimation of the highly localised convective rainfall by the gauges concerned (Wilson and Brandes, 1979).

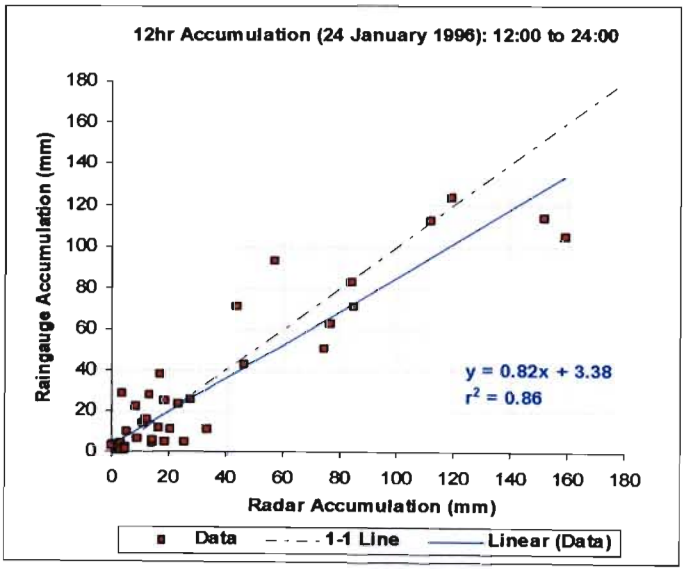


Figure 28: Scatter plot of radar and raingauge accumulations for a 12-hour accumulation period for the 24 January 1996 rain event. A strong correspondence exists between the radar and raingauge accumulation, especially for the high intensity rainfall.

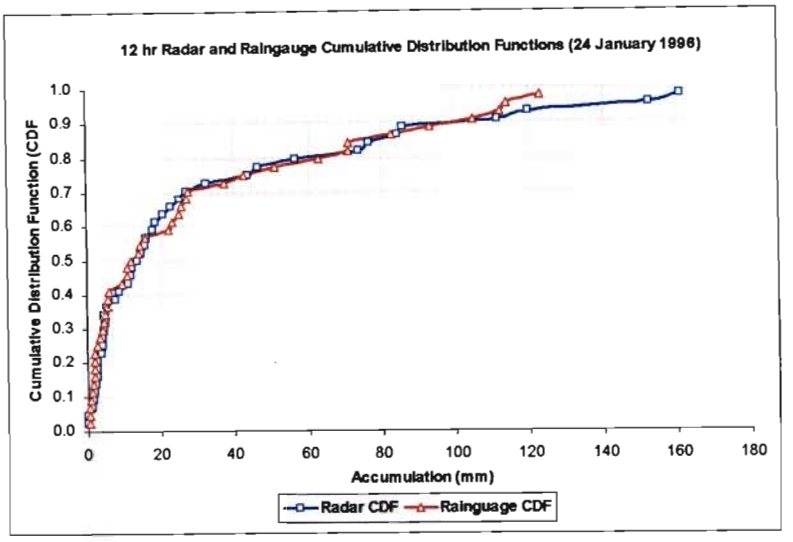


Figure 29: Cumulative Distribution Functions (CDF's) of the raingauge and radar accumulations for a 12-hour period. Accumulations values are from the 24 January 1996 (12:00 to 24:00). Testing at a significance level of 5% indicate that the distributions of the accumulations for the radar and raingauge are not dissimilar

Table 5 provides a summary of the statistical results returned for the raingauge and radar accumulation over a 12-hour period and two 6-hour periods. For the 6-hour period (18:00 to 24:00) and the 12-hour accumulation period there is a close correspondence between the radar and raingauge estimates. With the exception of one standard deviation, the means and standard deviations are statistically similar with the Kolmogrov-Smirnov test (K-S test) also indicating that the distributions of the accumulations for the radar and raingauge are not dissimilar, except for the 6-hour accumulation period from 12:00 to 18:00. The word "Accept" in Table 4 refers to the null hypothesis H_0 being accepted and H_1 rejected, and "Reject" refers to H_0 being rejected and H_1 accepted. Figure 29 shows an example of the cumulative distribution functions for 12-hour raingauge and radar accumulations where the distributions for this example were considered to be not dissimilar.

Table 5: Summary of statistical results for different accumulation periods for 24 January 1996 rain event for radar and raingauge data.

	6hr accumulation (12:00 to 18:00)		6hr accumulation (18:00 to 24:00)		12hr accumulation (12:00 to 24:00)	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	4.6	5.1	27.2	40.8	30.3	41.3
Raingauge	4.0	7.1	25.4	38.3	29.4	36.8
Accept / Reject H_0	Accept	Accept	Accept	Accept	Accept	Accept
r^2	0.08		0.88		0.86	
K-S TEST	Reject		Accept		Accept	

RAINFALL EVENT 2: 13 FEBRUARY 1996

The rain event on the 13 February 1996 consisted of a combination of convective and stratiform rain. The rainfall initially consisted of light stratiform rainfall but towards the evening and afternoon this area also experienced convective rainfall. In the 24-hour rain period 19% of the images were identified as containing bright band and corrected. The 24-hour accumulation for the 13 February 1996 is depicted in Figure 30 illustrating the combination of convective and stratiform rainfall that was recorded throughout the day. The exclusion of data directly above the radar applies in this instance as well. Two of the gauges were therefore excluded from the data analysis.

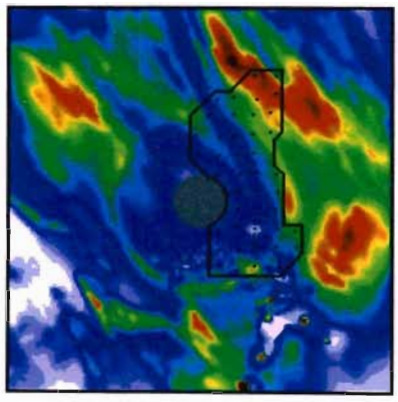


Figure 30: 24-hour rainfall accumulation from the Bethlehem radar on the 13 February 1996. An outline of the test catchment is indicated on the image as well as the raingauge locations. Dimensions of the image are 200 km square.

Figure 31 illustrates a scatter plot of the radar and raingauge areal average data for the 24-hour accumulation period of low intensity, wide spread, mostly stratiform rain, corresponding to Figure 30. The scatter plot indicates a strong correspondence between raingauges and radar estimates with a correlation coefficient, r^2, of 0.78 being returned. The higher rainfall accumulation values show a strong correspondence, while the lower rainfall values exhibit a weaker relationship. Again, the raingauges underestimate the highest two values of the rainfall when compared to the radar, inducing a bias because of the strong influence of the two largest values.

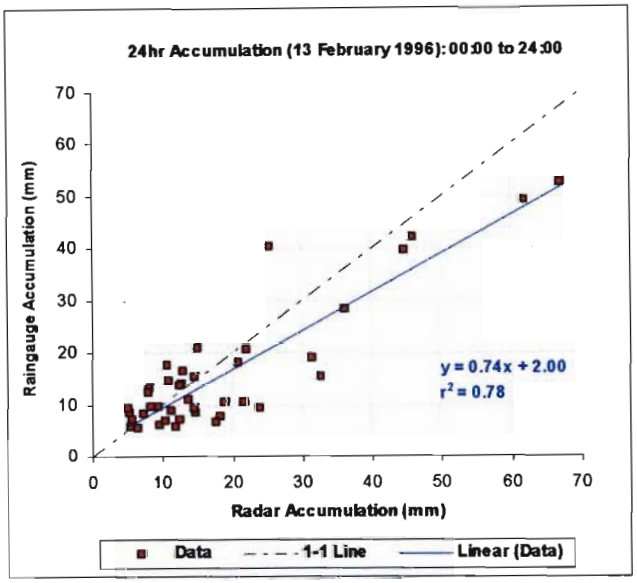


Figure 31: Scatter plot of radar and raingauge accumulations for a 24-hour accumulation period for the 13 February 1996 rain event. A strong correspondence exists between the radar and raingauge accumulation, especially for the high intensity rainfall.

Table 6 provides a summary of the statistical analysis of the comparisons between raingauges and radar accumulation for the full 24-hour period and two 12-hour periods. The mean and standard deviations for each of the accumulation periods closely correspond and are statistically similar, except for the mean value of the 24-hour accumulation. The correlation coefficient value, r^2, is also high except for the accumulation period from 12:00 to 18:00 which consisted of light stratiform rain. The cumulative distributions of accumulation values for each time period were also determined to be statistically similar via the K-S test.

Table 6: Summary of statistical results for different accumulation periods for 13 February 1996 rain event for radar and raingauge data.

	12hr accumulation (00:00 to 12:00)		12hr accumulation (12:00 to 24:00)		24hr accumulation (00:00 to 24:00)	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	1.1	1.4	17.0	14.5	18.1	14.7
Raingauge	1.0	1.5	14.8	11.9	15.8	12.2
Accept / Reject H_0	Accept	Accept	Accept	Reject	Accept	Accept
r^2	0.14		0.76		0.78	
K-S TEST	Accept		Accept		Accept	

The statistics of the comparatively high rainfall accumulations summarised in Table 5 and 6 indicate that the estimates of the rainfall at the ground from the radar images satisfactorily capture the amount and variability of the rainfall.

SUMMARY AND CONCLUSION

A technique has been presented that extrapolates radar volume scan data estimated at 1 km intervals above ground level to the earth's surface, and in the process estimates missing data in the CAPPI stack.

A rainfall classification was firstly done to separate the rainfall into stratiform and convective types. By separating the rainfall, climatological semivariograms for each rainfall type were defined, a procedure which ensures high computational efficiency with little loss in accuracy. The 2 km CAPPI is then examined to determine if a bright band is present and if so the pixels are adjusted on an individual basis. Both 3D Universal and Ordinary Cascade Kriging were then used to infill missing data in the CAPPI stack and finally provide a rainfall estimate at ground level. In the Kriging computations care was taken to ensure computational efficiency and stability.

The effectiveness of the above algorithm was tested on two different rainfall events exhibiting distinctly different types of rainfall by comparing raingauge and radar accumulation estimates at ground level. The technique demonstrated that both and high and moderate intensity rainfall exhibited similar estimates to raingauges whereas

stratiform rainfall showed a weaker correspondence to the raingauge estimates; the research is ongoing. The Cascade Kriging technique is undergoing further testing currently at the Innovation and Research Division of SAWS to determine if it is operationally sound to implement.

ACKNOWLEDGMENTS

This work is currently supported financially by the South African Water Research Commission (WRC) and the National Research Foundation (NRF). Thanks go to Scott Sinclair for providing the computer code for the Morphing Algorithm to compute the radar accumulation values. Thanks must also go to meteorologists Deon Terblanche and Pieter Visser of SAWS: METSYS at Bethlehem, South Africa for invaluable advice regarding the rainfall classification and bright band correction algorithms.

REFERENCES

Bras, R.L. and Rodriguez-Iturbe, I., (1985). *Random Functions and Hydrology*. Addison-Wesley Publishing Company. Ontario.

Ciach, G. J., (2002). Local Random Errors in Tipping-Bucket Rain Gauge Measurements. *Journal of Atmospheric and Oceanic Technology*. Notes and Correspondence, **20**, 752 – 759.

Chiles, J.P. and Delfiner, P., (1999). *Geostatistics - Modeling Spatial Uncertainty*. Wiley, New York, USA.

Cressie, N., (1993). *Statistics for Spatial Data*. Wiley-Interscience Publication. New York, USA.

Cressie, N. and Hawkins, D.M., (1980). Robust Estimation of the Variogram, 1. *Journal of the International Association for Mathematical Geology*, **12**, 115-125.

Creutin, J.D. and Obled, C., (1982). Objective Analysis and Mapping Techniques for Rainfall Fields: An Objective Comparison. *Water Resources Research*, **(18)** 2, 413-431.

Franco, M., Sempere-Torres, D., Sanchez-Diezma, R. and Andrieu, H., (2002). A methodology to identify the vertical profile of reflectivity from radar scans and to estimate the rainrate at ground at different distances. *Proceedings of ERAD (2002)*. 299-304.

Gordon, A.D., (1981). *Classification: Methods for the exploratory analysis of multivariate data*. Chapman and Hall. New York, USA.

Habib, E. and Krawjewski, W.F., (2002). Uncertainty Analysis of the TRMM Ground-Validation Radar-Rainfall Products: Application to the TEFLUN-B Field Campaign. *Journal of Applied Meteorology*, **41**, 558-572.

Hengl, T., Geuvelink, G.B.M. and Stein, A., (2003). Comparison of Kriging with external drift and regression-kriging. *Technical note, ITC*, Available online at http://www.itc.nl/library/Academic_output/

Jordan, P., Seed, A. and Austin, G., (2000). Sampling errors in radar estimates of rainfall. *Journal of Geophysical Research*, **Vol. 105**, **No. D2**, 2247-2257.

Journel, A.G. and Huijbrechts, Ch. J., (1978). *Mining Geostatistics*. Academic Press. London, UK.

Lack, S.A. and Fox, N.I., (2004). Errors in Surface Rainfall Rates Retrieved from Radar due to Wind Drift. *Sixth International Symposium on Hydrological Applications of Weather Radar*. Melbourne, Australia.

Lebel, T., Bastin, G., Obled, C. and Creutin, J.D., (1987). On the Accuracy of Areal Rainfall Estimation: A Case Study. *Water Resources Research*, **Vol. 23**, **No. 11**, 2123–2134.

Marshall J. S. and Palmer W. M., (1948). The distribution of raindrops with size, *Journal of Meteorology,* **5**, 165-166.

Mather, G.K., Terblanche, D.E. and Steffens, F.E., (1997). *National Precipitation Research Programme: Final Report for the Period 1993 – 1996*. WRC report No 726/1/97.

Matheron, G., (1962). Traite de Geostatistique Appliquee, Tome I. *Memories du Bureau de Recherches Geologiques et Minieres*, **No. 14**. Editions Technip, Paris.

Mittermaier, M.P., (1999). *Investigating the Characteristics of the Radar Reflectivity Profile*. MSc Eng dissertation, Civil Engineering, University of KwaZulu-Natal, Durban, Howard College.

Mittermaier, M. P., (2003). *Investigating synergies between weather radar data and mesoscale model output*. PhD Thesis, Department of Meteorology, University of Reading.

Pegram, G.G.S. and Clothier, A.N., (2001). High resolution space-time modelling of rainfall: the "string of beads" model. *Journal of Hydrology*, **241**, 26-41.

Sabyasachi, B., Gunst, R.F., Guertal, E.A. and Hartfield, M.I., (1997). The Effects of Influential Observations on Sample Semivariograms. *Journal of agriculture, biological, and environmental statistics*, **2**, 490-512.

Sanchez-Diezma, R., Zawadzki, I. and Sempere-Torres, D., (2000). Identification of the bright band through the analysis of volumetric radar data. *Journal of Geophysical Research*, **Vol. 105, N0. D2**, 2225 – 2236.

Seed, A.W. and Pegram, G.G.S., (2001). Using Kriging to infill gaps in radar data due to ground clutter in real time. *Proceedings Fifth International Symposium on Hydrological Applications of Weather Radar – Radar Hydrology*. Kyoto, Japan. 73-78.

Sempere-Torres, D., Sanchez-Diezma, R., Zawadzki, I. and Creutin, J.D., (2000). Identification of Stratiform and Convective Areas Using Radar Data with Application to the Improvement of DSD Analysis and Z-R Relationships. *Phys. Chem. Earth,* **25**, 985-990.

Sinclair, D.S. and Pegram, G.G.S., (2003). Combining traditional and remote sensing techniques of rainfall measurement as a tool for Hydrology, Agriculture and Water Resources Management. *Proceedings 11th SA National Hydrology Symposium*, Port Elizabeth, South Africa.

Steiner, M., Houze, R.A. and Yuter, S.E., (1995). Climatological Characterization of Three-Dimensional Storm Structure from Operational Radar and Rain Gauge Data. *Journal of Applied Meteorology*, **34**, 1978 – 2007.

Terblanche, D.E., Pegram, G.G.S. and Mittermaier, M.P., (2001). The development of weather radar as a research and operational tool for hydrology in South Africa. *Journal of Hydrology*, **241**, 3 – 25.

Visser P., (2003). *Spatial interpolation and Mapping of rainfall: 2. Radar and Satellite products*, Draft final report (K5/1152) to the Water Research Commission for the period March 2000 to March 2003. Water Research Commission. Pretoria. 19 – 29.

Wesson, S.M. and Pegram, G.G.S., (2004). Radar rainfall image repair techniques. *Hydrology and Earth System Sciences*, **8(2)**, 220 - 234.

Wilkinson, J.H., (1988). *The Algebraic Eigenvalue Problem*. Oxford Science Publication.

Wilson J.W. and Brandes E.A., (1979). Radar measurement of rainfall – A summary, *Bulletin of the American Meteorological Society*, **60(9)**, 1048-1058.

4.8 CHAPTER SUMMARY

In this chapter the operational method for infilling all missing data contained in a radar volume scan and providing a rainfall estimate at ground level was detailed was described. The initial step required was to provide a rainfall classification algorithm to separate the radar volume scan into convective and stratiform rain zones which have statistically different characteristics in order to fit climatological semivariograms. A novel bright band correction algorithm based on local gradients was presented to correct the reflectivity values contained on the 2 km CAPPI in order to improve rainfall extrapolations to ground level.

The methodology to determine the appropriate semivariogram parameters to use in each rainfall zone and a mixed rainfall zone was discussed and validated. The use of Universal and Ordinary Kriging to provide extrapolations in each rainfall zone was investigated and there effectiveness evaluated. An original approach entitled Cascade Kriging was adopted to extrapolate radar rainfall estimates to ground level. The technique to extrapolate radar rainfall estimates to ground level in this chapter was evaluated on two rainfall events which both indicated a good correspondence between radar extrapolations and raingauge estimates.

The Cascade Kriging technique to infill missing and contaminated radar volume scan data has been designed to specifically work on a real time basis and be suitable for online application at the South African Weather Service (SAWS) and is currently being implemented.

The next chapter details the testing of the efficiency and effectiveness of the proposed algorithm in this chapter to: (1) Infill contaminated ground clutter locations contained in a radar volume scan and; (2) a more detailed analysis and overview of the extrapolations of the radar rainfall data to ground level and statistical comparison of the radar to raingauge accumulations will be discussed.

CHAPTER 5
EXTENDED TESTING AND VALIDATION RESULTS

5.1 INTRODUCTION

In this chapter the results for the algorithms developed and explained in Chapter 4 are evaluated in terms of their effectiveness and efficiency. The first part of this chapter details four rain events whose volume scans (CAPPIs) were extrapolated to ground level as described in the paper contained in Chapter 4. The rain events are analysed in more detail than in the paper and an additional two rain events are evaluated to compare the differences between raingauge and radar accumulations. Finally the efficiency and accuracy of the Cascade Kriging algorithm in the process of infilling regions of identified ground clutter in a radar volume scan will be evaluated. The Bethlehem ground clutter map and three distinctly different rain events were used to test the algorithm over 24-hour periods.

5.2 RAINFALL ESTIMATION AT GROUND LEVEL - COMPARISON BETWEEN RADAR ESTIMATES AND BLOCK KRIGED RAINAGAUGE ESTIMATES

To determine the quality of the radar Cascade Kriging estimates at ground level a comparison was carried out between the radar accumulations and Block Kriging raingauge estimates over 9 km^2 areas surrounding the gauges over 6, 12 and 24-hour accumulation periods as previously explained in detail in Section 4.7, pages 152 to 153.

5.2.1 Statistical Testing Procedure

Various statistical comparisons were undertaken between the radar and raingauge accumulations to determine the accuracy of the radar accumulations at ground level. Firstly a scatter plot of the raingauge and radar accumulations was done to determine the correspondence between the two estimates and the correlation coefficient (r^2) computed to determine the strength of the relationship between the two estimates.

The Cumulative Distribution Functions (CDF) of the raingauge and radar accumulation estimates were plotted and the Kolmogrov-Smirnov (K-S) test applied to determine if the CDFs of the two distributions were different at a significance level of 5%, where the test is:

H_0: *Raingauge Distribution* = *Radar Distribution* against

H_1: *Raingauge Distribution* $\neq$ *Radar Distribution*

Many statistical tests, such as the tests to be described below, assume the data set is normally distributed, however the radar and raingauge accumulation data are typically approximately lognormally distributed (Pegram and Clothier, 2001). To ensure that the test data approximates a normal distribution a symmetrizing transform is performed. The Box-Cox transform is utilised to achieve this and is given by Equation (5.1):

$$x_T = \begin{cases} (x^T - 1)/T \\ \ln(x) \quad if\ T = 0 \end{cases} \tag{5.1}$$

T is chosen so:

$$\left(Mean(x_T) - Median(x_T)\right)/ d_F \cong 0$$

Where x_T are the transformed data, T is the transform variable and d_F is the distance between the lower and upper quartiles (Pegram, 2000). The Box-Cox transform was applied to all data sets used for the tests described below.

The mean values of the raingauge and radar accumulations were also computed and tested to see if they were different at a significance level of 5%. The null hypothesis being tested was that the mean difference (μ_d) between the two paired estimates are not significantly different from zero. The hypothesis tested is summarised below:

H_0: $\mu_d = 0$ against

H_1: $\mu_d \neq 0$

A 95% confidence interval for μ_d can be computed by Equation (5.2) to determine whether H_0 is rejected:

$$\mu_d \pm t_{(n-1;1-\alpha/2)} \cdot \frac{s_d}{\sqrt{n}} \tag{5.2}$$

Where μ_d is the mean difference between the two estimates, $t_{(n-1;1-\alpha/2)}$ the cumulative percentage point from the Student's t-distribution, n the sample size, α the significance level (in this case 5%) and s_d the standard deviation of the differences of the two estimates (Dixon and Massey, 1983: 130-131).

The sample variances of the raingauge and radar accumulations were also computed and tested to see if they were significantly different at a significance level of 5%. The hypothesis tested is:

H_0: $\sigma^2_{radar} = \sigma^2_{rainaguge}$ against

H_1: $\sigma^2_{radar} \neq \sigma^2_{rainaguge}$

To determine if the null hypothesis is to be rejected, the following computations are done. Firstly the *F*-critical values, $F_{(n-1;\ n-1;\ 1-\alpha/2)}$ and $1/F_{(n-1;\ n-1;\ 1-\alpha/2)}$, are determined which are the 97.5% points from the *F*-distribution, where n refers to the sample size and α the significance level. The ratio of the two sample variances, f, are calculated as shown by Equation (5.3).

$$f = \frac{s_1^2}{s_2^2} \tag{5.3}$$

If the ratio, f, lies between the upper and lower *F*-critical values then the hypothesis H_0 is accepted and H_1 rejected (Dixon and Massey, 1983: 118-119).

5.2.2 Results Rain Event 1 - 24 January 1996

Severe convective rainfall was recorded in the late afternoon and evening of the 24 January 1996 (Mather *et al*, 1997). The analysis for this event was conducted over a 12-hour period from 12:00 to 24:00. Figure 5.1 illustrates a 12-hour accumulation of the estimate at ground level and gives an indication of the extreme rainfall intensity during the 12-hour period. Indicated on Figure 5.1 is an outline of the catchment area as well as the raingauge locations. The region directly above the radar, a circle of 20 km radius, has not been infilled due to poor data quality close to the radar during this time period; two raingauges within this inner region were excluded from the study as described in Chapter 4.7, page 154.

6-HOUR ACCUMULATIONS

The 6-hour accumulations indicate that for the times 12:00 to 18:00 light and sparse rainfall was recorded over the catchment by the weather radar. For the second 6-hour accumulation the rainfall was very intense and predominantly in the Northern segment of the catchment. This is indicated by the two 6-hour accumulation images shown in Figure 5.1.

The scatter plots of the Block Kriged raingauge and radar accumulations are shown in Figure 5.2. For the first 6-hour period (12:00 to 18:00) there was a low correspondence between the raingauge and radar estimates with an r^2 value of 0.07 being returned; for the second 6-hour accumulation period (18:00 to 24:00) a higher correspondence between the two estimates occurred with an r^2 value of 0.88 returned. This is also indicated in Figure 5.2 where the CDFs of the raingauge and radar accumulations are plotted. The CDFs for the first 6-hour period is visibly different while for the second 6-hour period the CDFs are quite similar. This indicates that a greater correspondence exists between Block Kriged raingauge and radar accumulations for the more intense rainfall periods and a lower correspondence for the less intense stratiform rainfall periods.

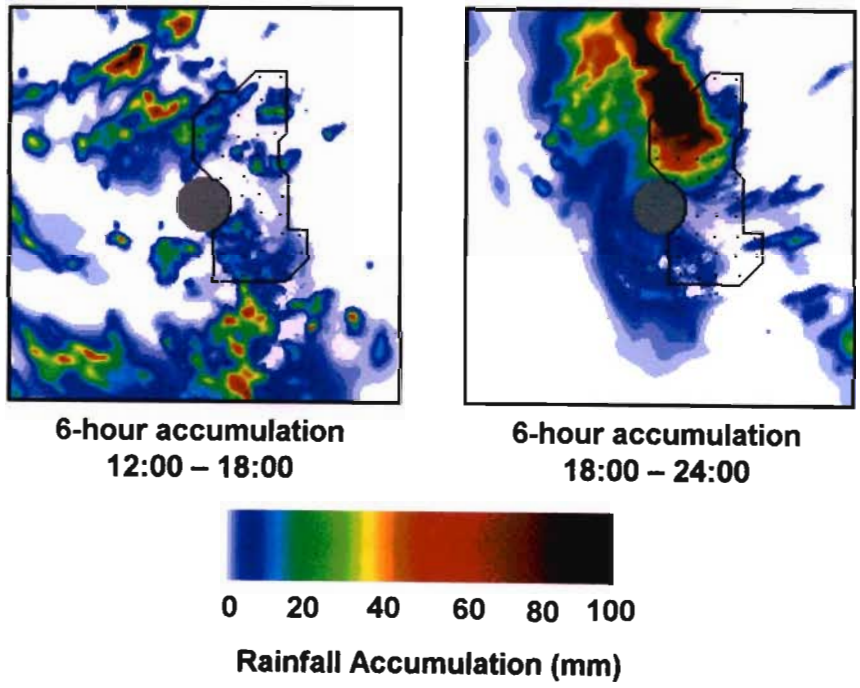


Figure 5.1: 6-hour accumulation's of rainfall for the 24 January 1996 for the times 12:00 to 18:00 and 18:00 to 24:00. An outline of the test catchment is indicated on the image as well as the raingauge locations. Sizes of images are 200 km square.

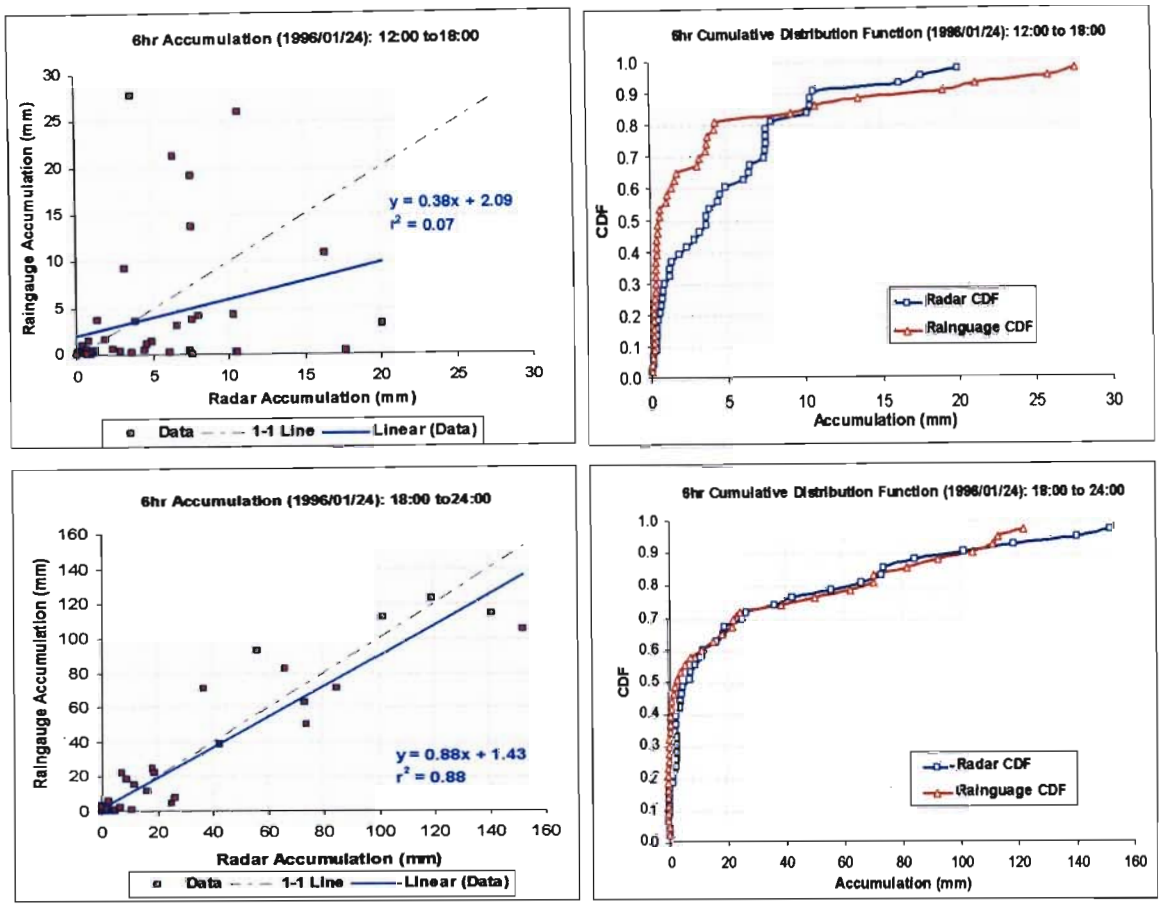
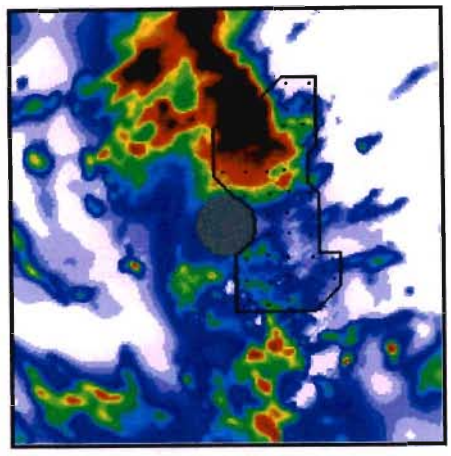


Figure 5.2: Scatter plots of Block Kriged raingauge and radar accumulation estimates over two 6-hour periods with the corresponding CDFs for the 24 January 1996.

12-HOUR ACCUMULATION

The 12-hour accumulation indicates that during the course of the 12-hour period the majority of the catchment experienced rainfall with the most intense rainfall occurring in the Northern region of the catchment. The rainfall accumulation values from 18:00 to 24:00 are indicated in Figure 5.3 with the outline of the test catchment and the location of the raingauges also indicated on the figure.



12-hour accumulation
18:00 to 24:00

Figure 5.3: 12-hour accumulation of rainfall for the 24 January 1996 for the period 12:00 to 24:00. An outline of the test catchment is indicated on the image as well as the raingauge locations. Size of image is 200 km square.

The scatter plot and CDF for the raingauge and radar accumulation estimates indicate a high correspondence between the two estimates. An r^2 value of 0.86 was returned for the scatter plot and the CDF plot also indicates a close correspondence between the raingauge and radar distributions. Table 5.1 shows that the all the mean values and standard deviations are not dissimilar. The distributions, except for the 12:00 to 18:00, periods are also not dissimilar. The highest pair of radar observations skew the linear fit and is probably due to an under estimation of the convective rainfall by the two raingauges (Wilson and Brandes, 1979).

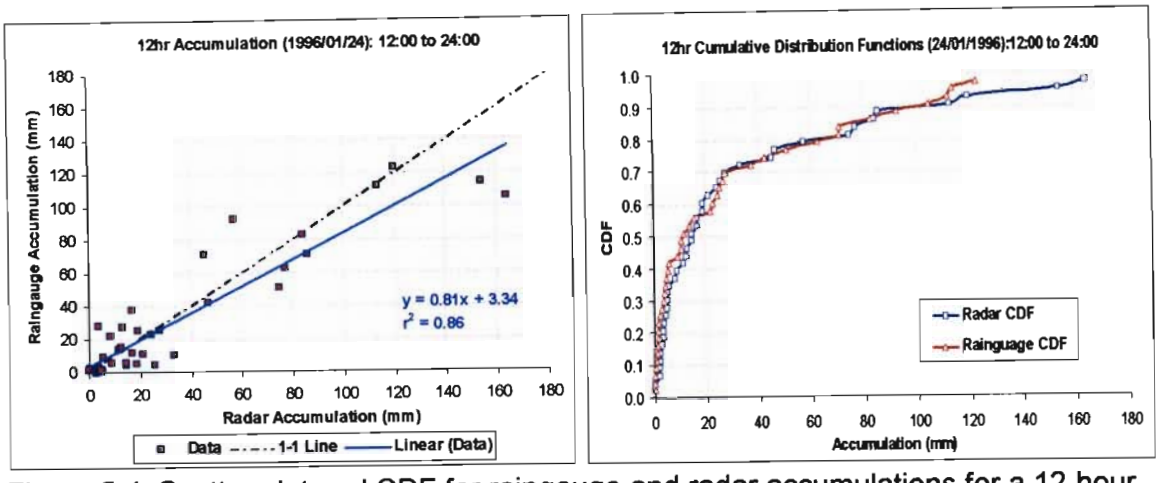


Figure 5.4: Scatter plot and CDF for raingauge and radar accumulations for a 12-hour accumulation period for the 24 January 1996. The two plots indicate a high correspondence between the raingauge and radar accumulations.

Table 5.1: Comparison of raingauge and radar accumulation values over 6 and 12-hour accumulation periods for the 24 January 1996.

	6 hr accumulation (12:00 to 18:00)		**6 hr accumulation (18:00 to 24:00)**		**12hr accumulation (12:00 to 24:00)**	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	4.6	5.1	27.2	40.8	30.3	41.3
Raingauge	4.0	7.1	25.4	38.3	29.4	36.8
Accept / Reject H_0	Accept	Accept	Accept	Accept	Accept	Accept
r^2	0.08		0.88		0.86	
K-S TEST	Reject		Accept		Accept	

Indicated in Figure 5.5 are the computation times to infill all the missing data contained in a radar volume scan including the estimate at ground level. The average time to provide these extrapolations is 53 seconds with the minimum computation time being 35 seconds and the maximum 87 seconds (1 minute 27 seconds). These times are well below the 5 minute intervals at which radar volume scan are received in South Africa. All computations for the rainfall estimations at ground level were done on a Pentium(R) 4 with a 2.40GHz CPU and 512MB of RAM.

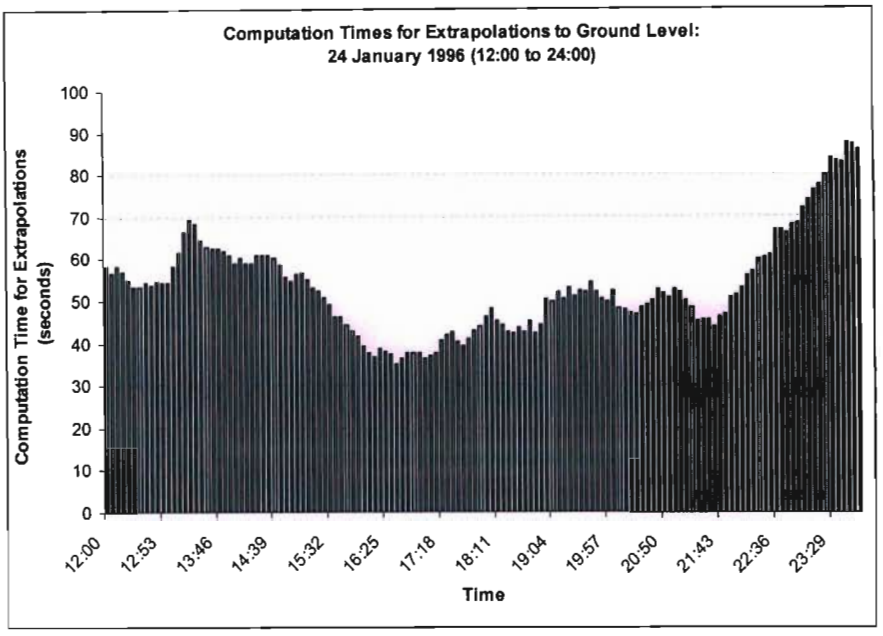


Figure 5.5: Computation time (seconds) for extrapolations to ground level for the 24 January 1996 from the time 12:00 to 24:00.

5.3.2 Results Test Event 2 - 13 February 1996

The rain event on the 13 February 1996 consisted of a combination of convective and stratiform rain. The rainfall initially consisted of light stratiform rainfall but in the evening also experienced intense convective rainfall. In the 24-hour period 19% of the images were identified as containing bright band and corrected. The 6-hour accumulation periods for the 13 February 1996 are depicted in Figure 5.6 illustrating the combination of convective and stratiform rainfall that was extrapolated to ground level throughout the day. The exclusion of data directly above the radar applies in this instance as well. Two of the gauges were therefore excluded from the data analysis.

6-HOUR ACCUMULATIONS

As indicated in Figure 5.6 during the first 6-hour period, 00:00 to 06:00, very light stratiform rainfall was recorded that was considered to be mostly insignificant. During the second 6-hour period, 06:00 to12:00, the rainfall intensity increased but still was predominantly stratiform and sparse. During the remaining 12-hours the rainfall increased in intensity with convective rainfall being recorded in the Northern segment of the catchment.

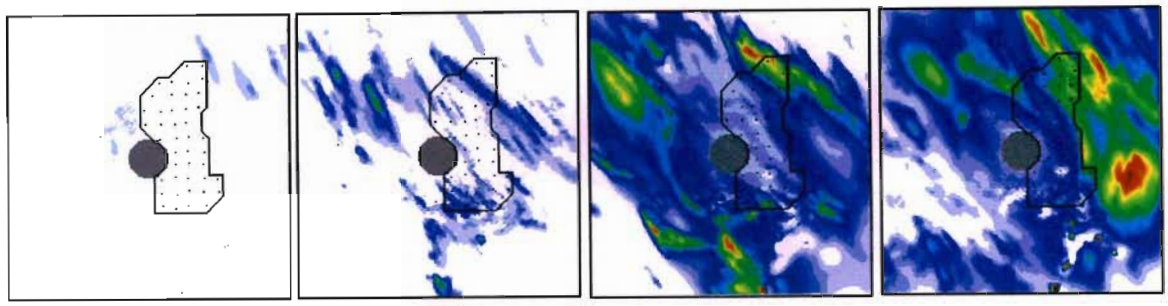


Figure 5.6: 6-hour accumulation images from the 13 February 1996. During the first 12-hour period the rainfall was generally sparse and stratiform; during the second 12-hour period the rainfall was predominantly convective. Sizes of images are 200 km square.

The scatter plots, indicated in Figure 5.7, show initially a low correspondence between Block Kriged raingauge and radar accumulation values, with an r^2 value of 0.18 being computed. As the rainfall intensified during the course of the day a stronger relationship between the raingauge and radar accumulation values developed, as the r^2 value increased to 0.58. The corresponding CDFs in Figure 5.7 also indicate an initially poor correspondence between the Block Kriged raingauge and radar distributions as in the 6-hour accumulation period from 06:00 to 12:00. The remaining 6-hour accumulation periods show a better correspondence between the Block Kriged rainaguge and radar distributions.

Table 5.2 provides a comparison of the statistics for the 6-hour accumulation periods described where all the mean values are considered not to be dissimilar except for one during the 12:00 to 18:00 period. All the standard deviations were considered to be not dissimilar. All the distributions, except for the 06:00 to 12:00, periods are not dissimilar.

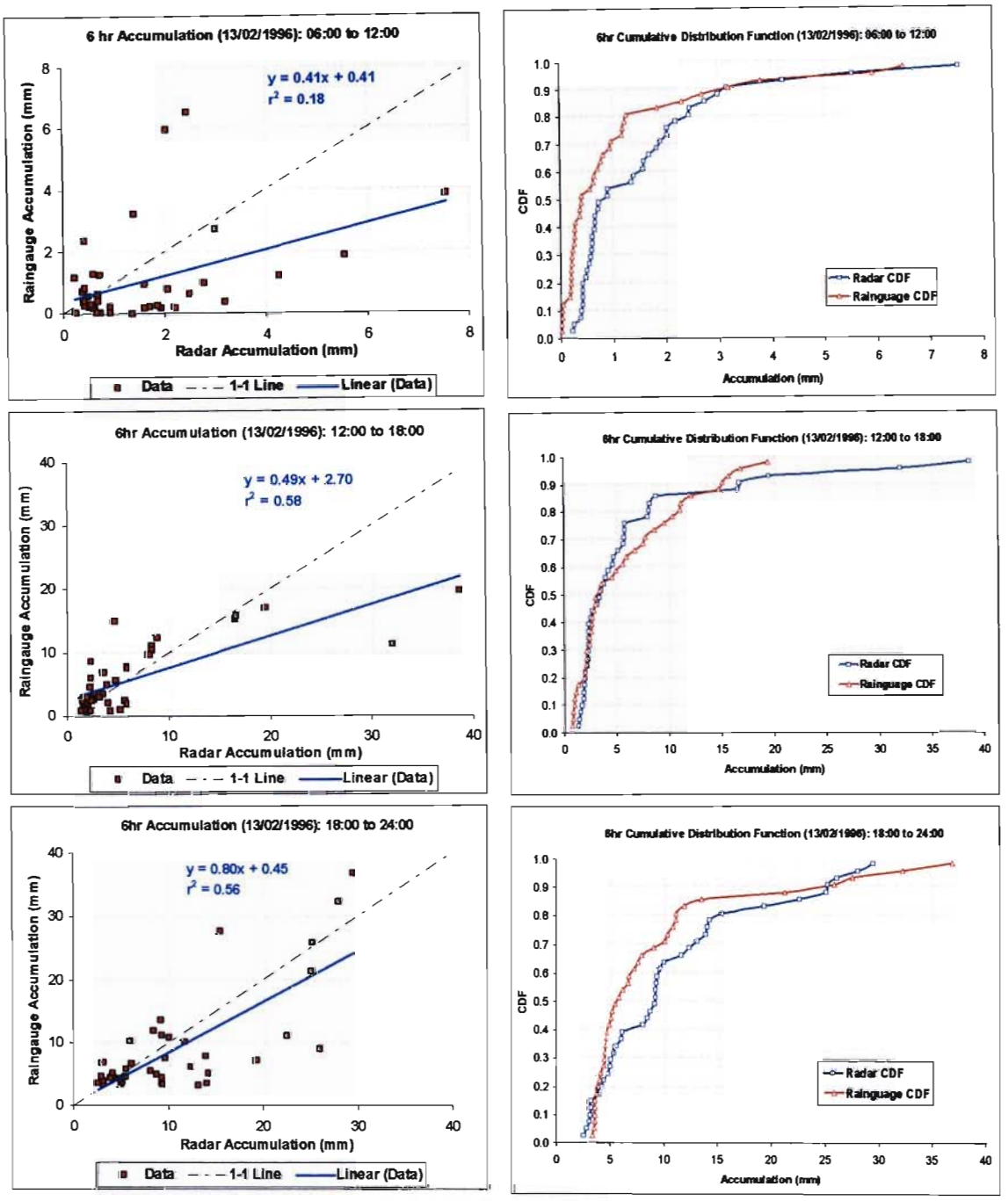


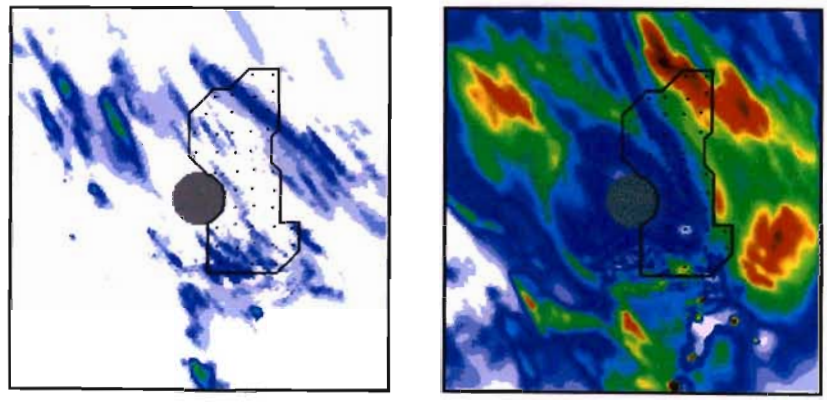
Figure 5.7: Scatter plots of raingauge and radar accumulation values over three 6-hour accumulation periods with the corresponding CDFs. Also indicated are the r^2 value and the linear relationship between the Block Kriged raingauge and radar accumulation values.

Table 5.2: Comparison of raingauge and radar accumulation values over three 6 -hour periods for the 13 February 1996.

	6hr accumulation (06:00 to 12:00)		6hr accumulation (12:00 to 18:00)		6hr accumulation (18:00 to 24:00)	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	1.52	1.53	6.29	7.99	10.64	7.71
Raingauge	1.04	1.49	5.81	5.18	8.99	8.25
Accept / Reject H_0	Accept	Accept	Reject	Accept	Accept	Accept
r^2	0.18		0.58		0.56	
K-S TEST	Reject		Accept		Accept	

12-HOUR ACCUMULATIONS

The 12-hour accumulations as illustrated in Figure 5.8 also indicate that sparse stratiform rainfall occurred in the hours 00:00 to 12:00 while in the second 12-hour period, 12:00 to 24:00, the rainfall was more widespread and consisted of a combination of convective and stratiform rainfall.



12-hour accumulation 00:00 to 12:00

12-hour accumulation 12:00 to 24:00

Figure 5.8: Accumulation images for 12-hour periods for the 13 February 1996. The accumulations for 00:00 to 12:00 indicate sparse stratiform rainfall while the second 12-hour period (12:00 to 24:00) indicates widespread stratiform and convective rainfall. Sizes of images are 200 km square.

The scatter plots in Figure 5.9 indicate an initially poor relationship between the raingauge and radar accumulations for the first 12-hour accumulation an r^2 value of 0.15 was

returned while for the second period an r^2 value of 0.76 was returned. Once again for the light stratiform rainfall (where the accumulations were less than 8 mm) the radar accumulations at ground level matched poorly with the raingauge accumulations. With the more intense rainfall (in excess of 20 mm) a far greater correspondence existed with the raingauges once again underestimating the two highest accumulation values. The CDFs also indicate a poor correspondence of the distributions for the first 12-hours and a closer correspondence for the more intense rainfall in the second 12-hour period.

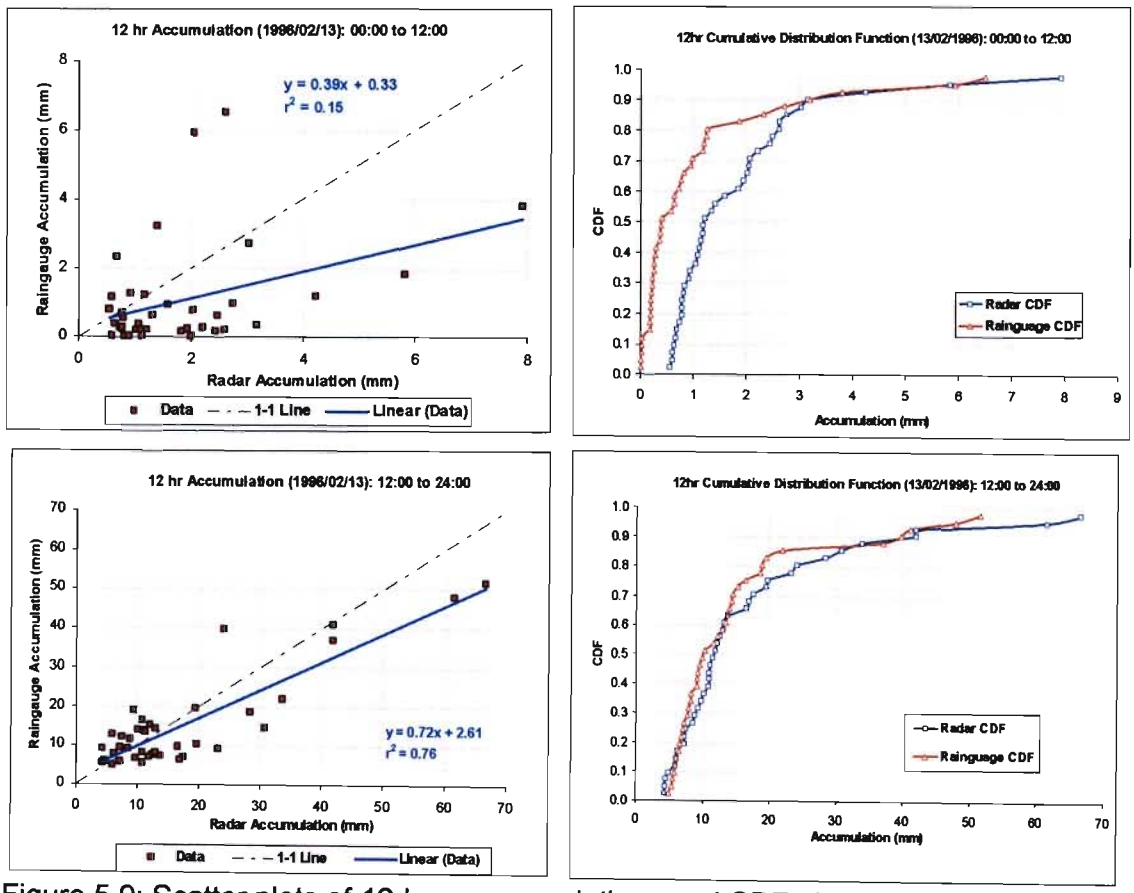
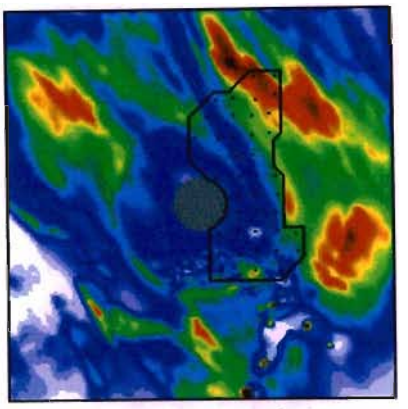


Figure 5.9: Scatter plots of 12-hour accumulations and CDFs for raingauge and radar accumulations in the time periods 00:00 to12:00 and 12:00 to 24:00

24-HOUR ACCUMULATION

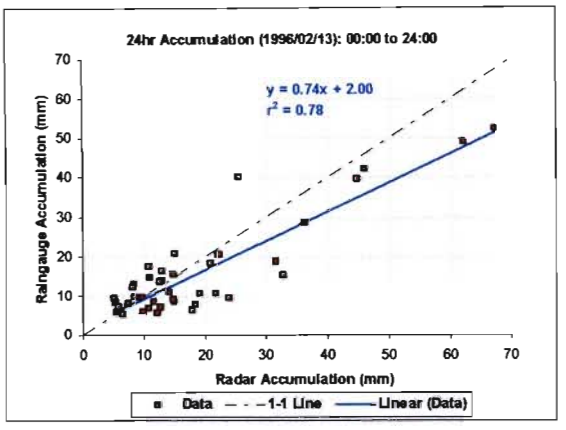
The 24-hour accumulation values, illustrated in Figure 5.10, indicate rainfall was recorded over the entire catchment with convective rainfall occurring in the Northern half of the catchment. Figure 5.11 indicates the scatter plot between raingauge and radar accumulation where a r^2 value of 0.78 was returned indicating a high correspondence

between the raingauge and radar accumulations. The CDF also indicates a close correspondence between the two estimates.



24-hour accumulation
00:00 to 24:00

Figure 5.10: Accumulation images for a 24-hour period for the 13 February 1996. An outline of the test catchment is indicated on the image as well as the raingauge locations. Size of image is 200 km square.



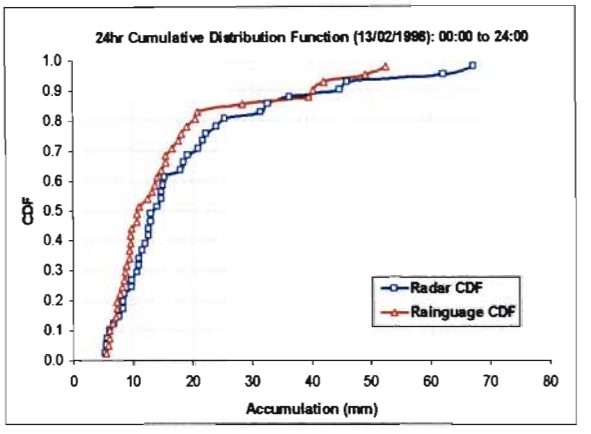


Figure 5.11: Scatter plot and CDF for a 24-hour accumulation period for the 13 February 1996. A strong correspondence exists between the radar and raingauge accumulations, especially for the high intensity rainfall.

Table 5.3 contains the statistical summary of the radar and raingauge accumulatons over two 12-hour accumulation periods and one 24-hour accumulation period. The results indicate that all the mean values over the accumulation periods are not dissimilar. All the standard deviations can also considered to be not dissimilar, except for one during the

12:00 to 24:00 periods. All the distribution from the 12-hour and 24-hour period can also be considered to be not dissimilar.

Table 5.3: Comparison of raingauge and radar accumulation values over 24 and two 12-hour periods for the 13 February 1996.

	12hr accumulation (00:00 to 12:00)		**12hr accumulation (12:00 to 24:00)**		**24hr accumulation (00:00 to 24:00)**	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	1.1	1.4	17.0	14.5	18.1	14.7
Raingauge	1.0	1.5	14.8	11.9	15.8	12.2
Accept / Reject H_0	Accept	Accept	Accept	Reject	Accept	Accept
r^2	0.14		0.76		0.78	
K-S TEST	Accept		Accept		Accept	

Indicated in Figure 5.12 are the computation times to infill all the missing data contained in the radar volume scan including the estimate at ground level. The average time to provide these extrapolations is 53 seconds with the minimum computation time being 14 seconds and the maximum 117 seconds (1 minute 57 seconds). These times are well below the 5 minute intervals at which radar volume scan are received in South Africa.

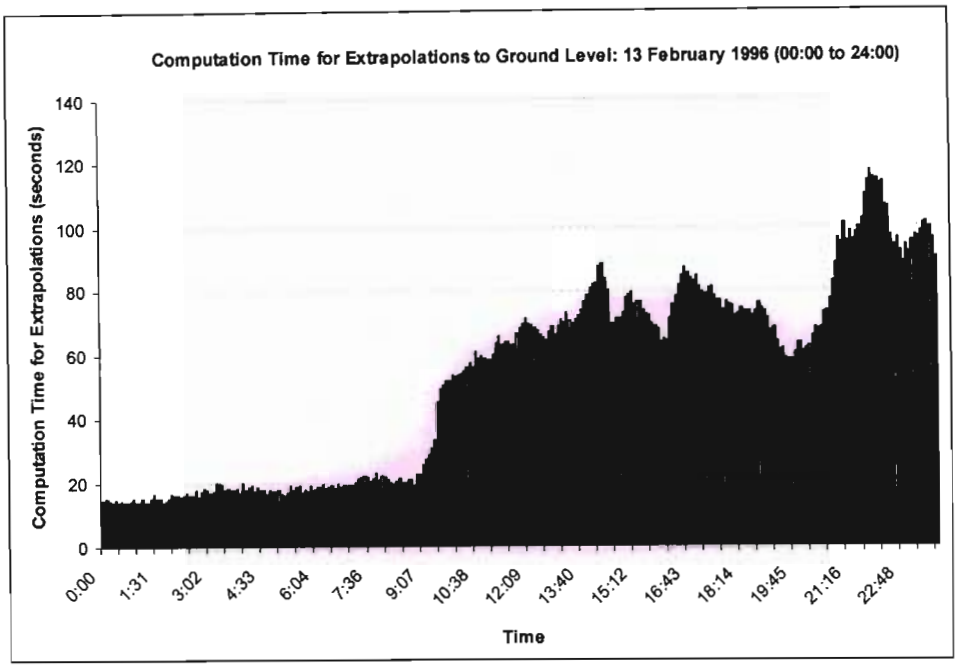


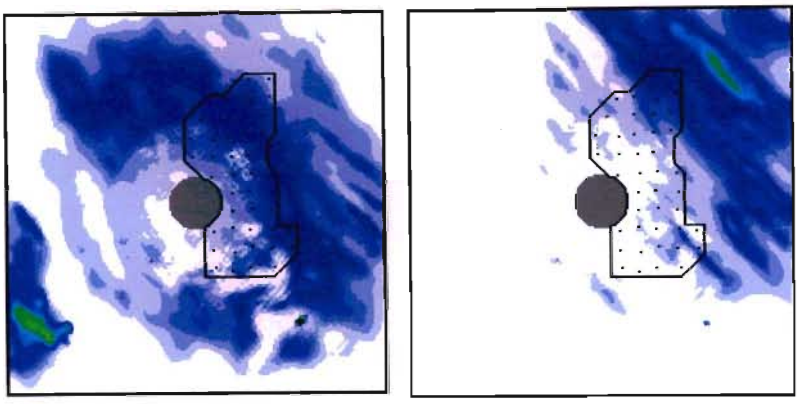
Figure 5.12: Computation time (seconds) for extrapolations to ground level for the 13 February 1996 from the time 00:00 to 24:00.

5.3.3 Results Test Event 3 - 27 January 1996

The rain event on the 27 January 1996 consisted of light widespread light stratiform rain throughout the time period 12:00 to 24:00. In the 12-hour rain period 14% of the images were identified as containing bright band and corrected. The exclusion of data directly above the radar applies once again; two of the gauges were therefore excluded from the data analysis.

6-HOUR ACCUMULATIONS

Two 6-hour accumulations were computed during this rain event from the time periods 12:00 to 18:00 and 18:00 to 24:00. The accumulation images for these time periods are illustrated in Figure 5.13 providing an indication of the light stratiform rainfall that was experienced over the 6-hour periods. During the first 6-hour period the majority of the rainfall occurred in the Eastern segment of the catchment and in the second 6-hour period in the Northern segment of the catchment.



6-hour accumulation 12:00 to 18:00

6-hour accumulation 18:00 to 24:00

Figure 5.13: 6-hour accumulation images for the time periods 12:00 to 18:00 and 18:00 to 24:00 for the 27 January 1996. Predominantly light stratiform rainfall was experienced over the catchment with the majority of the rainfall occurring in the Northern half of the catchment. Sizes of images are 200 km square.

The scatter plots of the raingauge and radar accumulation are contained in Figure 5.14 which indicates, in the period 12:00 to 18:00, a poor correspondence between the raingauges and radar accumulations with a r^2 value of 0.21 being returned and the CDFs indicating distinctly different distributions being returned. In the second accumulation period from 18:00 to 24:00 the rainfall and radar accumulations were values were far lower (below 10mm) than the first 6-hour accumulation period. In the accumulation period from 18:00 to 24:00 a high r^2 value was returned of 0.78 indicating a close correspondence between the raingauge and radar accumulation estimates, the CDFs returned also indicate a far higher correspondence between the two estimates.

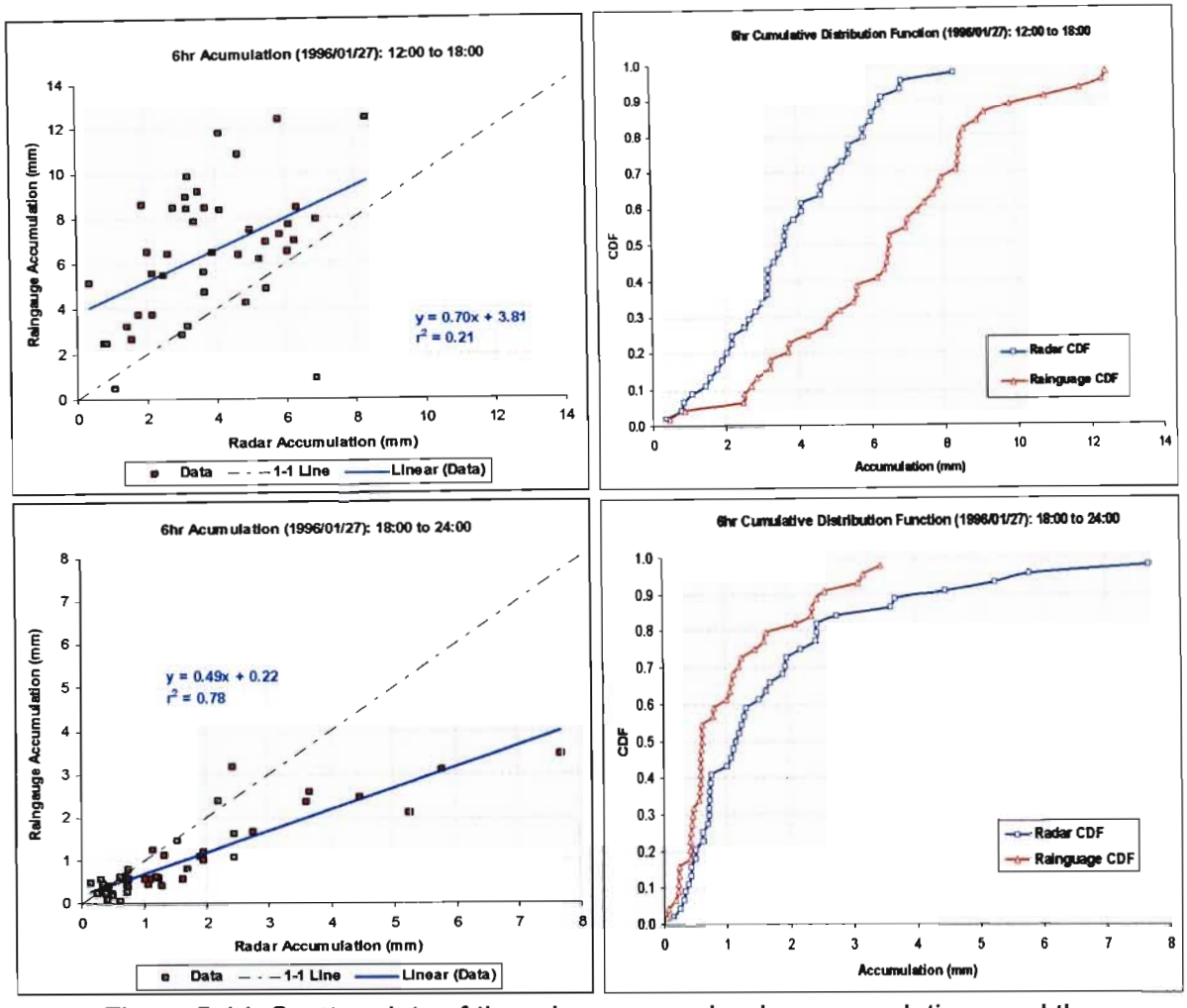
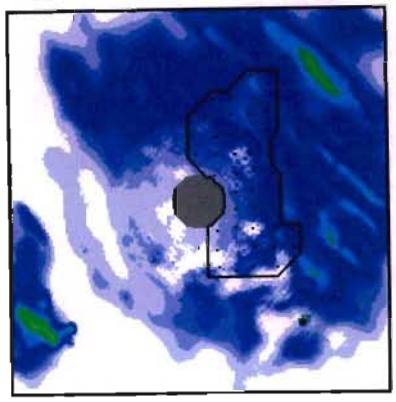


Figure 5.14: Scatter plots of the raingauge and radar accumulations and the corresponding CDFs for the 6-hour accumulation periods for the 27 January 1996.

12-HOUR ACCUMULATION

Figure 5.15 indicates the 12-hour accumulation image. The image indicates that stratiform rainfall was experienced over the entire catchment area over the accumulation period from 12:00 to 24:00. The majority of the stratiform rainfall occurred in the Northern half of the catchment area as indicated in Figure 5.15. The scatter plot of the raingauge and radar accumulation indicate a low correspondence between the radar and raingauge estimates with an r^2 value of 0.28 being returned with the radar indicating an underestimate of the rainfall accumulations at ground level. The CDFs also indicate the distributions returned are dissimilar.



12-hour accumulation
12:00 to 24:00

Figure 5.15: 12-hour accumulation image for the time period 12:00 to 24:00. The accumulations indicate light stratiform rainfall was experienced over the entire catchment with the majority of the rainfall occurring in the North Eastern segment of the catchment. Size of the image is 200 km square.

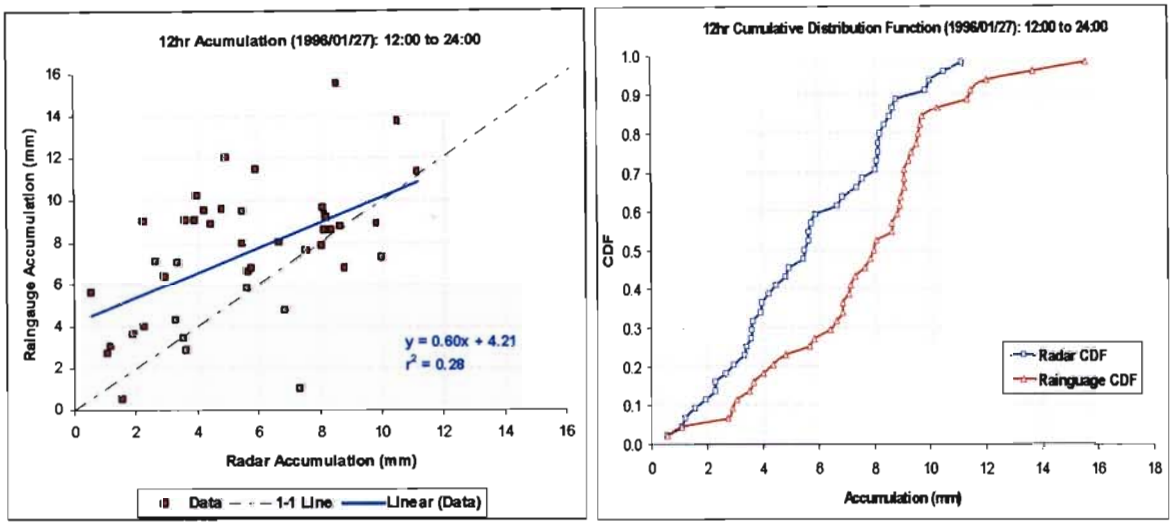


Figure 5.16: Scatter plot and Cumulative Distribution Functions (CDFs) for the 12-hour accumulation period from 12:00 to 24:00 for the 27 January 1996.

Table 5.4 summarises the statistical comparison between the raingauge and radar accumulations. All the mean values were considered to be dissimilar, while all the standard deviations were considered not to be dissimilar except for one during the 12:00 to 18:00 period. All the distributions were considered to be dissimilar. There was generally a low correspondence between the raingauge and radar estimates probably due to the

fact that the rainfall was light and straiform in nature and therefore more likely to be affected by atmospheric phenomena before reaching ground level such as wind drift.

Table 5.4: Comparison of raingauge and radar accumulation values over 12 and two 6-hour periods for the 27 January 1996.

	6hr accumulation (12:00 to 18:00)		**6hr accumulation (18:00 to 24:00)**		**12hr accumulation (12:00 to 24:00)**	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	3.8	1.9	1.7	1.6	5.5	2.8
Raingauge	6.5	2.9	1.0	0.9	7.5	3.2
Accept / Reject H_0	Reject	Reject	Reject	Accept	Reject	Accept
r^2	0.21		0.78		0.28	
K-S TEST	Reject		Reject		Reject	

Illustrated in Figure 5.17 are the computation times to infill all the missing data contained in a radar volume scan including the estimate at ground level. The average time to provide these extrapolations is 32 seconds with the minimum computation time being 13 seconds and the maximum 48 seconds. The computation times for the extrapolations to ground level are particular fast in this instance due to the fact that the rain is entirely stratiform and the quicker LU decomposition is utilised in the Cascade Kriging algorithm. These times are well below the 5 minute intervals at which radar volume scan are received in South Africa.

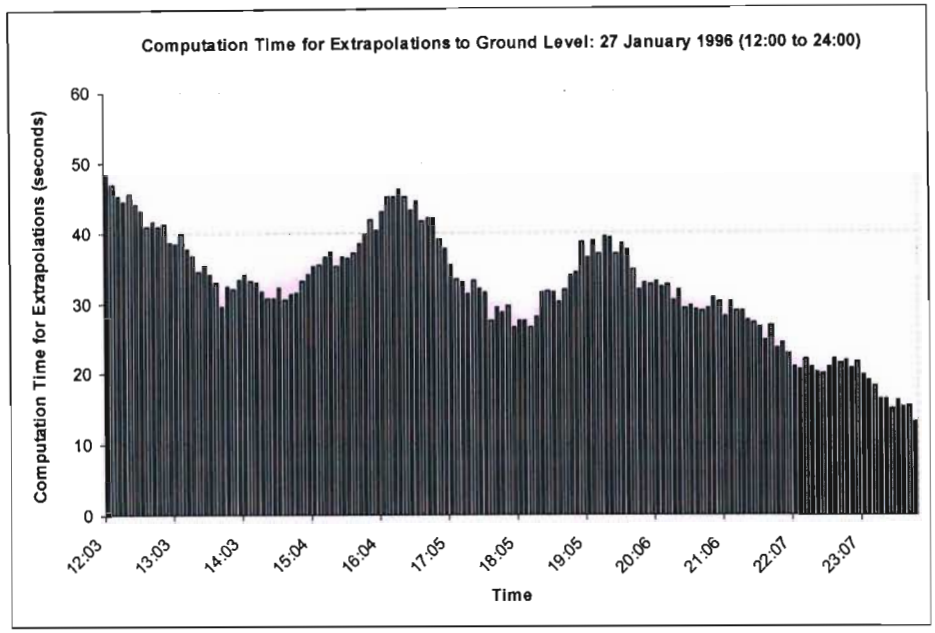


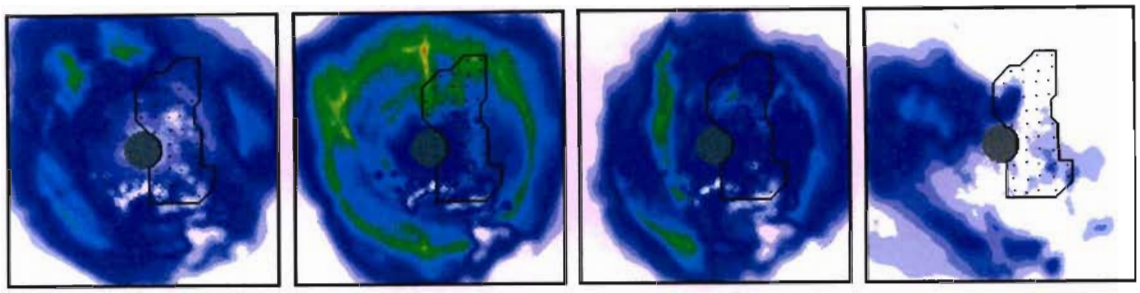
Figure 5.17: Computation time (seconds) for extrapolations to ground level for the 27 January 1996 from the time 12:00 to 24:00.

5.3.4 Results Test Event 4 - 17 December 1995

The rainfall event that occurred over the Liebenbergsvlei catchment on the 17 December 1995 was also predominantly stratiform with isolated convective rainfall occurring at sporadic intervals throughout the day. The rain event was affected by bright band for approximately 65% of the images throughout the day and corrected.

6-HOUR ACCUMULATIONS

The 6-hour accumulation images throughout the day are illustrated in Figure 5.18 indicating that in the first 6-hour period (00:00 to 06:00) was predominantly light but widespread over the catchment. The rainfall intensified over the second 6-hour period (06:00 to 12:00) and thereafter gradually decreased in intensity throughout the remaining two 6-hour periods. In Figure 5.18 it is significant that the base-scan of the radar misses the rain beyond approximately 75 km range as indicated in Figure 5.18, but the test area lies in this range.



6-hour accumulation 00:00 to 06:00 | **6-hour accumulation 06:00 to 12:00** | **6-hour accumulation 12:00 to 18:00** | **6-hour accumulation 18:00 to 24:00**

Figure 5.18: 6-hour accumulation images throughout the 17 December 1995 indicating the stratiform rainfall experienced throughout the day. Sizes of images are 200 km square.

The scatter plots and CDF's are indicated in Figure 5.19 for the 6-hour accumulation periods throughout the 17 December 1995. Generally a low correspondence exists between the raingauge and radar accumulation values with the r^2 value returned ranging between 0.22 and 0.39. The radar accumulations tended to be lower than the raingauge accumulation estimates indicating an underestimation of the rainfall by the radar.

Table 5.5 indicates a summary of the statistics computed over each of the four 6-hour periods. All the mean values and standard deviations were considered to be dissimilar except during the 06:00 and 12:00 period. All the distributions returned were also considered to be dissimilar except for the distributions returned for the 06:00 to 12:00 periods.

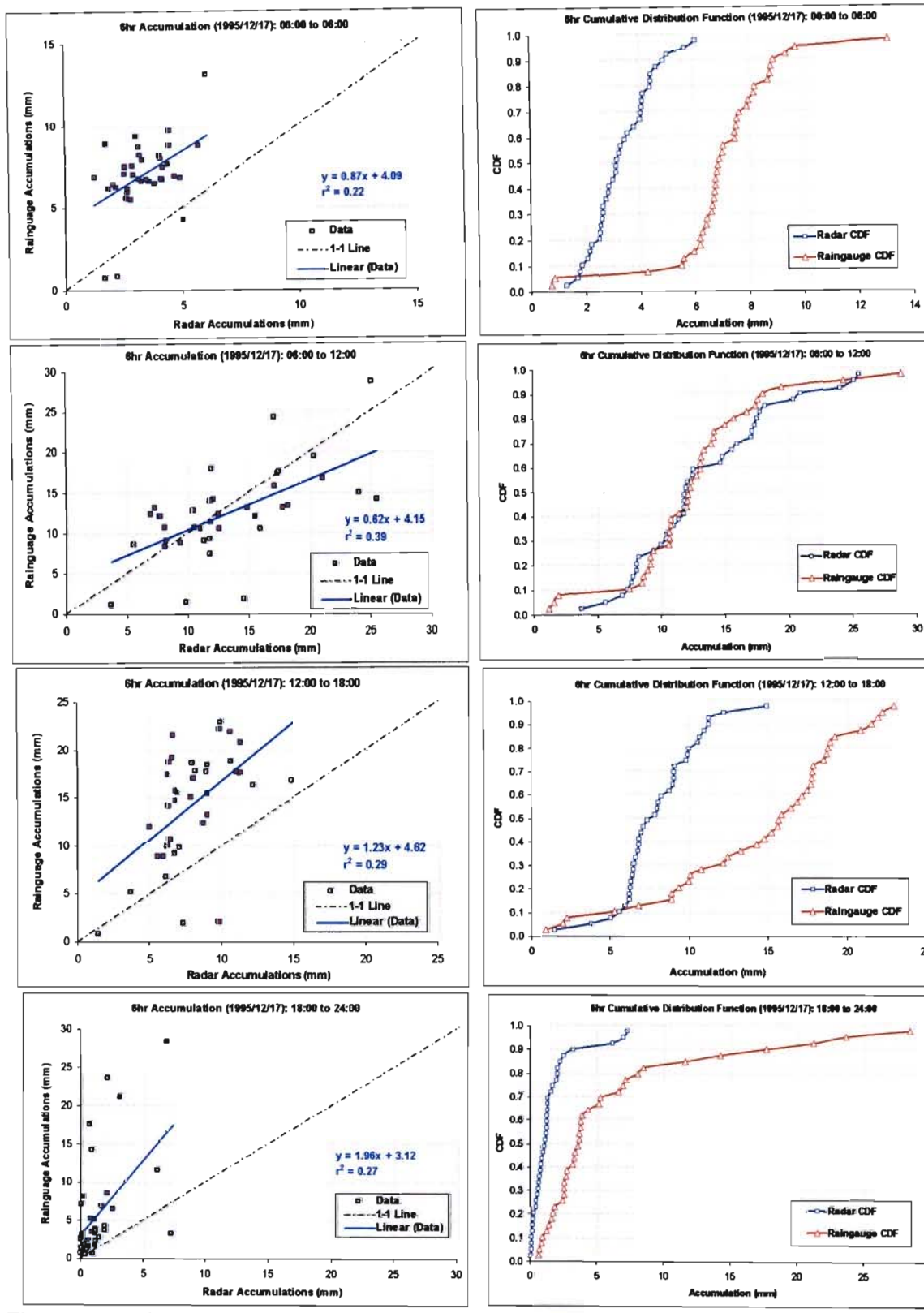


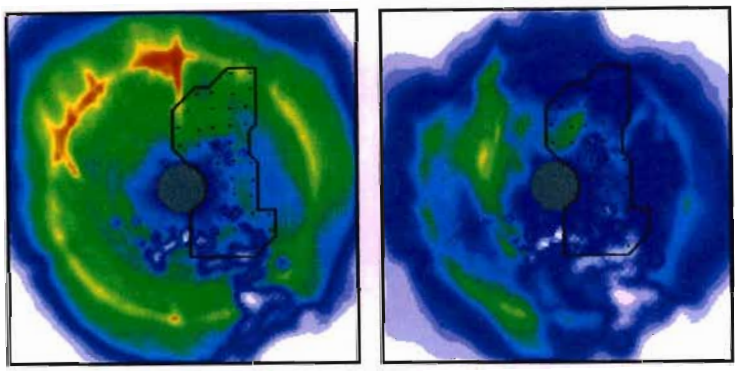
Figure 5.19: Scatter plots and CDFs for the 6-hour accumulation periods throughout the 17 December 1995.

Table 5.5: Comparison of raingauge and radar accumulation values over four 6-hour accumulation periods for the 17 December 1995.

	6hr accumulation (00:00 to 06:00)		6hr accumulation (06:00 to 12:00)		6hr accumulation (12:00 to 18:00)		6hr accumulation (18:00 to 24:00)	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	3.3	1.1	13.2	5.4	7.9	2.5	1.4	1.8
Raingauge	7.0	2.1	12.4	5.4	14.3	5.8	5.9	6.6
Accept / Reject H_0	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject
r^2	0.22		0.39		0.29		0.29	
K-S TEST	Reject		Accept		Reject		Reject	

12-HOUR ACCUMULATIONS

The accumulation images for the 12-hour accumulation periods on the 17 December 1995 illustrating the widespread stratiform rainfall over the two 12-hour accumulation periods throughout the day.



12-hour accumulation 00:00 to 12:00

12-hour accumulation 12:00 to 24:00

Figure 5.20: 12-hour accumulation periods for the 17 December 1995. The accumulation images indicate the widespread stratiform rainfall over the test catchment area over the 12-hour accumulation periods. Sizes of images are 200 km square.

Indicated in Figure 5.21 are the scatter plots and CDFs for the 12-hour accumulation periods for the 17 December 1995. An r^2 value of 0.36 was returned for the 00:00 to12:00 period and an r^2 value of 0.25 for the second 12-hour accumulation period. For the 12:00

to 24:00 period the radar accumulation values badly underestimated the rainfall accumulation values.

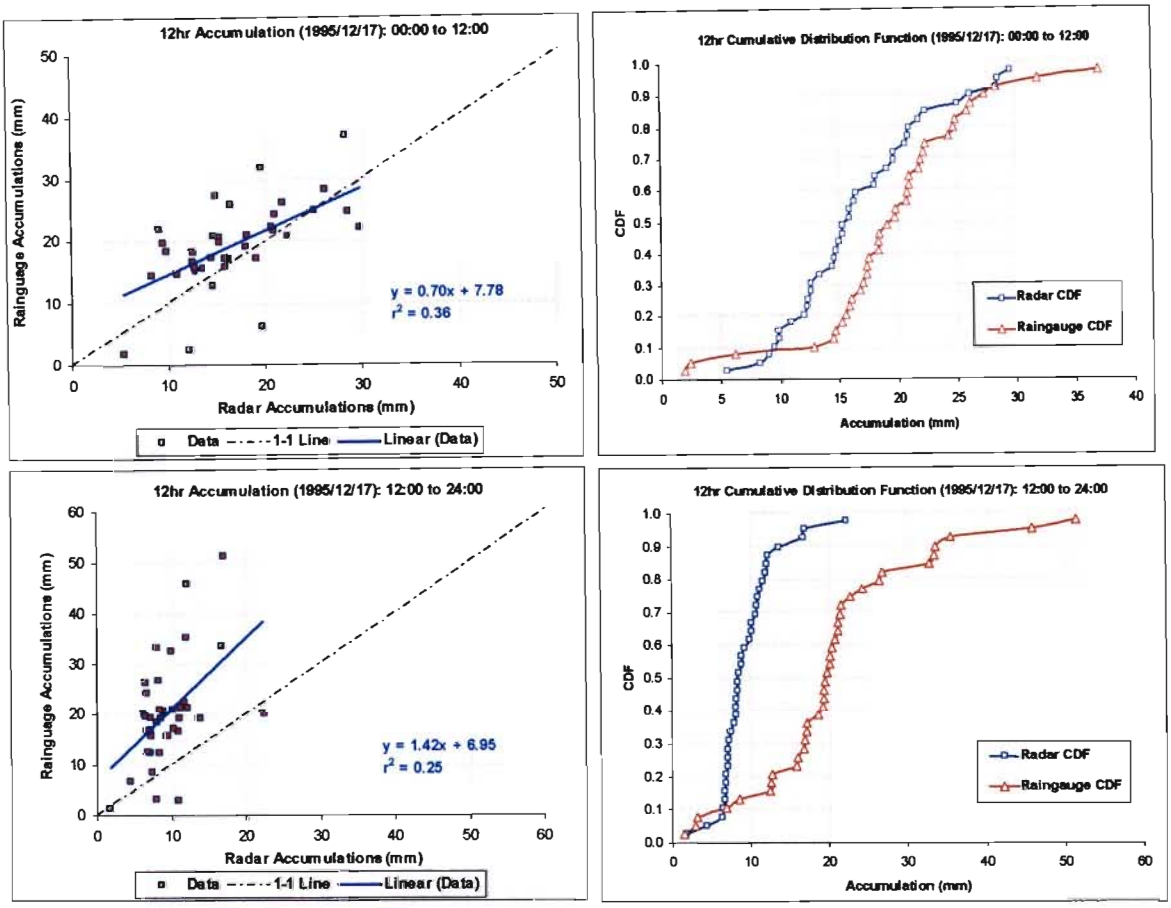
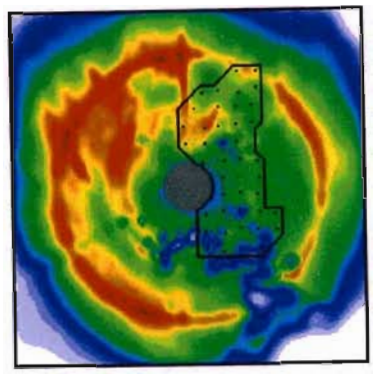


Figure 5.21: Scatter plots and CDF's for the two 12-hour accumulation periods for the 17 December 1995.

24-HOUR ACCUMULATION

The 24-hour accumulation image is indicated by Figure 5.22 illustrating the stratiform rainfall experienced over the catchment area throughout the 24-hour period. Figure 5.23 illustrates the scatter plots and CDFs for the 24-hour accumulation period, the scatter plot indicates a low correspondence between the radar and raingauge accumulation values with an r^2 value of 0.43 being returned. The scatter plot indicates that the radar underestimates the rainfall at ground level.



24-hour accumulation
00:00 to 24:00

Figure 5.22: 24-hour accumulation for the 17 December 1995. Size of image is 200 km square.

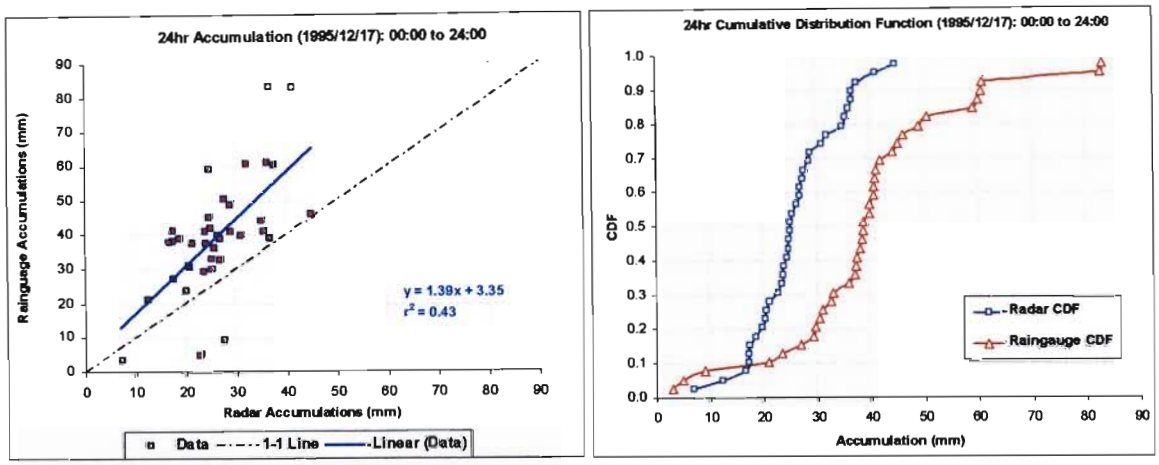


Figure 5.23: Scatter plots and CDFs for the 24-hour accumulation for the 17 December 1995.

Table 5.6 provides a summary of the statistics returned for the two 12-hour and 24-hour accumulation period for the 17 December 1995. For these accumulation periods all the mean values returned were considered to be dissimilar, all the standard deviations returned were also considered to be dissimilar except for one during the 00:00 to 12:00 period. All the distributions returned for this accumulation period were considered to be dissimilar. The poor correspondence between the radar and raingauge accumulations can be attributed to the fact that the rainfall is of low intensity and has severe bright band contamination. Both of theses factors combined decrease the likelihood that an accurate estimate of the rainfall at ground level can be obtained.

Table 5.6: Comparison of raingauge and radar accumulation values over two 12-hour and one 24-hour accumulation period for the 17 December 1995.

	12hr accumulation (00:00 to 12:00)		12hr accumulation (12:00 to 24:00)		24hr accumulation (00:00 to 24:00)	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Radar	16.6	5.9	9.4	3.7	26.2	7.9
Raingauge	19.4	6.9	20.3	10.4	39.6	16.8
Accept / Reject H_0	Reject	Accept	Reject	Reject	Reject	Reject
r^2	0.36		0.25		0.43	
K-S TEST	Reject		Reject		Reject	

Illustrated in Figure 5.24 are the computation times to infill all the missing data contained in a radar volume scan including the estimate at ground level. The average time to provide these extrapolations is 58 seconds with the minimum computation time being 24 seconds and the maximum 105 seconds (1 minute 45 seconds). Once again these times are well below the 5 minute intervals which are required for real time application.

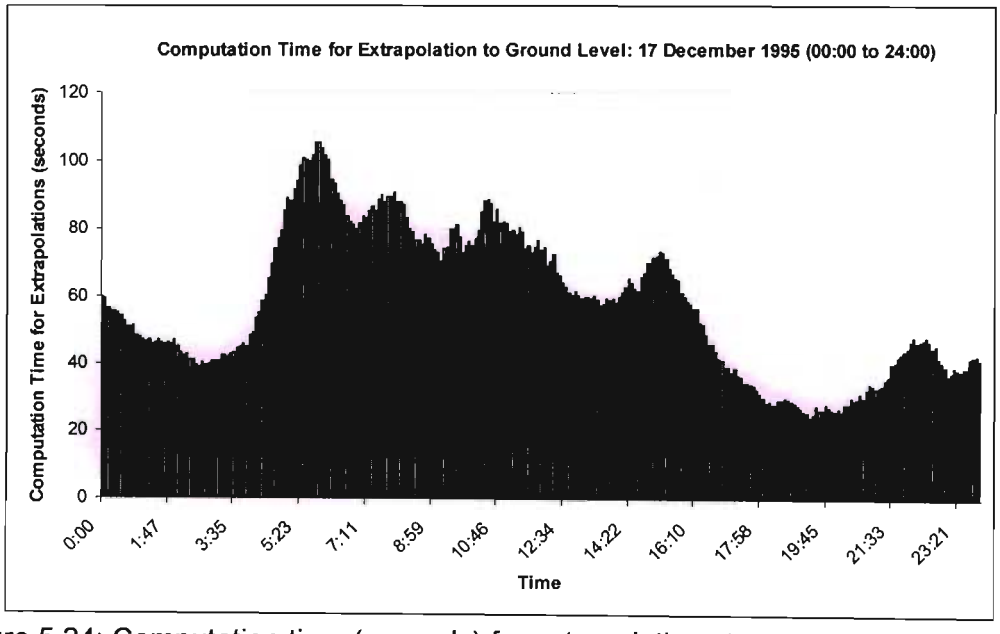


Figure 5.24: Computation time (seconds) for extrapolations to ground level for the 17 December 1995 from the time 00:00 to 24:00.

5.2.6 Brief Comparison to Alternative Methodologies

A brief overview of how the Cascade Kriging algorithm compared to two simpler methodologies to obtain the rainrate at ground level will be reviewed.

- Nearest Pixel - The first method simply uses the lowest (altitude) pixels reflectivity in a vertical profile as the estimate at ground level.
- Average Profile - The second method is currently employed by SAWS in the Daily Rainfall Mapping Project over South Africa to provide an estimate the rainfall at ground level. The method calculates the average of all the non-zero reflectivity values contained in a vertical column above the pixel, the average is then utilised as the estimate of the rainfall at ground level (Visser, 2003).

The three methods: Cascade Kriging, Nearest Pixel and Profile Average were compared over the four rain events reviewed in the previous sub-sections over the 6-hour, 12-hour and 24-hour accumulation periods. For each of the methods employed to obtain an estimate of the rainfall at ground level over each of the accumulation periods the MSSE was computed. The MSSE was determined by computing the sum of the difference squared between the Block Kriged raingauge accumulations and radar accumulations and dividing by the number of raingauges utilised.

The MSSE for each of the four rainfall events over different accumulation periods is depicted in Figure 25. For the rainfall event on the 24 January 1996 (top left bar chart) the Cascade Kriging algorithm did significantly better than the Profile Average method, whereas in comparison to the Nearest Pixel method the MSSEs returned were quite similar. For the rainfall event on the 13 February 1996 (top right bar chart) the Cascade Kriging algorithm returned significantly improved estimates over the Profile Average and Nearest Pixel methods, which is indicated by the lower MSSE returned over all of the accumulation periods. For the rainfall event on the 27 January 1996 (bottom left bar chart) the Cascade Kriging technique only provided improved estimates at ground level in comparison to the other techniques in one of the accumulation periods, 12:00 to 18:00, with however very light rainfall occurring in that rain event. For the rainfall event on the 17 December 1995 (bottom right bar chart) the Cascade Kriging algorithm provided better estimates in terms of the MSSE over the other techniques in all of the accumulation periods.

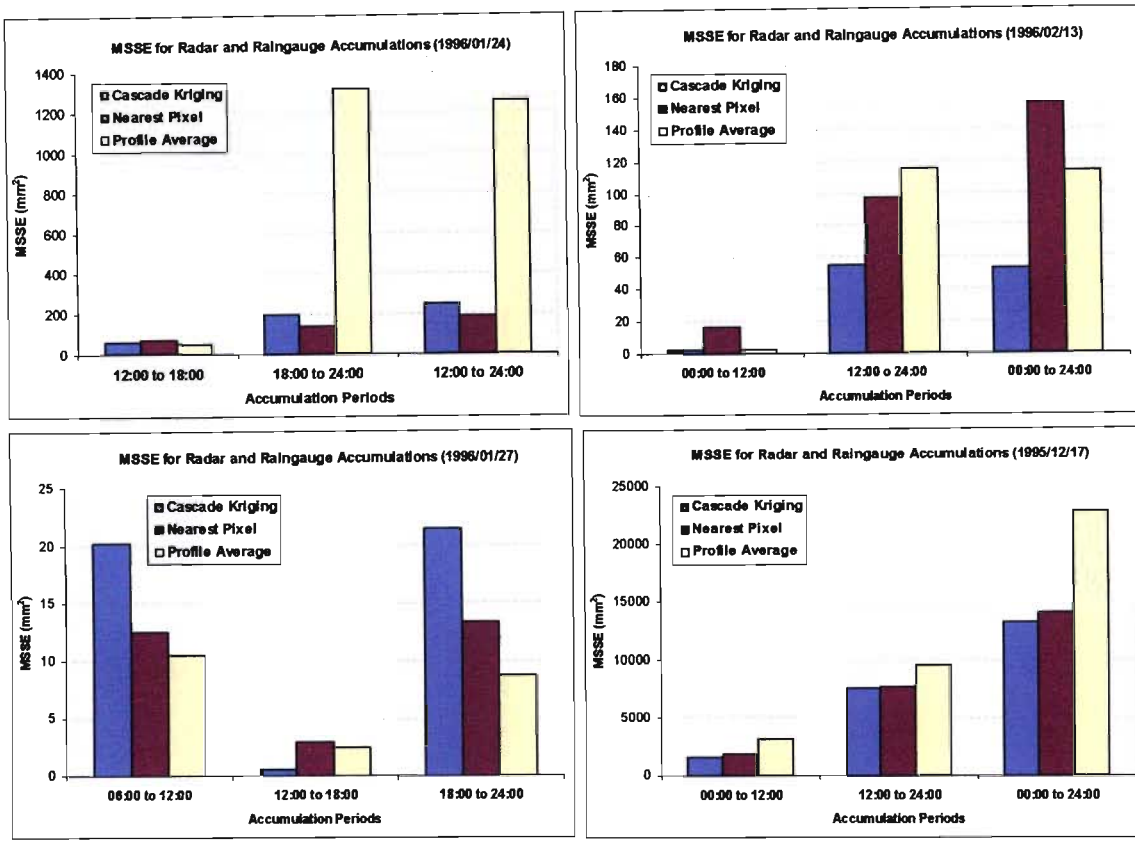


Figure 5.25: MSSE for Cascade Kriging, Nearest Pixel and Profile Average method to determine rainfall estimates at ground level.

The Cascade Kriging algorithm returned much improved estimates over the other techniques in rain events that consisted of a mixture of stratiform and convective rainfall over the accumulation periods over a range of rainfall events.

The average MSSE for each of the techniques over 6, 12 and 24-hour accumulation periods was computed and the results illustrated in Figure 5.26. As indicated in Figure 5.26 the Cascade Kriging algorithm returned superior estimates in terms of the MSSE over all the accumulation periods investigated; although it is marginally better than the Nearest pixel method it is much better than the Profile Averaging.

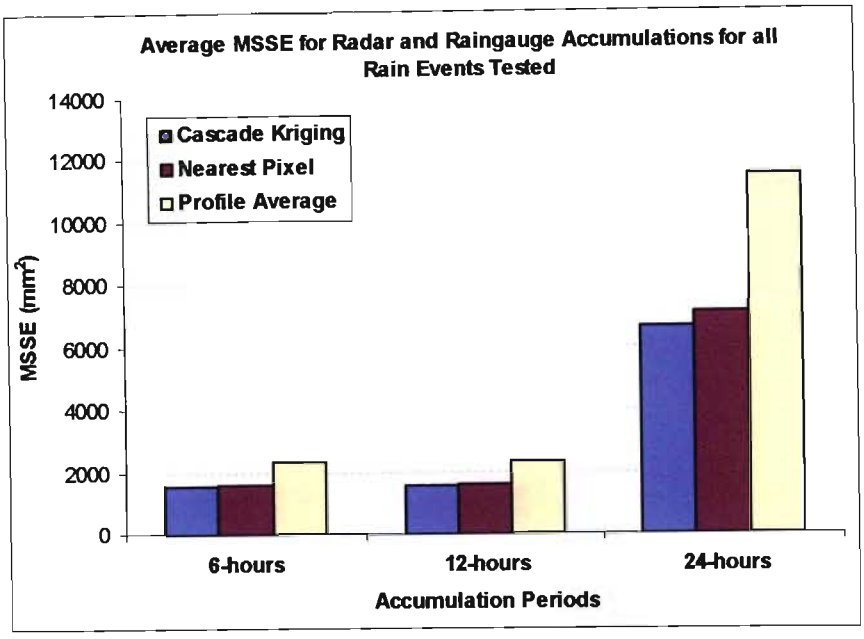


Figure 5.26: Average MSSE for radar and raingauge accumulations for all rain events tested over 6, 12 and 24-hour accumulation periods.

The Cascade Kriging algorithm however is not designed so as to solely provide an estimate of the rainfall at ground level but also to provide an estimate of the rainfall in any region of the radar volume scan that may be contaminated. The next section of this chapter details the testing and validation of the Cascade Kriging algorithm to infill ground clutter regions.

5.3 GROUND CLUTTER INFILLING

In order to test the efficiency and effectiveness of the Cascade Kriging algorithm in infilling ground clutter, a three dimensional picture of the Bethlehem ground clutter was obtained. The ground clutter map is mainly located in the South East corner of the radar volume scan data obtained from the Bethlehem weather radar, the ground clutter map contaminates the radar volume scan data from the CAPPI 1 km above ground level to the 5 km CAPPI and consists of a total of 1253 unknown data points. Indicated in Figure 5.27 is a 3D picture of the ground clutter map obtained from the Bethlehem weather radar illustrating the extent and nature of the ground clutter contamination.

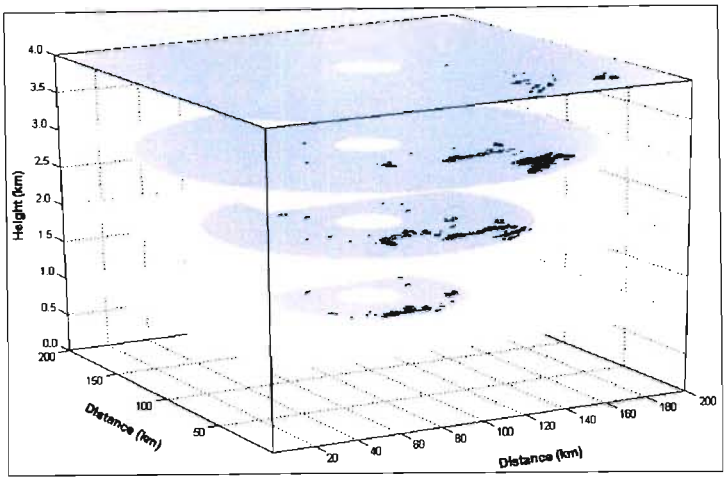


Figure 5.27: Bethlehem ground clutter situated in the South East corner of the radar volume scan contaminating the reflectivity data up to the 5 km CAPPI. The ground clutter comprises of 1253 unknown data points.

5.3.1 Statistical Testing Procedure

The testing procedure to determine the accuracy and computational efficiency of the Cascade Kriging algorithm to infill the ground clutter starts by rearranging the Bethlehem ground clutter map so that it is situated in the North East corner of the radar volume scan data and located in a region where no ground clutter contamination is present. The ground clutter map masks out observed radar reflectivity values in this region.

Three rain events from the Bethlehem weather radar were then selected that consisted of different rainfall types and tested over a 24-hour period; the set comprised approximately 280 instantaneous images for each 24-hour period. The rain events were from the following days:

- 17 December 1995 – a rain event that consisted of predominantly stratiform rain with isolated patches of light convective rainfall.
- 25 January 1996 – a rain event that consisted of intense localised convective rainfall throughout the majority of the 24-hour period and also regions of stratiform rainfall.
- 13 February 1996 – a rain event that initially consisted of sparse stratiform rainfall in the first 10 hours of the day but increased in intensity throughout the remainder of the day.

The missing data that are masked as ground clutter is estimated for each of the instantaneous radar volume scans on that particular day. Once all the ground clutter pixels have been estimated the reflectivity values are converted to rainrate values by the Marshall-Palmer equation and the data stored in a vector. The observed reflectivity data that were masked out are also converted to rainrate values and stored in separate vector. This procedure is illustrated in Figure 5.28 for the 2 km CAPPI for a single instantaneous radar volume scan.

At the end of the 24 hour rain event both the estimated and observed rainrate values for each of the instantaneous radar volume scan data has been recorded and stored in order to be able to compare the accumulations of the data to determine the accuracy of the infilling technique. The estimated and observed rainrate data are compared over 6, 12 and 24-hour accumulation periods. The time taken to compute and infill all missing and contaminated ground clutter data from the 18 km to the 1 km CAPPI is also recorded for each instantaneous radar volume scan.

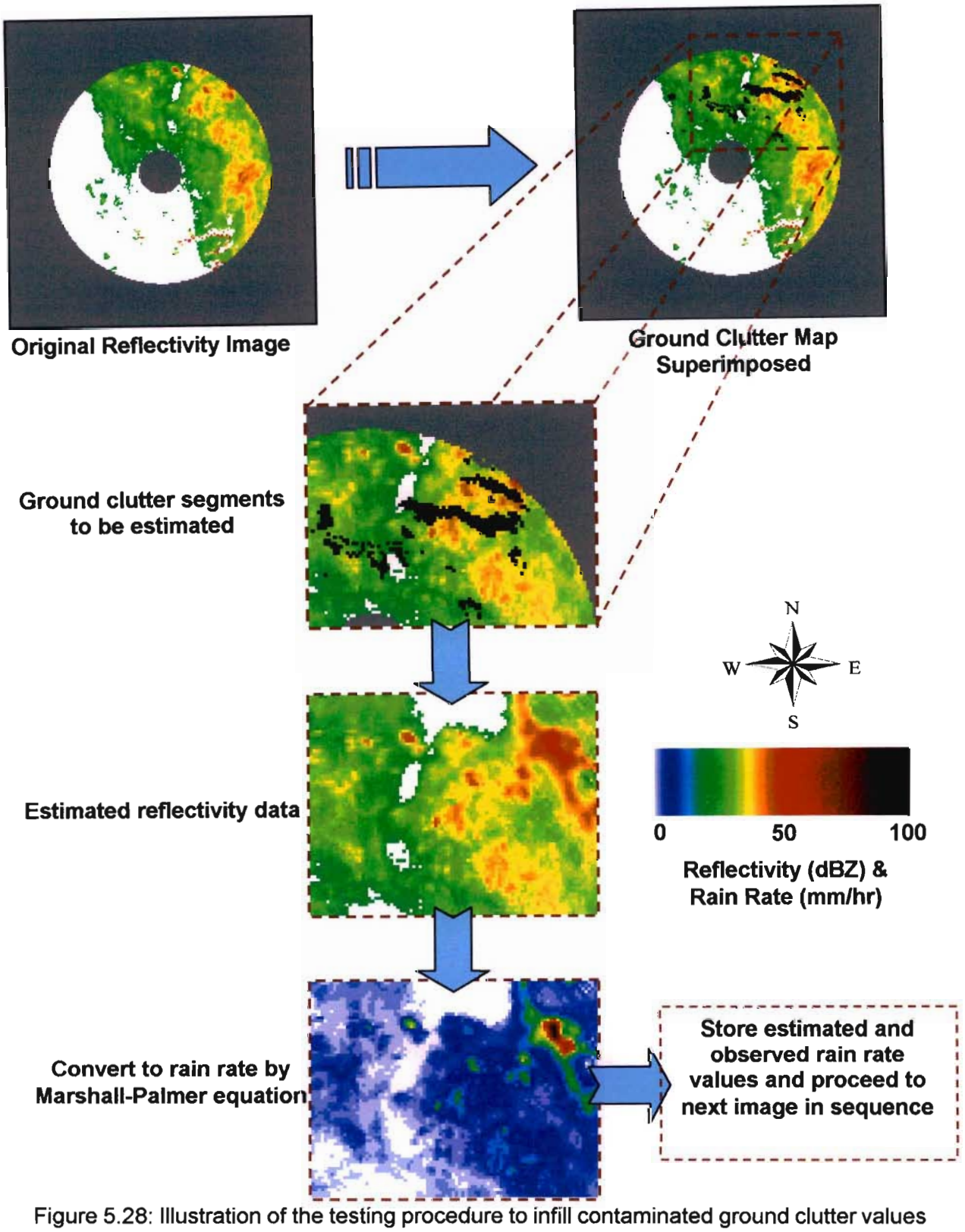


Figure 5.28: Illustration of the testing procedure to infill contaminated ground clutter values for a single instantaneous radar volume scan.

In order to undertake a meaningful statistical analysis between the observed and estimated values the following two factors need to be taken into account: firstly the very large sample size (n = 1253) would create an extremely small confidence interval when applying Equation (5.2) to a data set resulting in highly similar mean values being returned as dissimilar; secondly the data sets are highly spatially correlated.

Moolman (1985) derived a relation to take into account the correlation that exists within a univariate time series data set, where the derived formula is given by Equation (5.4).

$$\alpha(\phi) = 1 - \Phi\left[z_\alpha (1-\phi)\right] \tag{5.4}$$

Where z_α is the upper critical $100\alpha\%$ value, $\Phi(.)$ is the cumulative standard normal and $\alpha(\phi)$ is the significance level of the dependant series (Pegram, 2000). The ϕ value is determined from the climatological semivariograms where an approximate value of ϕ = 0.80 is utilised herein. It is proposed to apply Equation (5.4) in a 2D random field context. Although it may not be strictly correct as a test, there is no other appropriate test to be found.

In order to determine if the mean values of the observed and infilled data are dissimilar Equation (5.5) is utilised.

$$Z = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} \tag{5.5}$$

The tests were carried out at a significance level of 5%, with a Box-Cox transform (Equation (5.1)) applied to the data sets being tested. Where Z is the test statistic, $\bar{x}$ is the mean value of the Box-Cox transformed infilled rainfall at ground clutter locations, μ_0 is the mean value of the transformed observed rainfall at ground clutter locations, σ is the standard deviation of the observed rainfall and n the sample size. Where the hypothesis being tested is:

H_0: $\bar{x} = \mu_0$

H_1: $\bar{x} \neq \mu_0$

An example of how Equation (5.4) and (5.5) can be utilised to test the above hypothesis will now be given, where $\bar{x}$ = 4.22 mm, μ_0 = 4.14 mm, σ = 0.95 mm and n =1253. With the given data Z is calculated to be 2.98 which has a probability of exceedence of 0.0014 if the data are assumed independent spatially, which would mean that H_0 would be rejected although the means are very close (within 1.90% of each other). Employing Moolmans (1985) adjustment and multiplying the Z value (2.98) by (1-ϕ) yields a value of 0.596 which in turn has a probability of exceedence of 0.274. In this example it can therefore be cautiously concluded that mean values are not dissimilar and the hypothesis H_0 can be accepted with H_1 being rejected.

Apart from determining whether the mean values between the data sets are dissimilar other statistical tests are utilised to determine how accurately the ground clutter is infilled. Scatter plots of the estimated and observed data are plotted and the correlation coefficient computed (r^2), the Cumulative Distribution Functions (CDFs) of the observed and estimated accumulations are also plotted for comparison purposes. The standard deviation (σ^2), median, minimum value returned, maximum value returned and range computed to give an indication of the accuracy of the estimated data.

5.3.2 Results Rain Event 1 - 17 December 1995

The rain event from the 17 December 1995 was utilised as the first rain event to test the efficiency and accuracy of the Cascade Kriging algorithm to infill regions of ground clutter. The rain event consisted predominantly of stratiform rainfall with isolated periods of convective rainfall.

6-HOUR ACCUMULATIONS

The 6-hour accumulations for the observed and estimated rainfall were computed over the 24-hour rainfall period with the scatter plots indicated in Figure 5.29. The observed and estimated accumulation values are highly similar with the majority of the values lying along the 1-1 Line. The high correspondence between the accumulation values is reflected in the r^2 values returned which range from 0.72 to 0.94. The CDFs for the 6-hour accumulations are also plotted as indicated in Figure 5.30 which indicates that the distributions for the estimated and observed values match very closely for all the accumulation periods.

Table 5.7 provides a summary of the statistics returned for the accumulation periods indicating that all the statistics returned between the observed and estimated accumulation values are highly similar. To test whether the means were considered to be dissimilar the testing procedure described in Section 5.3.1 is utilised. All the mean values returned from the estimated data were considered not to be dissimilar to the observed data, with the hypothesis H_0 being accepted and H_1 rejected.

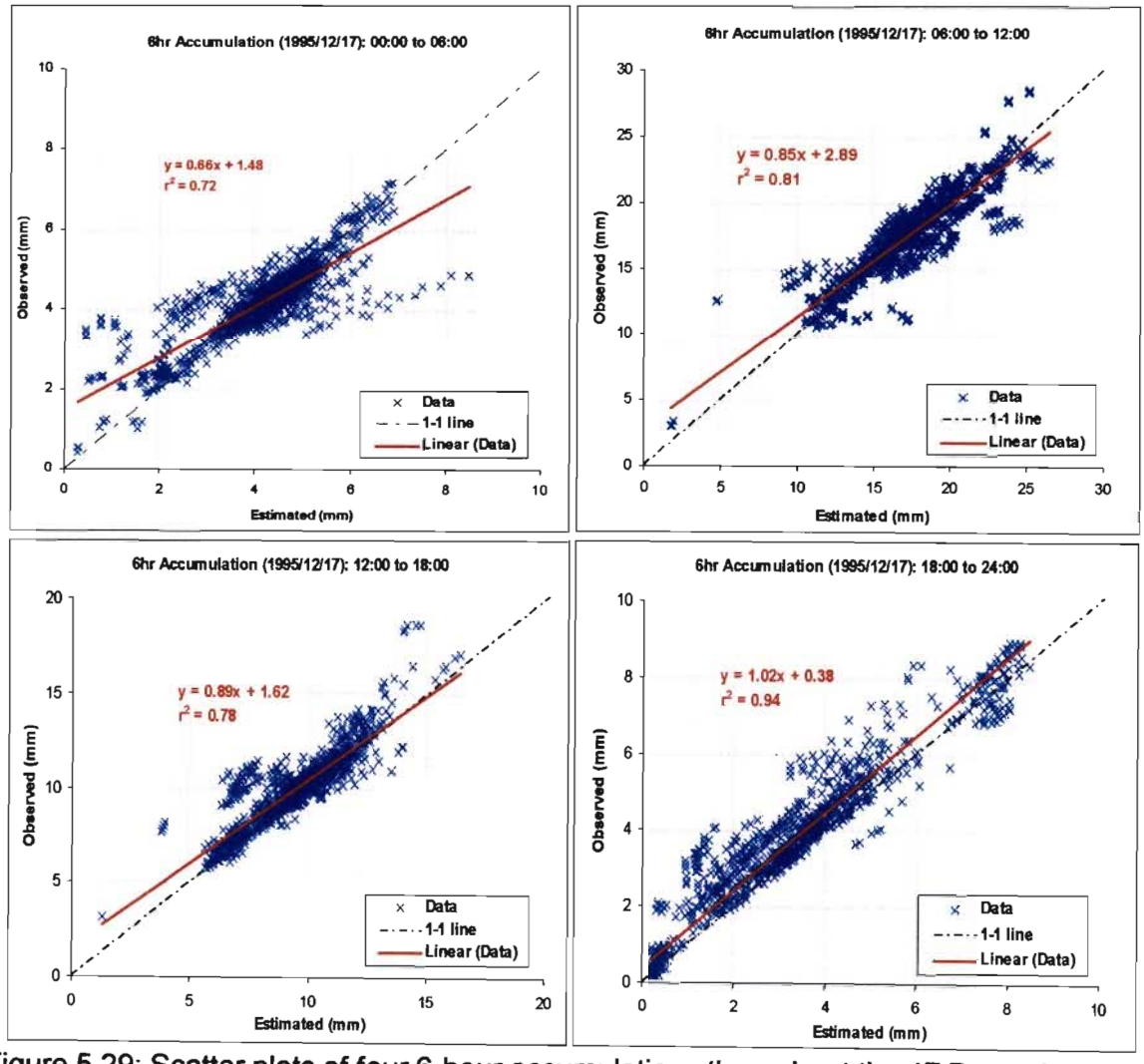


Figure 5.29: Scatter plots of four 6-hour accumulations throughout the 17 December 1995 of the observed and estimated accumulation values; the correlation coefficient (r^2) is also included on the plots.

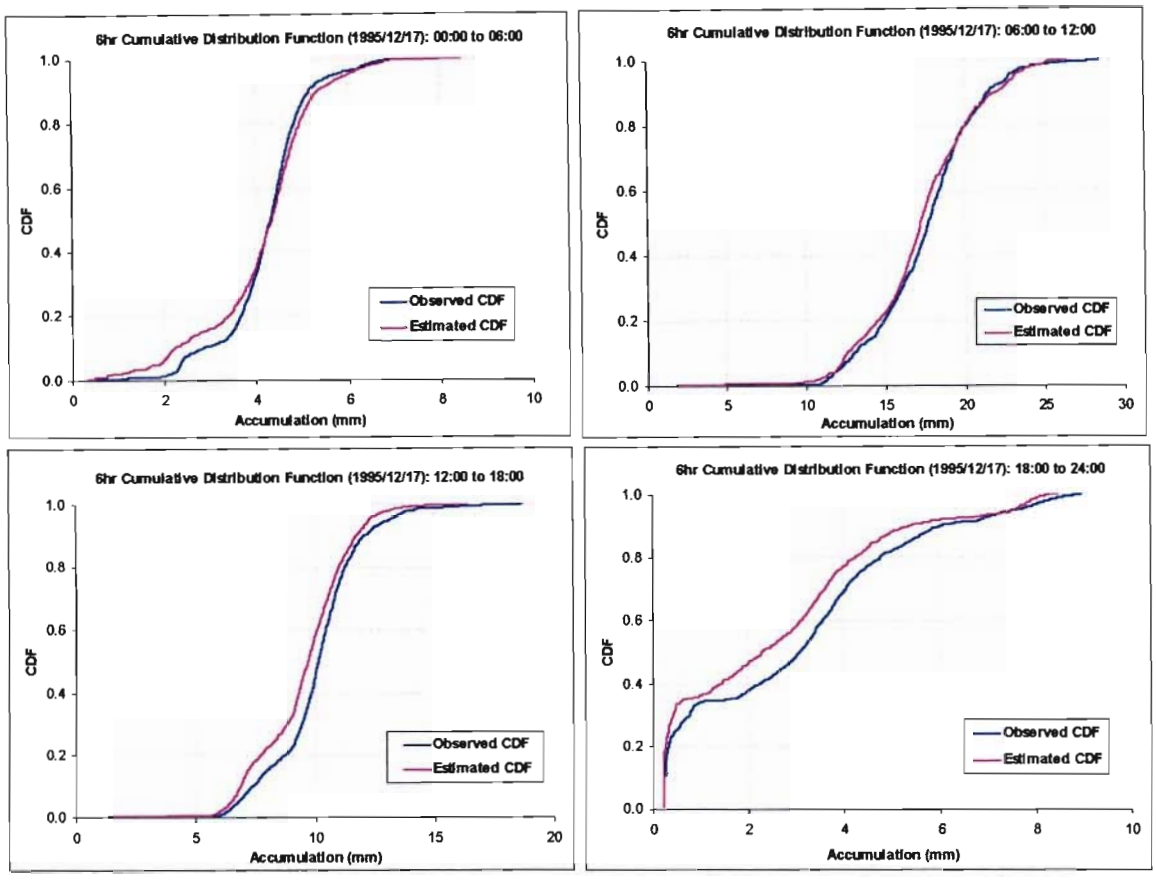


Figure 5.30: Cumulative Distribution Functions (CDFs) for the 6-hour accumulations returned for the 17 December 1995.

Table 5.7: Summary of the statistics returned for the 6-hour accumulation periods throughout the 17 December 1995.

	6hr Accums (00:00 to 06:00)		6hr Accums (06:00 to 12:00)		6hr Accums (12:00 to 18:00)		6hr Accums (18:00 to 24:00)	
	Obs.	**Est.**	**Obs.**	**Est.**	**Obs.**	**Est.**	**Obs.**	**Est.**
Mean	4.22	4.14	17.57	17.29	10.06	9.54	2.95	2.53
Stdev	0.95	1.21	3.29	3.49	1.89	1.89	2.32	2.21
Median	4.32	4.34	17.71	17.19	10.15	9.72	3.08	2.30
Min	0.44	0.28	3.16	1.83	3.21	1.33	0.22	0.22
Max	7.17	8.48	28.55	26.51	18.68	16.40	8.96	8.48
r^2	0.72		0.81		0.78		0.94	

12-HOUR ACCUMULATIONS

The 12-hour accumulations returned for the 17 December 1995 rain event also indicate a very close correspondence between the estimated and observed accumulation values. The scatter plots and CDFs for the two 12-hour accumulations are indicated in Figure 5.31 indicate the quality of the estimated rainfall accumulations. In the scatter plots r^2 values of 0.77 and 0.88 were returned indicating very similar values for the observed and estimated accumulations. The CDFs also indicate highly similar distributions returned for both estimated and observed accumulation values.

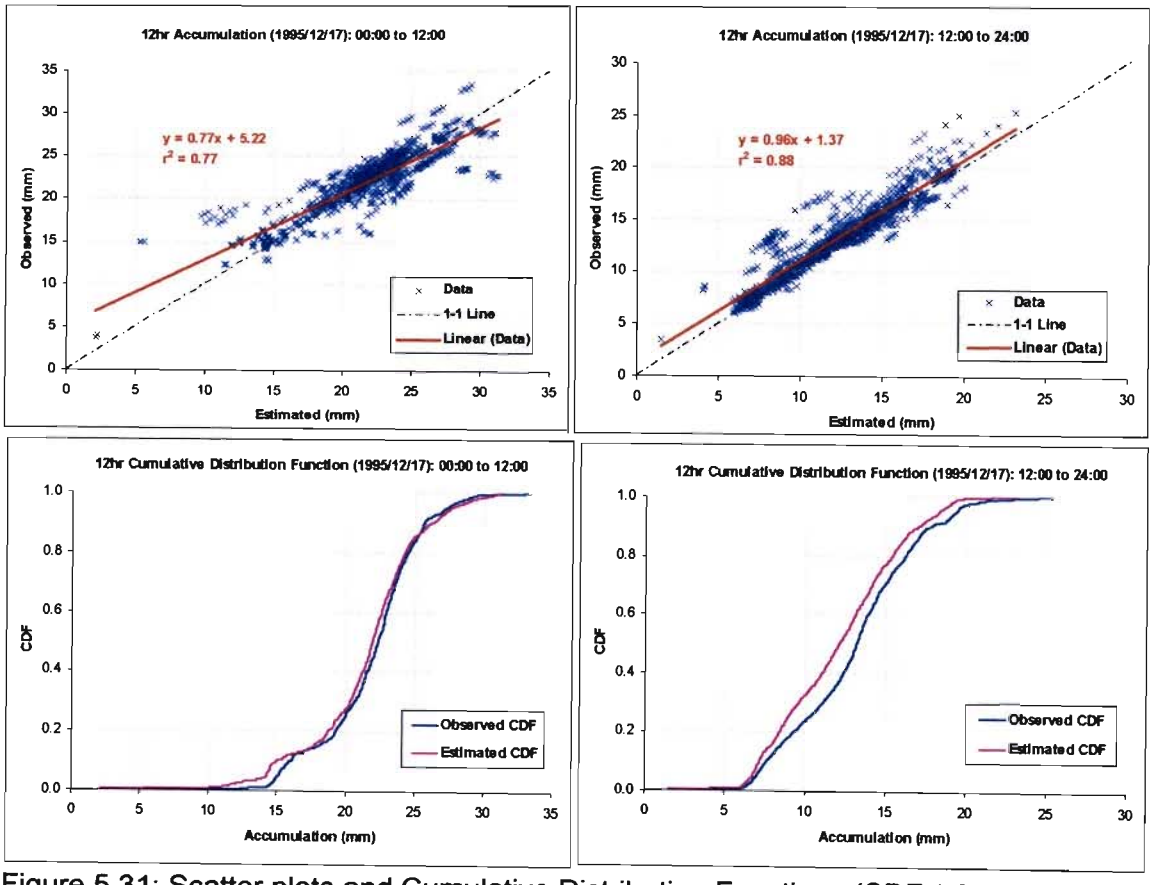


Figure 5.31: Scatter plots and Cumulative Distribution Functions (CDFs) for the two 12-hour accumulation periods for the 17 December 1995.

24-HOUR ACCUMULATION

The 24-hour accumulation returned also indicate a close correspondence between the observed and estimated accumulation as illustrated by the scatter plot and CDF in Figure 5.32 with an r^2 value of 0.70 being returned. The CDF also indicates a close correspondence between the two distributions. Table 5.8 gives a summary of the statistics for the 24-hour and 12-hour accumulation periods; all the statistics displayed indicate a close correspondence between the observed and estimated values. The mean values for the two 12-hour accumulation periods and the 24-hour accumulation period were all considered not to be dissimilar, where the hypothesis H_0 is accepted and H_1 rejected.

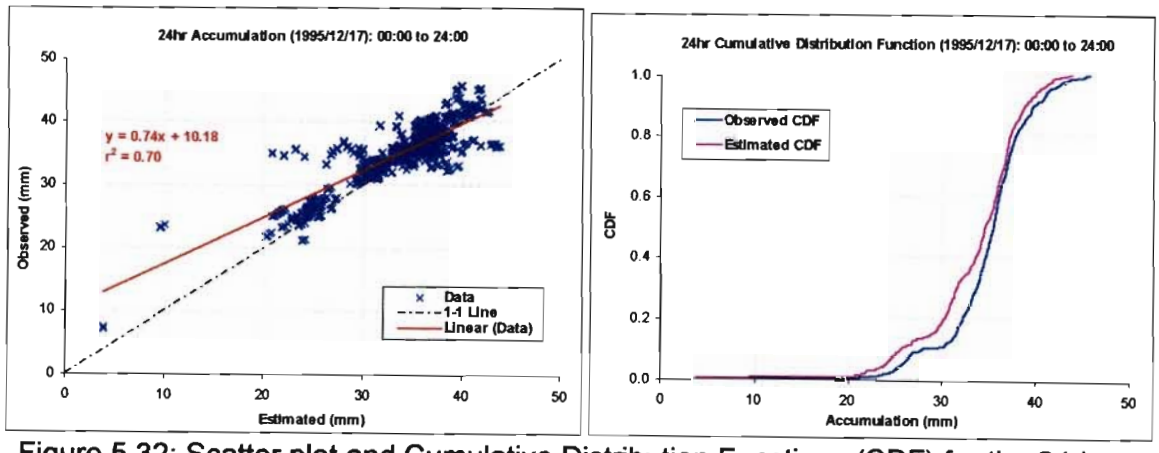


Figure 5.32: Scatter plot and Cumulative Distribution Functions (CDF) for the 24-hour accumulation period for the 17 December 1995.

Table 5.8: Summary of the statistics returned for the 24 and two 12-hour accumulation periods throughout the 17 December 1995.

	12hr Accumulations (00:00 to 12:00)		12hr Accumulations (12:00 to 24:00)		24hr Accumulations (00:00 to 24:00)	
	Observed	**Estimated**	**Observed**	**Estimated**	**Observed**	**Estimated**
Mean	21.79	21.43	13.02	12.07	34.81	33.50
Stdev	3.64	4.13	3.73	3.62	4.54	5.17
Median	22.32	21.92	13.20	12.13	35.35	34.51
Min	3.70	2.15	3.43	1.54	7.13	3.70
Max	33.33	31.31	25.30	23.14	45.68	43.77
r^2	0.77		0.88		0.70	

Illustrated in Figure 5.33 are the computation times to infill all the missing data contained in a radar volume scan, including ground clutter, from the 18 km CAPPI to the 1 km CAPPI. The average time to infill the missing data was 30 seconds with the minimum computation time being 9 seconds and the maximum 62 seconds (1 minute 2 seconds). These times are well below the 5 minute intervals which are required for real time application. All computations for the ground clutter infilling were done on a Pentium(R) 4 with a 2.40GHz CPU and 512MB of RAM.

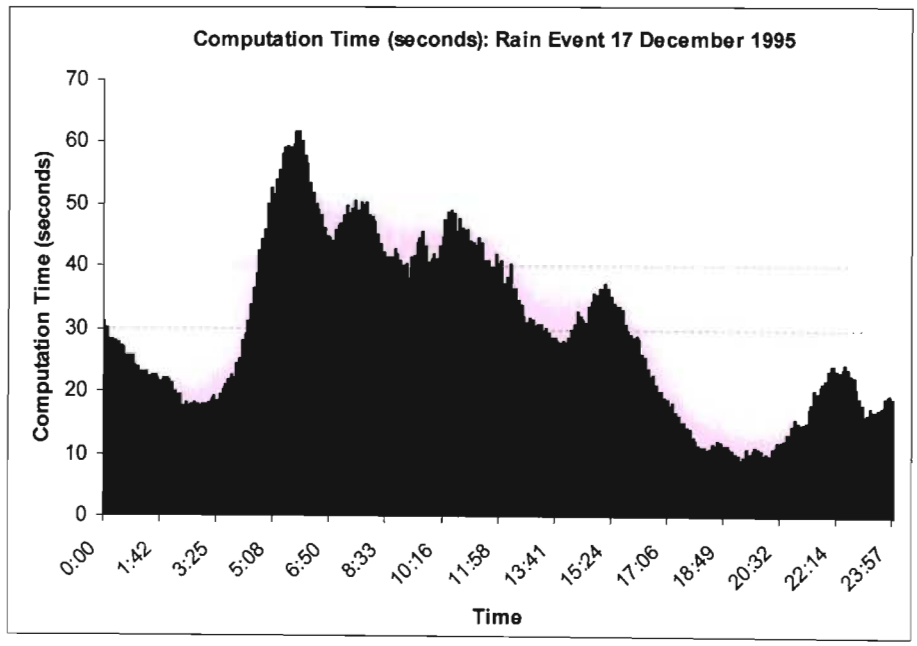


Figure 5.33: Computation time (seconds) to infill all missing data including ground clutter from the 18 km CAPPI to the 1 km CAPPI over the 24-hour period for the 17 December 1995.

5.3.3 Results Rain Event 2 - 25 January 1996

The rain event from the 25 January 1996 is also from the Bethlehem weather radar, for the first quarter of the day the rain event initially consisted of a combination of stratiform and convective rainfall. Throughout the second quarter the rainfall became more sparse with predominantly stratiform rainfall been recorded. Throughout the remainder of the day the rainfall increased slightly with intensity with a combination of convective and stratiform rainfall been recorded.

6-HOUR ACCUMULATIONS

The 6-hour accumulations returned from the 25 January 1996 indicate an extremely close correspondence between the observed and estimated values. The scatter plots of the observed and estimated accumulation values are given by Figure 5.34 which indicate the high accuracy of the returned estimates. The correlation coefficients (r^2) returned range from 0.80 to 0.97. The CDFs for the observed and estimated values, illustrated in Figure 5.35, also indicate a close correspondence between the observed and estimated accumulation values, especially during the 12:00 to 18:00 time period. A summary of the statistics for the 6-hour accumulations is given in Table 5.9 indicating the high accuracy of the returned estimates. The mean values returned were all considered not to be dissimilar except for the mean value returned for the 06:00 to 12:00 period.

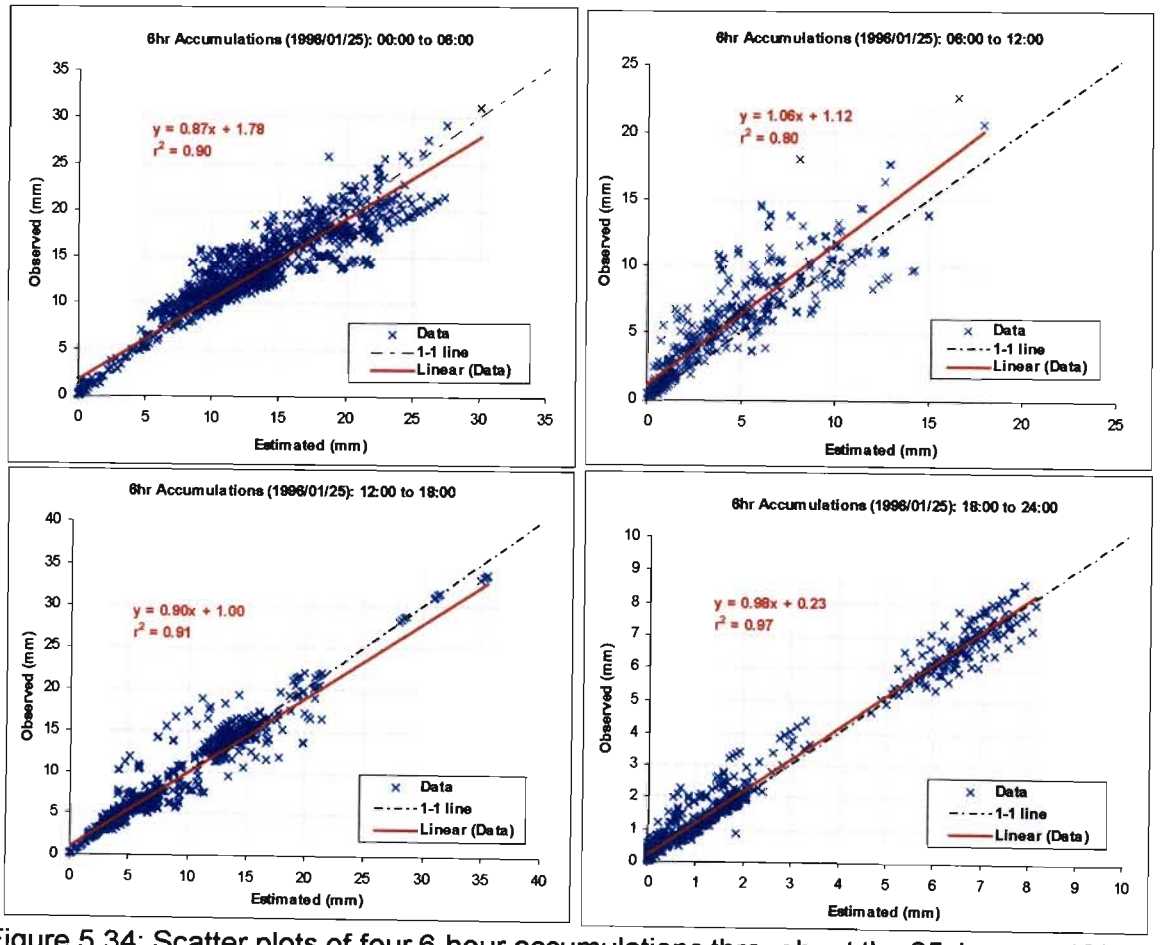


Figure 5.34: Scatter plots of four 6-hour accumulations throughout the 25 January 1996 of the observed and estimated accumulation values; the correlation coefficient (r^2) is also included on the plots.

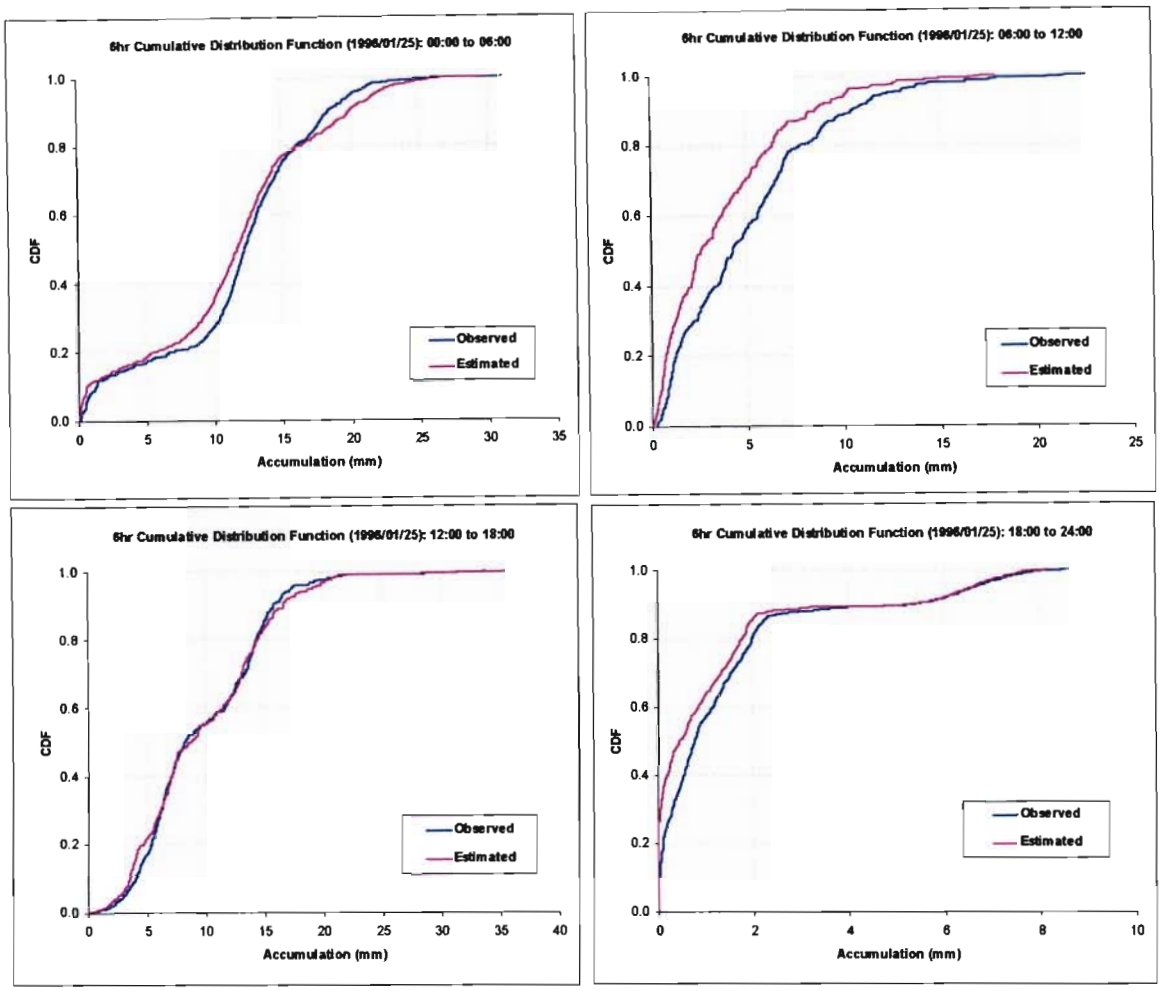


Figure 5.35: Cumulative Distribution Functions (CDFs) for the 6-hour accumulations returned for the 25 January 1996.

Table 5.9: Summary of the statistics returned for the 6-hour accumulation periods throughout the 25 January 1996.

	6hr Accums (00:00 to 06:00)		6hr Accums (06:00 to 12:00)		6hr Accums (12:00 to 18:00)		6hr Accums (18:00 to 24:00)	
	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.
Mean	11.57	11.28	5.03	3.70	9.75	9.75	1.47	1.26
Stdev	5.80	6.34	3.98	3.36	5.14	5.46	1.96	1.96
Median	12.18	11.68	4.25	2.59	8.18	8.54	0.75	0.44
Min	0.08	0.00	0.24	0.00	0.07	0.00	0.00	0.00
Max	31.01	30.07	22.57	17.87	33.75	35.42	8.58	8.15
r^2	0.91		0.80		0.90		0.97	

12-HOUR ACCUMULATIONS

The scatter plots and CDFs for the two 12-hour accumulations are shown in Figure 5.36. The two scatter plots indicate once again a high correspondence between the observed and estimated accumulation values. The computed r^2 values returned are 0.87 and 0.88 indicating a high level of correspondence between the two estimates. The CDFs also indicate a high level of correspondence between the estimated and observed distributions; the correspondence is especially high for the 12:00 to 24:00 period.

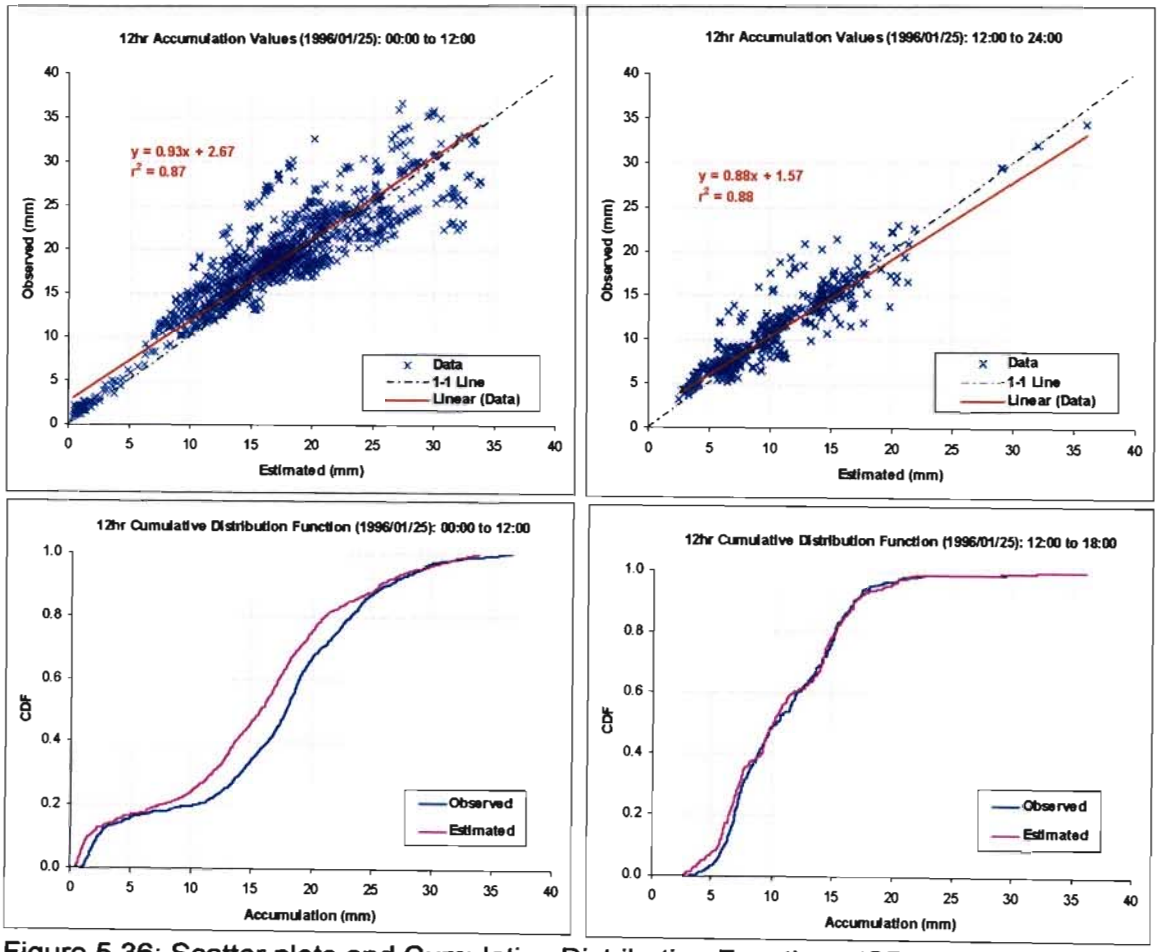


Figure 5.36: Scatter plots and Cumulative Distribution Functions (CDFs) for the two 12-hour accumulation periods for the 25 January 1996.

24-HOUR ACCUMULATION

The 24-hour accumulation values for the 25 January 1996 are illustrated in Figure 5.37, which gives an indication of the accuracy of the estimated rainfall accumulations at the ground clutter locations. The scatter plot of the observed and estimated accumulations returned an r^2 value of 0.81, a CDF of the accumulation values also indicate a highly

similar distribution for both the observed and estimated values. Table 5.10 provides a summary of the statistics returned for the 24-hour accumulation indicating the high quality of the estimates. The mean values returned for the observed and estimated accumulations over the 24-hour period were all not dissimilar, with the hypothesis H_0 being accepted and H_1 rejected.

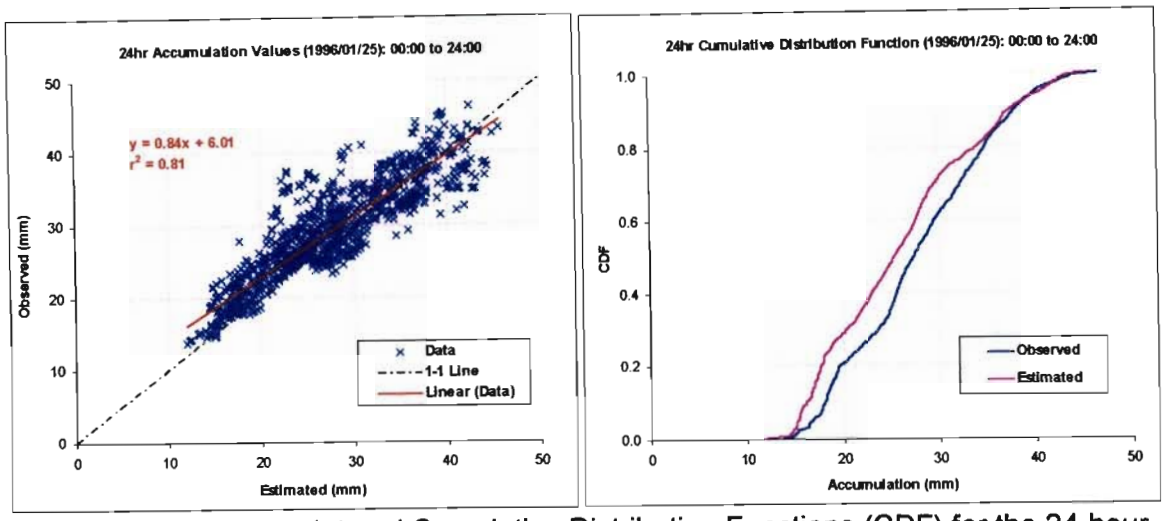


Figure 5.37: Scatter plot and Cumulative Distribution Functions (CDF) for the 24-hour accumulation period for the 25 January 1996.

Table 5.10: Summary of the statistics returned for the 24 and two 12-hour accumulation periods throughout the 25 January 1996.

	12hr Accumulations (00:00 to 12:00)		12hr Accumulations (12:00 to 24:00)		24hr Accumulations (00:00 to 24:00)	
	Observed	**Estimated**	**Observed**	**Estimated**	**Observed**	**Estimated**
Mean	16.60	14.97	11.21	11.01	27.81	25.98
Stdev	8.15	8.17	4.66	4.98	7.41	7.93
Median	17.88	15.76	10.49	10.15	27.37	25.33
Min	0.88	0.46	3.29	2.68	13.67	11.92
Max	36.60	33.86	34.39	36.08	46.38	45.69
r^2	0.87		0.88		0.81	

Illustrated in Figure 5.38 are the computation times to infill all the missing data contained in a radar volume scan including, including ground clutter, from the 18 km CAPPI to the 1 km CAPPI. The average time to infill the missing data was 33 seconds with the minimum computation time being 5 seconds and the maximum 80 seconds (1 minute 20 seconds).

These times are well below the 5 minute intervals which are required for real time application.

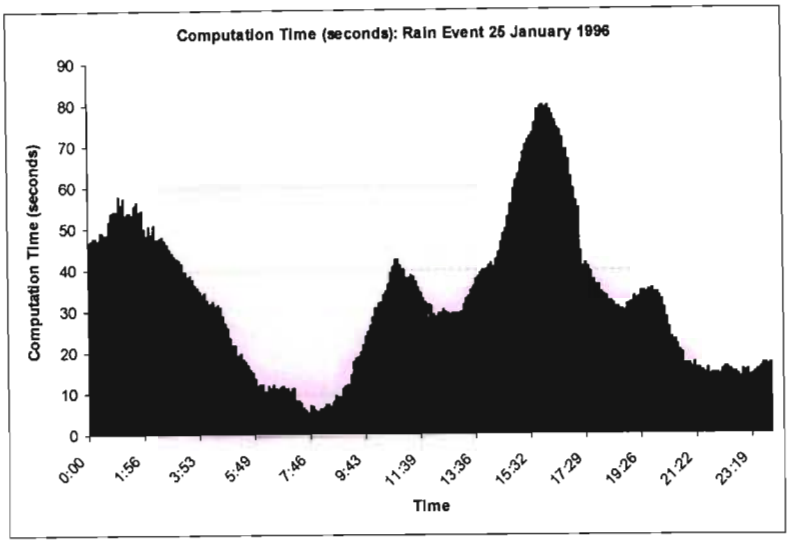


Figure 5.38: Computation time (seconds) to infill all missing data including ground clutter from the 18 km CAPPI to the 1 km CAPPI over the 24-hour period for the 25 January 1996.

5.3.4 Results Rain Event 3 - 13 February 1996

The rain event for the 13 February 1996 consisted of a combination of both stratiform and intense convective rainfall. Initially for the time period 00:00 to 09:00 the rain was very sparse and stratiform in nature. Throughout the remainder of the 13 February 1996 the rainfall became more widespread and increased in intensity.

6-HOUR ACCUMULATIONS

The scatter plots of the 6-hour accumulation values are illustrated in Figure 5.39. The sparse and stratiform nature of the rainfall is indicated in the scatter plot for the 00:00 to 06:00 period. For the lower rainfall accumulation values, in the periods 00:00 to 06:00 and 06:00 to 12:00, a weaker correspondence exists between the observed and estimated accumulation values with r^2 values of 0.49 and 0.72 being returned. For the more convective rainfall of greater intensity, in the periods 12:00 to 18:00 and 18:00 to 24:00, r^2 values 0.96 and 0.91 were returned. The CDFs for the observed and estimated rainfall is indicated in Figure 5.40 also indicating a closer correspondence between the observed and estimated accumulation values for the rainfall of greater intensity.

Given in Table 5.11 contains a summary of the statistics returned for each of the four 6-hour accumulation periods for the 13 February 1996. All the statistics returned for the observed and estimated accumulation periods indicate an accurate estimation of the rainfall at ground clutter locations. All the mean values of the observed and estimated accumulations were considered to be not dissimilar apart for the 06:00 to 12:00 period.

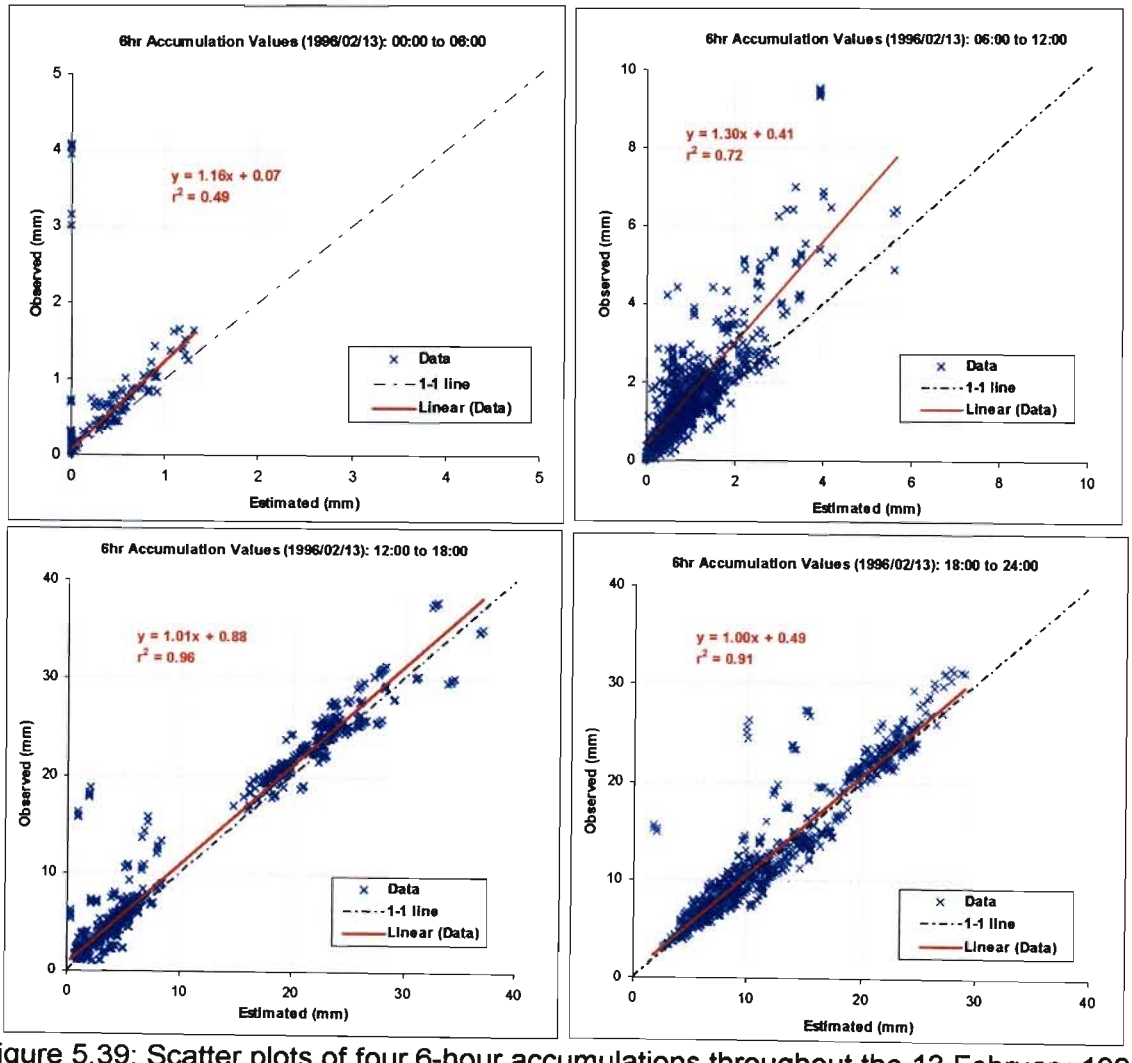


Figure 5.39: Scatter plots of four 6-hour accumulations throughout the 13 February 1996 of the observed and estimated accumulation values; the correlation coefficient (r^2) is also included on the plots.

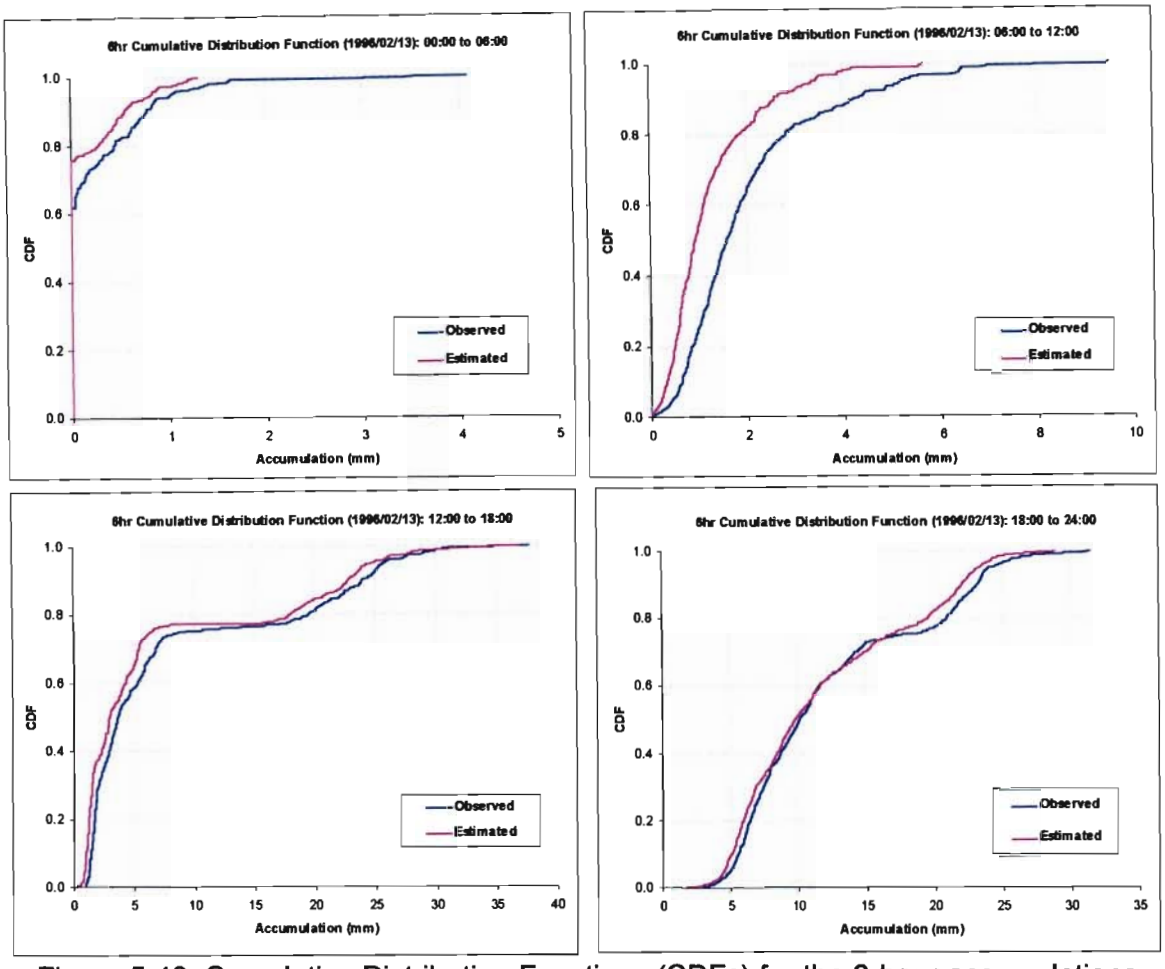


Figure 5.40: Cumulative Distribution Functions (CDFs) for the 6-hour accumulations returned for the 13 February 1996.

Table 5.11: Summary of the statistics returned for the 6-hour accumulation periods throughout the 13 February 1996.

	6hr Accums (00:00 to 06:00)		6hr Accums (06:00 to 12:00)		6hr Accums (12:00 to 18:00)		6hr Accums (18:00 to 24:00)	
	Obs.	**Est.**	**Obs.**	**Est.**	**Obs.**	**Est.**	**Obs.**	**Est.**
Mean	0.22	0.13	2.00	1.22	8.23	7.29	12.39	11.86
Stdev	0.47	0.28	1.50	0.98	8.94	8.67	6.71	6.38
Median	0.00	0.00	1.57	0.92	3.73	3.04	10.13	9.80
Min	0.00	0.00	0.00	0.00	1.07	0.31	3.05	1.81
Max	4.07	1.31	9.53	5.65	37.76	36.95	31.50	28.90
r^2	0.49		0.72		0.96		0.91	

12-HOUR ACCUMULATIONS

The 12-hour accumulation values also indicated a high level of correspondence between the observed and estimated rainfall accumulation values. The scatter plots and CDFs of the 12-hour accumulation are illustrated in Figure 5.41. The scatter plot for the initial 12-hour period indicates a reasonably strong correspondence between the observed and estimated accumulations with an r^2 value of 0.63, with the estimated values underestimating the observed values. In the second 12-hour period the correspondence between the two estimates is far greater with an r^2 value of 0.95 being returned.

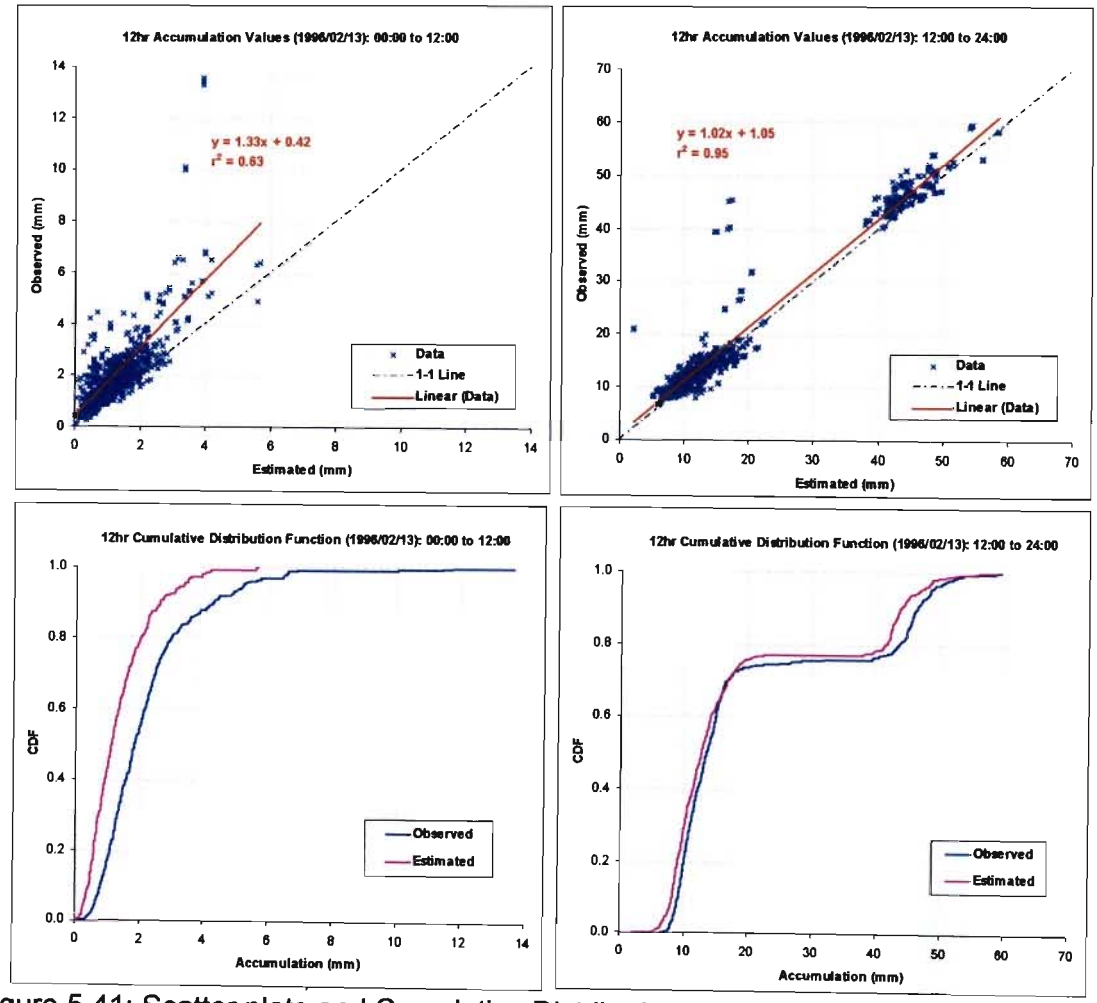


Figure 5.41: Scatter plots and Cumulative Distribution Functions (CDFs) for the two 12-hour accumulation periods for the 13 February1996.

24-HOUR ACCUMULATION

The 24-hour accumulation scatter plot and CDF are illustrated in Figure 5.42. The scatter plot indicates a close correspondence between the estimated and observed accumulation

values with an r^2 value of 0.93 been returned, however indicated in the plot is that the infilled or estimated rainfall is slightly underestimating the observed values. The CDF also indicates a close correspondence between the observed and estimated distributions. Table 5.12 indicates the statistics returned for the two 12-hour and 24-hour accumulation period indicating the accuracy of the estimated results. The mean values returned for the observed and estimated accumulations over the two 12 and one 24-hour periods were all considered to be not dissimilar, except for the 12-hour period from 00:00 to 12:00.

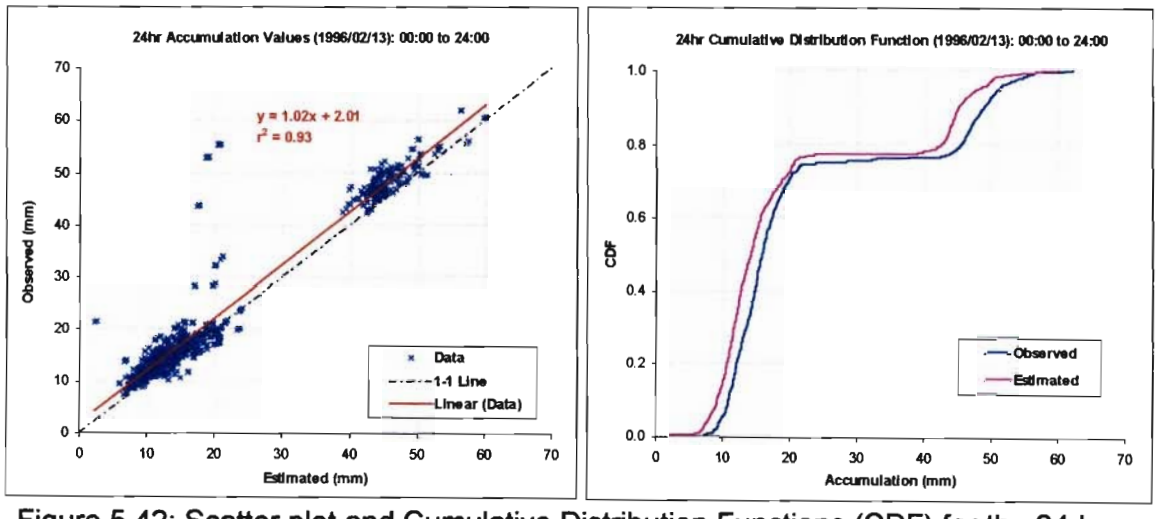


Figure 5.42: Scatter plot and Cumulative Distribution Functions (CDF) for the 24-hour accumulation period for the 13 February 1996.

Table 5.12: Summary of the statistics returned for the 24 and two 12-hour accumulation periods throughout the 13 February1996.

	12hr Accumulations (00:00 to 12:00)		12hr Accumulations (12:00 to 24:00)		24hr Accumulations (00:00 to 24:00)	
	Observed	**Estimated**	**Observed**	**Estimated**	**Observed**	**Estimated**
Mean	2.22	1.35	20.62	19.15	22.84	20.50
Stdev	1.60	0.95	14.88	14.16	15.02	14.23
Median	1.83	1.12	13.45	12.72	15.79	14.12
Min	0.13	0.00	6.60	2.12	7.16	2.12
Max	13.58	5.65	59.50	58.66	62.00	59.98
r^2	0.63		0.95		0.93	

Illustrated in Figure 5.43 are the computation times to infill all the missing data contained in a radar volume scan including, including ground clutter, from the 18 km CAPPI to the 1 km CAPPI. The average time to infill the missing data was 29 seconds with the minimum computation time being 5 seconds and the maximum 71 seconds (1 minute 11 seconds). These are once again well below the 5 minute intervals which are required for real time application.

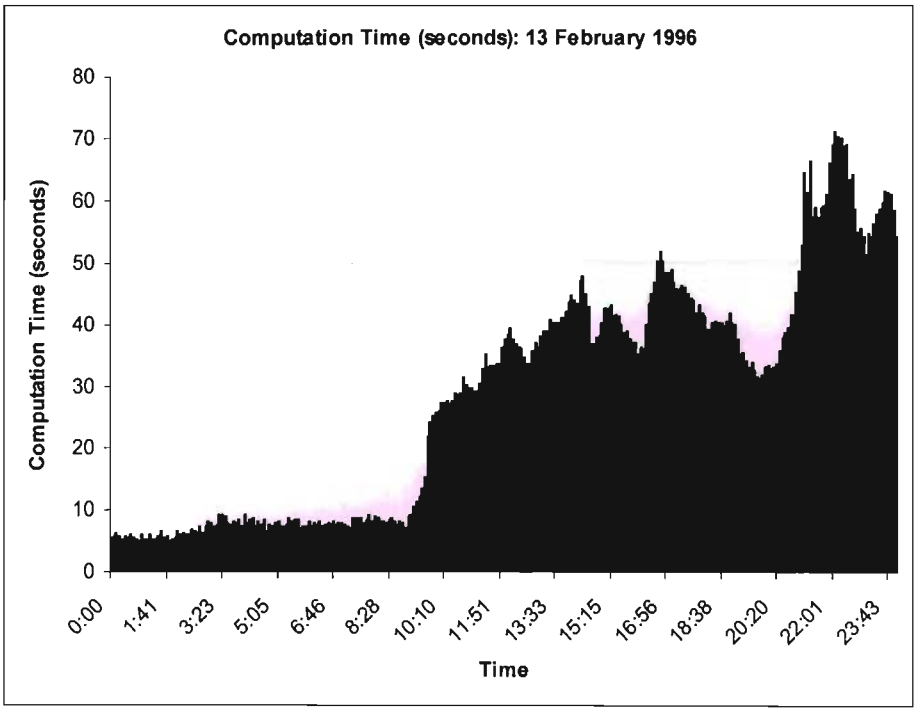


Figure 5.43: Computation time (seconds) to infill all missing data including ground clutter from the 18 km CAPPI to the 1 km CAPPI over the 24-hour period for the 13 February 1996.

5.4 CHAPTER SUMMARY

This chapter contains methods for testing and validating the Cascade Kriging algorithm's ability to provide an accurate indication of rainfall at ground level and its ability to infill regions of contaminated data.

To test the Cascade Kriging algorithms ability to provide a real time and accurate estimate of the rainfall at ground level the radar accumulation estimates is statistically compared to the Block Kriged raingauge accumulations. This is done over four distinctly different rain events that contained rainfall consisting of intense convective rainfall to very light

stratiform rainfall. The results indicate that a close correspondence exist between the radar and raingauge accumulations for convective rainfall. However there is a relatively poor correspondence between the raingauge and radar accumulations for light stratiform rainfall, with the radar accumulations generally underestimating the rainfall at ground level. This could be due to the fact that higher intensity rainfall is less likely to be affected by atmospheric phenomena such as wind drift where as rainfall of low intensity may be more susceptible to these atmospheric conditions.

The Cascade Kriging algorithm was also tested on its ability to infill ground clutter regions on three different rain events with rainfall ranging from stratiform to convective. The analysis indicates a high level of accuracy provided by the Cascade Kriging algorithm to infill ground clutter on both stratiform and convective rainfall types. There is also no indication that there is degradation in the quality of the estimates when infilling ground clutter on the lower CAPPI level where the controls consist of previously estimated target points contained on upper CAPPI levels.

The computation times to infill all the missing data in the volume scan and providing an estimate of the rainfall at ground level is well below the 5 minutes (time between volume scans) required for real time application in South Africa. Overall the results indicate that the technique replaces regions of contaminated data in the radar volume scan with a reasonable level of accuracy and can also provide an estimate of the rainfall at ground level which captures the spatial nature and intensity of the rainfall with an acceptable amount of accuracy.

CHAPTER 6
CONCLUSION AND RECOMMENDATIONS

An algorithm has been presented in this thesis to improve the quality of weather radar estimates on a real time basis. After initial classification by reflectivity thresholding the algorithm first utilises a bright band correction to adjust reflectivity values on the 2 km CAPPI. The reflectivity estimates are then progressively extrapolated to ground level in the process infilling all unknown and contaminated data in the radar volume scan. This is achieved by utilising a combination of Ordinary and Universal Kriging depending on rainfall type where climatological semivariograms are utilised. Computational stability and efficiency are ensured by using a moving nearest neighbourhood approach and also utilising SVD matrix rank reduction techniques to trim small singular values that are the cause of numerical instabilities.

The algorithm was tested on a variety of weather types over different accumulation periods. The results indicate that the Cascade Kriging algorithm infills ground clutter regions with a high degree of accuracy and in a computationally efficient manner that makes it fast enough and suitable for real time implementation. The comparison of raingauge and weather radar accumulation estimates at ground level indicates that the extrapolations to ground level capture the magnitude of the rainfall relatively accurately. The results indicate that the proposed algorithm works well for convective rainfall but less accurately for less intense stratiform rainfall.

Based on the accuracy of the results obtained and the computational efficiency of the algorithm, SAWS has taken the decision to implement the Cascade Kriging algorithm on the South African weather radar network to improve the data quality returned. The computer code is currently being adapted so as to run on the computer system utilised at the Bethlehem weather office run by the Innovation and Research Division of SAWS.

Recommendations for further research and development for the Cascade Kriging algorithm include the following:

- The estimate at ground level at present is extrapolated to a surface at radar level which is a constant altitude above mean sea level. The ground level estimates can

be improved by extrapolating to a Digital Elevation Model (DEM) which models the natural topography surrounding the radar site.

- Raingauges might usefully be placed in regions contaminated with ground clutter in order to determine the accuracy of the estimates at ground level in these areas.
- Research needs to be undertaken to determine the accuracy of the rainfall estimates at ground level with range from the weather radar for various rainfall types. This is especially relevant for stratiform rainfall which has limited vertical extent.
- The Marshall-Palmer Z-R relationship should be revaluated as suggested by Visser (2003) in order to improve the correspondence between the raingauges and radar estimates at ground level, now that the best radar estimates are available.
- Further testing of the correspondence between radar and raingauges estimates at ground level could be investigated by using data sets from other countries where more extensive raingauge networks exist.
- The bright band correction algorithm needs to be adjusted so as to be site specific, as the bright band may affect higher CAPPIs (above 2 km) in other regions of South Africa, especially near the warmer coast.

APPENDIX

COMPUTER CODE FOR OPERATIONAL INFILLING

Contained in this section is the computer code written in C++ for the operational infilling of all missing data contained in the radar volume scan. A flow chart of the algorithm and the sequence of the procedures carried out in the code are illustrated in Figure 4.38 on pages 104 and 105. The computer code contained below is currently being adapted by the Innovation and Research Division of SAWS so as to operate on their computer systems.

```
/************************************************************
      Cascade_Final_Version.cpp
      Cascade Kriging algorithm to infill all missing
      data contained in a radar volume scan.
      Date: 5 November 2005
      Author: Stephen Wesson
************************************************************/

#include <iostream>
#include <fstream>
#include <vector>
#include <algorithm>
#include <cmath>
#include <string>
#include <sstream>
#include <cstdlib>
#include <numeric>

//#if defined _MSC_VER // Microsoft Visual C++
  // Must define default file access as binary
  // This is assumed to be true by the mdv libraries.
  #include <fcntl.h>
  extern int _fmode = _O_BINARY;
  #pragma warning(disable : 4508) // return from main()
//#endif

#include "boost/numeric/ublas/matrix.hpp"
#include "boost/numeric/ublas/vector.hpp"
#include "boost/numeric/ublas/io.hpp"
#include "mdv/mdv_handle.h"
#include "mdv/mdv_read.h"
#include "mdv/mdv_user.h"
#include "mdv/mdv_write.h"
typedef boost::numeric::ublas::matrix<double> Matrix;
typedef boost::numeric::ublas::vector<double> Vector;
namespace ublas = boost::numeric::ublas;

// Read in the mdv files
typedef ublas::matrix<double> dbl_mat;
std::vector<dbl_mat> read_mdv_field( std::string const&, int );
void mdv_field_dim( char const*, int, unsigned int&, unsigned int&, unsigned int&
);

// local functions
void write_mdv( char const*, std::vector<dbl_mat> const&,  int );
void convert_to_byte( MDV_handle_t&, std::vector<dbl_mat> const&, ui08 *** );
```

```
// Matrix Inversion by use of LU decompostion
#define NRANSI
#include "nrutil.h"
#undef NRANSI
#ifdef __cplusplus
extern "C" {
  #endif
  void ludcmp (float **a, int n, int *indx, float *d);
  void lubksb (float **a, int n, int *indx, float b[]);
  #ifdef __cplusplus
}
#endif

// Matrix Inversion by use of SVD
Matrix inverse_NRIC(Matrix const&cov);

// Read in the mdv files
typedef ublas::matrix<double> dbl_mat;
std::vector<dbl_mat> read_mdv_field( std::string const&, int );
void mdv_field_dim( char const*, int, unsigned int&, unsigned int&, unsigned int&
);

// vector to store temporary values
class Temp_Data {
public: int y_control_temp;
        int x_control_temp;
        int z_control_temp;
        double dBZ_control_temp;
        double dist_to_target_temp;
};

// vector to store control data
class Control_Data {
public: int y_control;
        int x_control;
        int z_control;
        double dBZ_control;
        double dist_to_target;
};

// vector to store target data location and estimate of target
class Target_Data {
public: int y_target;
        int x_target;
        double dBZ_target;
};

// bright band correction function call
void BB_correction_algorithm (std::vector< ublas::matrix<double> >& radar_data,
int nrows_radar, int ncols_radar, double total_points_1, double total_points_2,
double total_points_3, const int No_Data);

// search algorithm to locate nearest controls to a target
void spherical_search (int z, const int No_Data, int y, int x, int range1, int
range2, int lev, int nrows_radar, int ncols_radar, std::vector<
ublas::matrix<double> >const& radar_data, std::vector<Control_Data>&controls,
std::vector<Temp_Data>&temps, double upper);

// Ordinary Kriging function call stratiform controls
double Ordinary_Kriging_Strat (std::vector<Control_Data>const&controls, int y,
int x, int z);

// Universal Kriging function call for mixed rainfall where
```

```
// controls are predominantly stratiform
double Universal_Kriging_Mixed_Strat (std::vector<Control_Data>const&controls,
int y, int x, int z, double r0, double h0, double H);

// Universal Kriging function call for mixed rainfall where
// controls are predominantly convective
double Universal_Kriging_Mixed_Convec (std::vector<Control_Data>const&controls,
int y, int x, int z, double r0, double h0, double H);

// Ordinary Kriging function call for convective rainfall
double Ordinary_Kriging_Convec (std::vector<Control_Data>const&controls, int y,
int x, int z);

// semivariogram for controls consisting entirely of convective rainfall
inline double Cov_Convec_to_Convec (double r, double h);

// semivariogram for controls consisting entirely of stratiform rainfall
inline double Cov_Strat_to_Strat (double r, double h);

// semivariogram for controls of mixed controls (convective & stratiform
// rainfall)
inline double Cov_Mixed (double r, double h, double r0, double h0, double H);

void usage()
{
  std::cout << "Cascade Kriging Data Infilling Algorithm by Stephen Wesson
               [wessons@ukzn.ac.za]\n"
            << "Version 1.0 - November 2005\n\n"
            << "usage: cascade_kriger <input_mdv_file> <output_mdv_file>\n\n";
}

int main(int argc, char *argv[])
{
  if ( argc != 3 ){
    usage();
    exit(0);
  }

  int a, b, c;
  int const field = 0;
  const int No_Data = 255;
  const int No_Rain = 0;
  const double Convec_Threshold = 35.0;

  // determing the dimensions of the mdv file
  // find the number of rows, columns & number of levels in the mdv file
  unsigned int nrows_radar, ncols_radar, max_level_radar;
  mdv_field_dim(argv[1], field, nrows_radar, ncols_radar, max_level_radar);

  // vector of 19 matrices that stores the radar volume scan data
  std::vector< ublas::matrix<double> > radar_data;
  radar_data = read_mdv_field(argv[1], field);

  // determine the number of data points on 1, 2 & 3km CAPPIs
  // to be used in Bright Band Correction to determine the percentage rain on
  // each CAPPI
  double total_points_1 = 0;
  double total_points_2 = 0;
  double total_points_3 = 0;
  for (b=0; b<nrows_radar; b++) {
    for (c=0; c<ncols_radar; c++) {
      if (radar_data[1][b][c] != No_Data) {
        total_points_1++;
      }
```

```
    if (radar_data[2][b][c] != No_Data) {
      total_points_2++;
    }
    if (radar_data[3][b][c] != No_Data) {
      total_points_3++;
    }
  }
}

// all radar reflectivity data thresholdeded from the 1 to 18km CAPPI
// is set to 0dBZ if less than or equal to 18dBZ
for (a=1; a<=max_level_radar; a++) {
  for (b=0; b<nrows_radar; b++) {
    for (c=0; c<ncols_radar; c++) {
      if (radar_data[a][b][c]<=18) {
        radar_data[a][b][c] = 0;
      }
    }
  }
}

// bright band correction function call
BB_correction_algorithm(radar_data, nrows_radar, ncols_radar, total_points_1,
total_points_2, total_points_3, No_Data);

// vector to store location and estimated value of targets
std::vector<Target_Data> targets;
// vector to store "control data"
std::vector<Control_Data> controls;
// vector to store "temporary data"
std::vector<Temp_Data> temps;

// loop from top to bottom (18km to 1km) through the CAPPIs
for (int z=max_level_radar; z>0; z--) {

// infill two upper most levels (18 & 17km) where no data located with 0dBZ
  if (z>(max_level_radar-2)) {
    // search through CAPPI for unknown data points
    for (int y=0; y<nrows_radar; y++) {
      for (int x=0; x<ncols_radar; x++) {
        if (radar_data[z][y][x]==No_Data) {
          radar_data[z][y][x] = No_Rain;
        }
      }
    }
  }

// Infill all missing data on 16km to 1km CAPPIs with Kriging estimates
else {
  // search through CAPPI for unknown data points
  for (int y=0; y<nrows_radar; y++) {
    for (int x=0; x<ncols_radar; x++) {
      if (radar_data[z][y][x]==No_Data) {
        // number of controls to infill unknown data points
        int Control_Points = 25;
        int level = z;
        // range values for row & column search iterations
        int range01 = 1;
        int range02 = 0;
        int range11 = 1;
        int range12 = 0;
        int range21 = 1;
        int range22 = 0;
        int range31 = 1;
```

```
int range32 = 0;
int range41 = 1;
int range42 = 0;
// upper range bound
double upper0 = 1.999;
double upper1 = 1.999;
double upper2 = 2.999;
double upper3 = 3.999;
double upper4 = 4.999;
// initiate search algorithm at this point
// searches outwards until 25 nearest controls located
  do {
    if (level>=z) {
      const int lev = z;
      spherical_search(z, No_Data, y, x, range01, range02, lev,
      nrows_radar, ncols_radar, radar_data, controls, temps, upper0);
      upper0+=1;
      range01++;
      range02++;
    }
    if (level>=z) {
      const int lev = (z+1);
      spherical_search(z, No_Data, y, x, range11, range12, lev,
      nrows_radar, ncols_radar, radar_data, controls, temps, upper1);
      upper1+=1;
      range11++;
      range12++;
    }
    if (level>=(z+1)) {
      const int lev = (z+2);
      spherical_search(z, No_Data, y, x, range21, range22, lev,
      nrows_radar, ncols_radar, radar_data, controls, temps, upper2);
      upper2+=1;
      range21++;
      range22++;
    }
    if (level>=(z+2)) {
      const int lev = (z+3);
      spherical_search(z, No_Data, y, x, range31, range32, lev,
      nrows_radar, ncols_radar, radar_data, controls, temps, upper3);
      upper3+=1;
      range31++;
      range32++;
    }
    if (level>=(z+3)) {
      const int lev = (z+4);
      spherical_search(z, No_Data, y, x, range41, range42, lev,
      nrows_radar, ncols_radar, radar_data, controls, temps, upper4);
      upper4+=1;
      range41++;
      range42++;
    }
    level++;
  } while (Control_Points>controls.size());

// reducing the number of controls to exactly 25
if (controls.size()>Control_Points) {
  do {
    double dist_max = controls[0].dist_to_target;
    for (b=0; b<controls.size(); b++) {
      if (dist_max < controls[b].dist_to_target) {
        dist_max = controls[b].dist_to_target;
      }
    }
```

```
    for (c=0; c<controls.size(); c++) {
      if (controls[c].dist_to_target==dist_max) {
        controls.erase(controls.begin()+c);
        break;
      }
    }
  } while (controls.size()>Control_Points);
}
// check to see if the sum of controls is equal to zero
// also count the number of convective pixels located
double dBZ_SUM=0;
int convec_counter = 0;
for (b=0; b<controls.size(); b++) {
  dBZ_SUM+=controls[b].dBZ_control;
  if (controls[b].dBZ_control>=Convec_Threshold) {
    convec_counter++;
  }
}

// estimating the target based on controls selected
// if controls sum to 0dBZ target assigned a 0dBZ value
if (dBZ_SUM==No_Rain) {
  Target_Data values;
  values.y_target = y;
  values.x_target = x;
  values.dBZ_target = No_Rain;
  targets.push_back(values);
}
// if controls all stratiform
if ((convec_counter==0) && (dBZ_SUM!=No_Rain)) {
  Target_Data values;
  values.y_target = y;
  values.x_target = x;
  values.dBZ_target = Ordinary_Kriging_Strat(controls, y, x, z);
  targets.push_back(values);
}
// rainfall consists of stratiform & convective controls
if ((convec_counter>0) && (convec_counter<controls.size())) {
  // computing new semivariogram paramters based on ratios of
  // convective and stratiform pixels located
  double Num_Convec = convec_counter;
  double Num_Strat = Control_Points - convec_counter;
  double Num_Controls = Control_Points;
  double r0 = ((3.38*Num_Convec)+(8.40*Num_Strat))/Num_Controls;
  double h0 = ((4.11*Num_Convec)+(2.56*Num_Strat))/Num_Controls;
  double H = ((1.78*Num_Convec)+(1.43*Num_Strat))/Num_Controls;

  // majority of pixels located are stratiform
  if (Num_Convec<(Num_Controls/2)) {
    Target_Data values;
    values.y_target = y;
    values.x_target = x;
    values.dBZ_target = Universal_Kriging_Mixed_Strat(controls, y, x,
    z, r0, h0, H);
    targets.push_back(values);
  }
  // majority of pixels located are convective
  else {
    Target_Data values;
    values.y_target = y;
    values.x_target = x;
    values.dBZ_target = Universal_Kriging_Mixed_Convec(controls, y,
    x, z, r0, h0, H);
    targets.push_back(values);
```

```
          }
        }
        // rainfall consists entirely of convective pixels
        if (convec_counter==controls.size()) {
          Target_Data values;
          values.y_target = y;
          values.x_target = x;
          values.dBZ_target = Ordinary_Kriging_Convec(controls, y, x, z);
          targets.push_back(values);
        }

        // clearing the vectors used to store the data
        controls.clear();
        temps.clear();

      } // end of "if" loop
    } // end of "x" / "col" loop
  } // end of "y" / "row" loop

  // infill all missing data on CAPPI "z" at this point
  for (b=0; b<targets.size(); b++) {
    radar_data[z][targets[b].y_target][targets[b].x_target] =
    targets[b].dBZ_target;
  }

  // clear "Target_Data" vector
  targets.clear();

  } // end of "else" loop
} // end of "z" loop


/*************************************************************
  FINAL STEP: ESTIMATING REFLECTIVITY AT GROUND LEVEL:
  KRIGED ESTIMATES ABOVE USED TO INFILL AT GROUND LEVEL
*************************************************************/
for (int y=0; y<nrows_radar; y++) {
  for (int x=0; x<ncols_radar; x++) {
    // number of controls to infill unknown data points
    int Control_Points = 25;
    int z = 0;
    int level = 1;

    // Range values for row & column search iterations
    int range01 = 1;
    int range02 = 0;
    int range11 = 1;
    int range12 = 0;
    int range21 = 1;
    int range22 = 0;
    int range31 = 1;
    int range32 = 0;
    int range41 = 1;
    int range42 = 0;
    // Upper range bound
    double upper0 = 1.999;
    double upper1 = 2.999;
    double upper2 = 3.999;
    double upper3 = 4.999;
    double upper4 = 5.999;

    // initiate search algorithm at this point
    // searches outwards until 25 nearest controls located
    do {
```

```
  if (level>=(z+1)) {
    const int lev = z+1;
    spherical_search(z, No_Data, y, x, range01, range02, lev,
    nrows_radar, ncols_radar, radar_data, controls, temps, upper0);
    upper0+=1;
    range01++;
    range02++;
  }
  if (level>=(z+2)) {
    const int lev = z+2;
    spherical_search(z, No_Data, y, x, range11, range12, lev,
    nrows_radar, ncols_radar, radar_data, controls, temps, upper1);
    upper1+=1;
    range11++;
    range12++;
  }
  if (level>=(z+3)) {
    const int lev = z+3;
    spherical_search(z, No_Data, y, x, range21, range22, lev,
    nrows_radar, ncols_radar, radar_data, controls, temps, upper2);
    upper2+=1;
    range21++;
    range22++;
  }
  if (level>=(z+4)) {
    const int lev = z+4;
    spherical_search(z, No_Data, y, x, range31, range32, lev,
    nrows_radar, ncols_radar, radar_data, controls, temps, upper3);
    upper3+=1;
    range31++;
    range32++;
  }
  if (level>=(z+5)) {
    const int lev = z+5;
    spherical_search(z, No_Data, y, x, range41, range42, lev,
    nrows_radar, ncols_radar, radar_data, controls, temps, upper4);
    upper4+=1;
    range41++;
    range42++;
  }
  level++;
} while (Control_Points>controls.size());

// Reducing the number of controls to the exact number required
if (controls.size()>Control_Points) {
  do {
    double dist_max = controls[0].dist_to_target;
      for (b=0; b<controls.size(); b++) {
        if (dist_max < controls[b].dist_to_target) {
          dist_max = controls[b].dist_to_target;
        }
      }
    for (c=0; c<controls.size(); c++) {
      if (controls[c].dist_to_target==dist_max) {
        controls.erase(controls.begin()+c);
        break;
      }
    }
  } while (controls.size()>Control_Points);
}

// check to see if the sum of controls is equal to zero
// also count the number of convective pixels located
double dBZ_SUM=0;
```

```
      int convec_counter = 0;
      for (b=0; b<controls.size(); b++) {
        dBZ_SUM+=controls[b].dBZ_control;
        if (controls[b].dBZ_control>=Convec_Threshold) {
          convec_counter++;
        }
      }

      // estimating the target based on controls selected
      // if controls located sum to 0dBZ target assigned a 0dBZ value
      if (dBZ_SUM==No_Rain) {
        radar_data[z][y][x] = No_Rain;
      }
      // if rainfall all stratiform
      if ((convec_counter==0) && (dBZ_SUM!=No_Rain)) {
        radar_data[z][y][x] = Ordinary_Kriging_Strat(controls, y, x, z);
      }
      // rainfall consists of stratiform & convective pixels
      if ((convec_counter>0) && (convec_counter<controls.size())) {
        // computing new semivariogram paramters based on ratios of
        // convective and stratiform pixels located
        double Num_Convec = convec_counter;
        double Num_Strat = Control_Points - convec_counter;
        double Num_Controls = Control_Points;
        double r0 = ((3.38*Num_Convec)+(8.40*Num_Strat))/Num_Controls;
        double h0 = ((4.11*Num_Convec)+(2.56*Num_Strat))/Num_Controls;
        double H = ((1.78*Num_Convec)+(1.43*Num_Strat))/Num_Controls;

        // majority of pixels located are stratiform
        if (Num_Convec<(Num_Controls/2)) {
          radar_data[z][y][x] = Universal_Kriging_Mixed_Strat(controls, y, x,
          z, r0, h0, H);
        }
        // majority of pixels located are convective
        else {
          radar_data[z][y][x] = Universal_Kriging_Mixed_Convec(controls, y,
          x, z, r0, h0, H);
        }
      }
      // rainfall consists entirely of convective pixels
      if (convec_counter==controls.size()) {
        radar_data[z][y][x] = Ordinary_Kriging_Convec(controls, y, x, z);
      }
      // clear "controls" and "temps" vector
      controls.clear();
      temps.clear();
    }
  }
  // point where the mdv file is created and written
  write_mdv( argv[2], radar_data, nrows_radar );
} // end of int main()

// bright band correction function call
void BB_correction_algorithm (std::vector< ublas::matrix<double> >& radar_data,
int nrows_radar, int ncols_radar, double total_points_1, double total_points_2,
double total_points_3, const int No_Data)
{
  int y, x;

  // input text file that stores previous (@ t-1) weighted means and standard
  // deviation values
  // initial values: Weighted Mean (0.50) CAPPI 1km = 24.24 dBZ
  // Weighted Mean (0.50) CAPPI 2km = 24.73 dBZ
  // Weighted Mean (0.50) CAPPI 3km = 24.00 dBZ
```

```
// Weighted Standard Deviation (0.50) CAPPI 1km = 2.63 dB
// Weighted Mean (0.05) CAPPI 1km = 24.24 dBZ
// Weighted Mean (0.05) CAPPI 2km = 24.73 dBZ
// Weighted Mean (0.05) CAPPI 3km = 24.00 dBZ
// Weighted Standard Deviation (0.05) CAPPI 1km = 2.63 dB
std::ifstream BB_in;
BB_in.open("D:/Bethlehem_Code/Cascade_Kriging_Final/BB_values.txt");
double BB_data;
std::vector<double> BB_info;
while (BB_in >> BB_data){
  BB_info.push_back(BB_data);
}
double cappi1_adj = BB_info[0];
double cappi2_adj = BB_info[1];
double cappi3_adj = BB_info[2];
double stdev1_adj = BB_info[3];
double cappi1_wt = BB_info[4];
double cappi2_wt = BB_info[5];
double cappi3_wt = BB_info[6];
double stdev1_wt = BB_info[7];

// compute the statistics of stratiform rain on each level
// (1) Mean @ 1, 2 & 3km CAPPIs
// (2) Standard Deviation @ 1km CAPPI
// (3) Percentage of rainfall on each CAPPI
double n1 = 0;
double n2 = 0;
double n3 = 0;
double rain_sum1 = 0;
double rain_sum2 = 0;
double rain_sum3 = 0;
double rain_sumsqr1 = 0;

for (y=0; y<nrows_radar; y++) {
  for (x=0; x<ncols_radar; x++) {
    if ((radar_data[1][y][x]!=No_Data) && (radar_data[1][y][x]>18) &&
    (radar_data[1][y][x]<35)) {
      n1++;
      rain_sum1+=radar_data[1][y][x];
      rain_sumsqr1+=((radar_data[1][y][x])*(radar_data[1][y][x]));
    }
    if ((radar_data[2][y][x]!=No_Data) && (radar_data[2][y][x]>18) &&
    (radar_data[2][y][x]<35)) {
      n2++;
      rain_sum2+=radar_data[2][y][x];
    }
    if ((radar_data[3][y][x]!=No_Data) && (radar_data[3][y][x]>18) &&
    (radar_data[3][y][x]<35)) {
      n3++;
      rain_sum3+=radar_data[3][y][x];
    }
  }
}

// compute the percentage of stratiform rainfall on each CAPPI
double percent_1 = (n1/total_points_1)*100.0;
double percent_2 = (n2/total_points_2)*100.0;
double percent_3 = (n3/total_points_3)*100.0;
// compute the mean & standard deviation of rainfall on each CAPPI
double mean_1 = (rain_sum1/n1);
double mean_2 = (rain_sum2/n2);
double mean_3 = (rain_sum3/n3);
double stdev_1 = sqrt(((n1*rain_sumsqr1)-(rain_sum1*rain_sum1))/(n1*(n1-
1.0)));
```

```
const double min_rain = 10.0;

// if percentage of rainfall greater then 10% check to see if Bright Band
// exists
if (percent_1>=min_rain && percent_2>=min_rain && percent_3>=min_rain) {
  // compute weighted means and standard deviations
  const double lambda = 0.50;
  double mean_lambda_1 = (lambda*(mean_1))+((1.0-lambda)*(cappi1_adj));
  double mean_lambda_2 = (lambda*(mean_2))+((1.0-lambda)*(cappi2_adj));
  double mean_lambda_3 = (lambda*(mean_3))+((1.0-lambda)*(cappi3_adj));
  double stdev_lambda_1 = (lambda*(stdev_1))+((1.0-lambda)*(stdev1_adj));
  const double delta = 0.05;
  double mean_delta_1 = (delta*(mean_1))+((1.0-delta)*(cappi1_wt));
  double mean_delta_2 = (delta*(mean_2))+((1.0-delta)*(cappi2_wt));
  double mean_delta_3 = (delta*(mean_3))+((1.0-delta)*(cappi3_wt));
  double stdev_delta_1 = (delta*(stdev_1))+((1.0-delta)*(stdev1_wt));

  // check to see if bright band exists
  if ((mean_lambda_1>mean_lambda_3) && (mean_lambda_2>=mean_lambda_1)) {
    for (y=0; y<nrows_radar; y++) {
      for (x=0; x<ncols_radar; x++) {
        // check each stratiform pixel in turn
        if ((radar_data[2][y][x]!=255) && (radar_data[2][y][x]>18) &&
        (radar_data[2][y][x]<35)) {
          double grad_crit = -((radar_data[3][y][x]-
          (mean_lambda_1+(1.0*stdev_lambda_1)))/2.0);
          double grad_obs = -((radar_data[3][y][x]-radar_data[2][y][x]));
          if (grad_obs>=grad_crit) {
            double new_pixel_value = ((radar_data[3][y][x] +
            mean_lambda_1)/2.0);
            if (new_pixel_value>=18.0) {
              radar_data[2][y][x] = new_pixel_value;
            }
            else {
              radar_data[2][y][x] = 0.0;
            }
          }
        }
      }
    }
  }

  cappi1_adj = mean_lambda_1;
  cappi2_adj = mean_lambda_2;
  cappi3_adj = mean_lambda_3;
  stdev1_adj = stdev_lambda_1;
  cappi1_wt = mean_delta_1;
  cappi2_wt = mean_delta_2;
  cappi3_wt = mean_delta_3;
  stdev1_wt = stdev_delta_1;

  // percentage of rainfall less then 10% no bright band correction
  else {
    const double delta = 0.05;
    double mean_delta_1 = (delta*(cappi1_wt))+((1.0-delta)*(cappi1_adj));
    double mean_delta_2 = (delta*(cappi2_wt))+((1.0-delta)*(cappi2_adj));
    double mean_delta_3 = (delta*(cappi3_wt))+((1.0-delta)*(cappi3_adj));
    double stdev_delta_1 = (delta*(stdev1_wt))+((1.0-delta)*(stdev1_adj));
    cappi1_wt = mean_delta_1;
    cappi2_wt = mean_delta_2;
    cappi3_wt = mean_delta_3;
    stdev1_wt = stdev_delta_1;
  }
```

```
// write weighted mean and standard deviation values to text file
std::ofstream BB_out;
BB_out.open("D:/Bethlehem_Code/Cascade_Kriging_Final/BB_values.txt");
BB_out << cappi1_adj << "\n" << cappi2_adj << "\n" << cappi3_adj << "\n"
       << stdev1_adj << "\n" << cappi1_wt << "\n" << cappi2_wt << "\n"
       << cappi3_wt << "\n" << stdev1_wt;
} // end of Bright Band function call

// spherical search pattern
void spherical_search (int z, const int No_Data, int y, int x, int range1, int
range2, int lev, int nrows_radar, int ncols_radar, std::vector<
ublas::matrix<double> >const& radar_data, std::vector<Control_Data>&controls,
std::vector<Temp_Data>&temps, double upper)
{
  const int ROWS = nrows_radar-1;
  const int COLS = ncols_radar-1;
  int row1 = y-range1;
  int row2 = x-range1;
  int row3 = y+range1;
  int row4 = x+range1;
  int col1 = y-range2;
  int col2 = x-range1;
  int col3 = y+range2;
  int col4 = x+range1;

  // Edge Constraints
  if (row2<=0)      { row2=0; }
  if (row4>=COLS)   { row4=COLS; }
  if (col1<=0)      { col1=0; }
  if (col3>=ROWS)   { col3=ROWS; }

  // "range2==0" indicates first search on level
  if ((range2==0) && (radar_data[lev][y][x]!=No_Data)) {
    Temp_Data values;
    values.x_control_temp = x;
    values.y_control_temp = y;
    values.z_control_temp = lev;
    values.dBZ_control_temp = radar_data[lev][y][x];
    values.dist_to_target_temp = fabs(lev-z);
    temps.push_back(values);
  }
  // row searching algorithm: move incremental outwards row by row
  for (int c=row2; c<=row4; c++) {
    if (row1>=0) {
      if (radar_data[lev][row1][c]!=No_Data) {
        Temp_Data values;
        values.x_control_temp = c;
        values.y_control_temp = row1;
        values.z_control_temp = lev;
        values.dBZ_control_temp = radar_data[lev][row1][c];
        values.dist_to_target_temp = sqrt((pow((y-row1),2))+(pow((x-
        c),2))+(pow((lev-z),2)));
        temps.push_back(values);
      }
    }
    if (row3<ROWS) {
      if (radar_data[lev][row3][c]!=No_Data) {
        Temp_Data values;
        values.x_control_temp = c;
        values.y_control_temp = row3;
        values.z_control_temp = lev;
        values.dBZ_control_temp = radar_data[lev][row3][c];
        values.dist_to_target_temp = sqrt((pow((y-row3),2))+(pow((x-
        c),2))+(pow((lev-z),2)));
```

```
            temps.push_back(values);
          }
        }
      }
      // column searching algorithm: move incremental outwards column by column
      for (int r=col1; r<=col3; r++) {
        if (col2>=0) {
          if (radar_data[lev][r][col2]!=No_Data) {
            Temp_Data values;
            values.x_control_temp = col2;
            values.y_control_temp = r;
            values.z_control_temp = lev;
            values.dBZ_control_temp = radar_data[lev][r][col2];
            values.dist_to_target_temp = sqrt((pow((y-r),2))+(pow((x-
            col2),2))+(pow((lev-z),2)));
            temps.push_back(values);
          }
        }
        if (col4<COLS) {
          if (radar_data[lev][r][col4]!=No_Data) {
            Temp_Data values;
            values.x_control_temp = col4;
            values.y_control_temp = r;
            values.z_control_temp = lev;
            values.dBZ_control_temp = radar_data[lev][r][col4];
            values.dist_to_target_temp = sqrt((pow((y-r),2))+(pow((x-
            col4),2))+(pow((lev-z),2)));
            temps.push_back(values);
          }
        }
      }

      // moving all the values from "temp" to "control" vectors if point lie in
      // specified radii
      std::vector<int>point;
      for (int row=0; row<temps.size(); row++) {
        if((temps[row].z_control_temp=lev) &&
        (temps[row].dist_to_target_temp<upper)){
          Control_Data values;
          values.x_control = temps[row].x_control_temp;
          values.y_control = temps[row].y_control_temp;
          values.z_control = temps[row].z_control_temp;
          values.dBZ_control = temps[row].dBZ_control_temp;
          values.dist_to_target = temps[row].dist_to_target_temp;
          controls.push_back(values);
          point.push_back(row);
        }
      }

      // Erasing all the "temp" values that have been moved to "control" vectors
      int counter=0;
      for (int col=0; col<point.size(); col++) {
        temps.erase(temps.begin()+point[col]-counter);
        counter++;
      }

      point.clear();
    } // end of spherical search function call

    // Ordinary Kriging stratiform rainfall function call
    double Ordinary_Kriging_Strat (std::vector<Control_Data>const&controls, int y,
    int x, int z)
    {
      int row, col;
```

```
  int m = controls.size();
  float d;
  int n = m+1;
  float **a = matrix(1,n,1,n);
  float *b = vector(1,n);
  int *indx = ivector(1,n);

  // computing the coefficient matrix (G) control to controls
  for (row=0; row<m; row++) {
    for (col=0; col<m; col++) {
      double r = (pow((controls[row].y_control-
      controls[col].y_control),2)+pow((controls[row].x_control-
      controls[col].x_control),2));
      double h = fabs(controls[row].z_control-controls[col].z_control);
      a[row+1][col+1] = Cov_Strat_to_Strat(r, h);
    }
  }

  for (row=0; row<n; row++) {
    a[row+1][n] = 1;
    a[n][row+1] = 1;
  }
  a[n][n] = 0;

  // computing g(s) vector target to controls
  for (row=0; row<n; row++) {
    double r = (pow((y-controls[row].y_control),2)+pow((x-
    controls[row].x_control),2));
    double h = fabs(z-controls[row].z_control);
    b[row+1] = Cov_Strat_to_Strat(r, h);
  }
  b[n] = 1;

  // LU Decomposition
  ludcmp (a, n, indx, &d);
  lubksb (a, n, indx, b);

  // Vector of controls values
  Vector dBZ_Values (m+1);
  for (row=0; row<m+1; row++) {
    dBZ_Values[row]=controls[row].dBZ_control;
  }
  dBZ_Values[m]=0;

  // vector of weighting values "b" determined by LU decomposition copied to
  // Weights vector
  Vector Weights (n);
  for (row=0; row<n; row++) {
    Weights[row] = b[row+1];
  }

  // Computing the estimate of the final Kriged Value
  double Kriged_Value = (ublas::inner_prod(Weights, dBZ_Values));
  if (Kriged_Value<=18.0) { // dBZ threshold
    Kriged_Value = 0;
  }

  free_matrix(a,1,n,1,n);
  free_vector(b,1,n);
  free_ivector(indx,1,n);

  return Kriged_Value;
} // end of Oridinary Kriging Stratiform function call
```

```
// mixed rain: majority of controls are stratiform rainfall
double Universal_Kriging_Mixed_Strat (std::vector<Control_Data>const&controls,
int y, int x, int z, double r0, double h0, double H)
{
  int row,col;
  const double Convec_Threshold = 35.0;
  int n = controls.size();
  int GROWS = n+1+2;
  int GCOLS = n+1+2;
  Matrix GMATRIX (GROWS, GCOLS);

  // computing the coefficient matrix (G)
  for (row=0; row<n; row++) {
    for (col=0; col<n; col++) {
      double r = (pow((controls[row].y_control-
      controls[col].y_control),2)+pow((controls[row].x_control-
      controls[col].x_control),2));
      double h = fabs(controls[row].z_control-controls[col].z_control);
      GMATRIX[row][col] = Cov_Mixed(r, h, r0, h0, H);
    }
  }
  for (row=0; row<n; row++) {
    GMATRIX[row][n] = 1;
    GMATRIX[n][row] = 1;
  }
  for (row=n; row<(n+1+2); row++) {
    for (col=n; col<(n+1+2); col++) {
      GMATRIX[row][col] = 0;
    }
  }

  // convecitive & stratiform binary switch for convective and stratiform
  // pixels e.g. if convective [C S] = [1 0]; e.g. if stratiform [C S] = [0 1]
  for (row=0; row<n; row++) {
    if (controls[row].dBZ_control>=Convec_Threshold) {
      GMATRIX[n+1][row] = 1;
      GMATRIX[n+2][row] = 0;
      GMATRIX[row][n+1] = 1;
      GMATRIX[row][n+2] = 0;
    }
    else {
      GMATRIX[n+1][row] = 0;
      GMATRIX[n+2][row] = 1;
      GMATRIX[row][n+1] = 0;
      GMATRIX[row][n+2] = 1;
    }
  }

  // Inverting the coefficeint Matrix by Singular Value Decomposition (SVD)
  Matrix IGMATRIX = inverse_NRIC(GMATRIX);

  // computing g(s) vector, target to controls
  // majority pixels stratiform therefore [C S] = [0 1]
  Vector gvalues (n+1+2);
  for (row=0; row<n; row++) {
    double r = (pow((y-controls[row].y_control),2)+pow((x-
    controls[row].x_control),2));
    double h = fabs(z-controls[row].z_control);
    gvalues[row] = Cov_Mixed(r, h, r0, h0, H);
  }
  gvalues[n]=1;
  gvalues[n+1]=0;
  gvalues[n+2]=1;
```

```
  // vector of controls values
  Vector dBZ_Values (n+1+2);
  for (row=0; row<n; row++) {
    dBZ_Values[row]=controls[row].dBZ_control;
  }
  dBZ_Values[n]=0;
  dBZ_Values[n+1]=0;
  dBZ_Values[n+2]=0;

  // Computing the estimate of the final Kriged value
  double Kriged_Value = (ublas::inner_prod((ublas::prod(IGMATRIX, gvalues)),
  dBZ_Values));
  if (Kriged_Value<=18.0) { // dBZ threshold
    Kriged_Value = 0;
  }

  return Kriged_Value;
} // end of Universal_Kriging_Mixed_Strat function call

// mixed rain, majority of controls are convective rainfall
double Universal_Kriging_Mixed_Convec (std::vector<Control_Data>const&controls,
int y, int x, int z, double r0, double h0, double H)
{
  int row, col;
  const double Convec_Threshold = 35.0;
  int n = controls.size();
  int GROWS = n+1+2;
  int GCOLS = n+1+2;
  Matrix GMATRIX (GROWS, GCOLS);

  // computing the coefficient matrix (G)
  for (row=0; row<n; row++) {
    for (col=0; col<n; col++) {
      double r = (pow((controls[row].y_control-
      controls[col].y_control),2)+pow((controls[row].x_control-
      controls[col].x_control),2));
      double h = fabs(controls[row].z_control-controls[col].z_control);
      GMATRIX[row][col] = Cov_Mixed(r, h, r0, h0, H);
    }
  }
  for (row=0; row<n; row++) {
    GMATRIX[row][n] = 1;
    GMATRIX[n][row] = 1;
  }
  for (row=n; row<(n+1+2); row++) {
    for (col=n; col<(n+1+2); col++) {
      GMATRIX[row][col] = 0;
    }
  }

  // convecitive & stratiform binary switch for convective and stratiform
  // pixels e.g. if convective [C S] = [1 0]; e.g. if stratiform [C S] = [0 1]
  for (row=0; row<n; row++) {
    if (controls[row].dBZ_control>=Convec_Threshold) {
      GMATRIX[n+1][row] = 1;
      GMATRIX[n+2][row] = 0;
      GMATRIX[row][n+1] = 1;
      GMATRIX[row][n+2] = 0;
    }
    else {
      GMATRIX[n+1][row] = 0;
      GMATRIX[n+2][row] = 1;
      GMATRIX[row][n+1] = 0;
      GMATRIX[row][n+2] = 1;
```

```
    }
  }

  // inverting the coefficient matrix by Singular Value Decomposition (SVD)
  Matrix IGMATRIX = inverse_NRIC(GMATRIX);

  // computing g(s) vector, target to controls
  // majority pixels convective therefore [C S] = [1 0]
  Vector gvalues (n+1+2);
  for (row=0; row<n; row++) {
    double r = (pow((y-controls[row].y_control),2)+pow((x-
    controls[row].x_control),2));
    double h = fabs(z-controls[row].z_control);
    gvalues[row] = Cov_Mixed(r, h, r0, h0, H);
  }
  gvalues[n]=1;
  gvalues[n+1]=1;
  gvalues[n+2]=0;

  // vector of control values
  Vector dBZ_Values (n+1+2);
  for (row=0; row<n; row++) {
    dBZ_Values[row]=controls[row].dBZ_control;
  }
  dBZ_Values[n]=0;
  dBZ_Values[n+1]=0;
  dBZ_Values[n+2]=0;

  // Computing the estimate of the final Kriged value
  double Kriged_Value = (ublas::inner_prod((ublas::prod(IGMATRIX, gvalues)),
  dBZ_Values));
  if (Kriged_Value<=18.0) { // dBZ threshold
    Kriged_Value = 0;
  }

  return Kriged_Value;
} // end of function call for Universal_Kriging_Mixed_Convec

// Ordinary Kriging convective rainfall function call
double Ordinary_Kriging_Convec (std::vector<Control_Data>const&controls, int y,
int x, int z)
{
  int row,col;
  int n = controls.size();
  int GROWS = n+1;
  int GCOLS = n+1;
  Matrix GMATRIX (GROWS, GCOLS);

  // computing the coefficient matrix (G)
  for (row=0; row<n; row++) {
    for (col=0; col<n; col++) {
      double r = (pow((controls[row].y_control-
      controls[col].y_control),2)+pow((controls[row].x_control-
      controls[col].x_control),2));
      double h = fabs(controls[row].z_control-controls[col].z_control);
      GMATRIX[row][col] = Cov_Convec_to_Convec(r, h);
    }
  }
  for (row=0; row<n; row++) {
    GMATRIX[row][n] = 1;
    GMATRIX[n][row] = 1;
  }
  GMATRIX[n][n] = 0;
```

```
  // inverting the covariance matrix
  Matrix IGMATRIX = inverse_NRIC(GMATRIX);

  // computing g(s) vector, target to controls
  Vector gvalues (n+1);
  for (row=0; row<n; row++) {
    double r = (pow((y-controls[row].y_control),2)+pow((x-
    controls[row].x_control),2));
    double h = fabs(z-controls[row].z_control);
    gvalues[row] = Cov_Convec_to_Convec(r, h);
  }
gvalues[n]=1;

  // vector of controls values
  Vector dBZ_Values (n+1);
  for (row=0; row<n; row++) {
    dBZ_Values[row]=controls[row].dBZ_control;
  }
  dBZ_Values[n]=0;

  // computing the estimate of the final Kriged value
  double Kriged_Value = (ublas::inner_prod((ublas::prod(IGMATRIX, gvalues)),
  dBZ_Values));
  if (Kriged_Value<=18.0) { // dBZ threshold
    Kriged_Value = 0;
  }

  return Kriged_Value;
}

// 3D Semivariogram Function call for Convective pixels. Centroid values from
sensitivity analysis (Wesson & Pegram, 2005, NHESS paper)
inline double Cov_Convec_to_Convec (double r, double h)
{
  const double r0 = 3.38;
  const double h0 = 4.11;
  const double H = 1.78;
  double dist = sqrt(((r)/(r0*r0))+(pow((h/h0),2)));
  double g = 1-exp(-(pow((dist),H)));
  return g;
}

// 3D Semivariogram Function call for Stratiform pixels. Centroid values from
sensitivity analysis (Wesson & Pegram, 2005, NHESS paper)
inline double Cov_Strat_to_Strat (double r, double h)
{
  const double r0 = 8.40;
  const double h0 = 2.56;
  const double H = 1.27;
  double dist = sqrt(((r)/(r0*r0))+(pow((h/h0),2)));
  double g = 1-exp(-(pow((dist),H)));
  return g;
}

// 3D Semivariogram Function call for mixed Convective to Stratiform pixels
// Centroid values from sensitivity analysis (Wesson & Pegram, 2005, NHESS
// paper)
inline double Cov_Mixed (double r, double h, double r0, double h0, double H)
{
  double dist = sqrt(((r)/(r0*r0))+(pow((h/h0),2)));
  double g = 1-exp(-(pow((dist),H)));
  return g;
}
```

```
// Write a data array to an mdv file with path specified by file_name. Assumes
// the field is square at each level.
void write_mdv(char const* file_name, std::vector<dbl_mat> const& radar_data, int
nrows_radar)
{
  int ny = nrows_radar;
  int nx = ny;
  int nz = radar_data.size();
  MDV_handle_t mdv;
  MDV_init_handle(&mdv);

  // set up master header
  MDV_init_master_header(&mdv.master_hdr);
  mdv.master_hdr.data_dimension = 3;
  mdv.master_hdr.data_collection_type = MDV_DATA_MIXED;
  mdv.master_hdr.native_vlevel_type = MDV_VERT_TYPE_SURFACE;
  mdv.master_hdr.vlevel_type = MDV_VERT_TYPE_SURFACE;
  mdv.master_hdr.vlevel_included = true;
  mdv.master_hdr.grid_order_direction = MDV_ORIENT_NS_WE;
  mdv.master_hdr.grid_order_indices = MDV_ORDER_XYZ;
  mdv.master_hdr.n_fields = 1;
  mdv.master_hdr.max_nx = nx;
  mdv.master_hdr.max_ny = ny;
  mdv.master_hdr.max_nz = nz;
  mdv.master_hdr.n_chunks = 0;
  mdv.master_hdr.sensor_lon = (float)28.16;
  mdv.master_hdr.sensor_lat = (float)-28.10;
  mdv.master_hdr.sensor_alt = (float)1.719;
  mdv.master_hdr.time_centroid = 0;

  sprintf(mdv.master_hdr.data_set_info, "Radar volume scan data with all
  missing data infilled");
  sprintf(mdv.master_hdr.data_set_name, "MRL5 infilled data");
  sprintf(mdv.master_hdr.data_set_source, "Infilled by Cascade Kriging
  Algorithm");
  MDV_alloc_handle_arrays(&mdv, mdv.master_hdr.n_fields, mdv.master_hdr.max_nz,
  mdv.master_hdr.n_chunks);

  // set up field headers
  int i;
  for (i = 0; i < mdv.master_hdr.n_fields; ++i) {
    MDV_field_header_t *fhdr = mdv.fld_hdrs + i;
    MDV_init_field_header(fhdr);
    fhdr->nx = nx;
    fhdr->ny = ny;
    fhdr->nz = nz;
    fhdr->proj_type = MDV_PROJ_FLAT;
    // fhdr->encoding_type = MDV_PLANE_RLE8;
    fhdr->encoding_type = MDV_INT8;
    fhdr->bad_data_value = 255;
    fhdr->missing_data_value = 255;
    fhdr->data_element_nbytes = 1;
    fhdr->field_data_offset = 0;
    fhdr->volume_size = fhdr->nx*fhdr->ny*fhdr->nz*sizeof(ui08);
    fhdr->proj_origin_lon = (float)28.16;
    fhdr->proj_origin_lat = (float)-28.10;
    fhdr->grid_dy = 1;
    fhdr->grid_dx = 1;
    fhdr->grid_dz = 1;
    fhdr->grid_minx = -nx/2 + 0.5;
    fhdr->grid_miny = -ny/2 + 0.5;
    fhdr->grid_minz = mdv.master_hdr.sensor_alt + 1;
    fhdr->proj_rotation = 0;
```

```
    sprintf(fhdr->field_name_long, "MRL5 infilled data");
    sprintf(fhdr->field_name, "MRL5 data");
    sprintf(fhdr->units, "dBZ");

    MDV_vlevel_header_t *vhdr = mdv.vlv_hdrs + i;
    MDV_init_vlevel_header(vhdr);

    vhdr->vlevel_type[0] = MDV_VERT_TYPE_Z;
    vhdr->vlevel_params[0] = fhdr->grid_minz;
  }

  // convert to character array
  ui08 ***c_array = new ui08 **[mdv.master_hdr.n_fields];
  int ifield;
  for (ifield = 0; ifield < mdv.master_hdr.n_fields; ++ifield) {
    MDV_field_header_t *fhdr = mdv.fld_hdrs + ifield;
    c_array[ifield] = new ui08 *[fhdr->nz];
    int iz;
    for (iz = 0; iz < fhdr->nz; ++iz) {
      c_array[ifield][iz] = new ui08[fhdr->nx*fhdr->ny];
    }
  }

  convert_to_byte(mdv, radar_data, c_array);
  MDV_write_all(&mdv, const_cast<char*>(file_name), MDV_PLANE_RLE8);

  // must free c_array _before_ calling MDV_free_handle !!
  for(ifield = 0; ifield < mdv.master_hdr.n_fields; ++ifield){
    int iz;
    for(iz = 0; iz < mdv.fld_hdrs[ifield].nz; ++iz) {
      delete [] c_array[ifield][iz];
    }
    delete [] c_array[ifield];
  }
  delete [] c_array;
  MDV_free_handle(&mdv);
}

// Convert the floating point values in the pixel_vec to values which
// fit into a single byte.
void convert_to_byte(MDV_handle_t &mdv, std::vector<dbl_mat> const&
radar_data, ui08 *** c_array){
  double bad_data = 255;
  int iz;

  // determine the min and max, initialize with extreme values
  for ( int ifield = 0; ifield < mdv.master_hdr.n_fields; ++ifield ) {
    float min = 1000000;
    float max = -1000000;

    for ( iz = 0; iz < mdv.fld_hdrs[ifield].nz; ++iz ){
      for ( int iy = 0; iy < mdv.fld_hdrs[ifield].ny; ++iy ){
        for ( int ix = 0; ix < mdv.fld_hdrs[ifield].nx; ++ix ){
          if ( radar_data[iz][iy][ix] != bad_data ){
            if ( min > radar_data[iz][iy][ix] ){
              min = radar_data[iz][iy][ix];
            }
            else if ( max < radar_data[iz][iy][ix] ){
              max = radar_data[iz][iy][ix];
            }
          }
        }
      }
    }
```

```
    }

    float range = max - min;
    mdv.fld_hdrs[ifield].scale = (float)(range/250.0); // 250 bins
    mdv.fld_hdrs[ifield].bias = (float)(min);
    for(iz = 0; iz < mdv.fld_hdrs[ifield].nz; ++iz){
      for(int ixy = 0; ixy < mdv.fld_hdrs[ifield].ny *
      mdv.fld_hdrs[ifield].nx; ++ixy){
        double row = floor(ixy/mdv.fld_hdrs[ifield].ny);
        double col = ixy - row*mdv.fld_hdrs[ifield].ny;

        if ( radar_data[iz][row][col] == bad_data ){
          c_array[ifield][iz][ixy] = (ui08)(
          mdv.fld_hdrs[ifield].missing_data_value );
        }
        else{
          c_array[ifield][iz][ixy] = (ui08)((radar_data[iz][row][col] -
          mdv.fld_hdrs[ifield].bias)/mdv.fld_hdrs[ifield].scale + 0.5);
        }
      }
      mdv.field_plane[ifield][iz] = c_array[ifield][iz];
    }
  }
}
```

REFERENCES

Aitken, A.C. (1967). *Determinants and Matrices*. Oliver and Boyd Ltd. London.

Austin, G.L. (2001). History of Radar Meteorology. *Radar hydrology for real time flood forecasting: Proceedings of an advanced study course,* Griffith, R.J., Cluckie, I.D., Austin, G.L. and Han, D, eds., European Commission. Luxembourg, Belgium. 3-13.

Battan, L.J. (1973). *Radar Observation of the Atmosphere*. University of Chicago Press. Chicago.

Bell, T.L. (1987). A Space-Time Stochastic Model of Rainfall for Satellite Remote Sensing Studies, *J. Geophys. Res.,* **Vol. 92**, No. D8, P9631.

Berenguer, M., Sempere-Torres, D., Sanchez-Diezma, R. and Pascual, R (2005). Identification of clutter echoes using a fuzzy logic technique. *32nd Conference on Radar Meteorology*. Albuquerque, New Mexico.

Bras, R.L. and Rodriguez-Iturbe, I. (1985). *Random Functions and Hydrology*. Addison-Wesley Publishing Company. Ontario.

Chiles, J.P. and Delfiner, P. (1999). *Geostatistics – Modelling Spatial Uncertainty*. Wiley-Intersicence Publication.

Ciach, G. J. (2002). Local Random Errors in Tipping-Bucket Rain Gauge Measurements. *Journal of Atmospheric and Oceanic Technology. Notes and Correspondence*, **20**, 752-759.

Collier, C.G. (1986). Accuracy of Rainfall Estimates by Radar, Part 1: Calibration by Telemetering Raingauges. *Journal of Hydrology*, **83**, 207-223.

Crespi, M., Monai, M., Pesci, A.D. and Trolese, A. (1994). Improvements in Clutter Reduction and Precipitation Intensity Measurements for Monte Grande Radar. *COST 75 Weather Radar Systems International Seminar.* Brussels, Belgium. 381-392.

Cressie, N. (1993). *Statistics for Spatial Data.* Wiley-Interscience Publication. New York.

Cressie, N. and Hawkins, D.M. (1980). Robust Estimation of the Variogram, 1. *Journal of the International Association for Mathematical Geology*, **12**, 115-125.

Creutin, J.D. and Obled, C. (1982). Objective Analysis and Mapping Techniques for Rainfall Fields: An Objective Comparison. *Water Resources Research*, **(18) 2**, 413-431.

Deyzel, I.T.H. (2003). *Spatial Interpolation and Mapping of Rainfall Simar: Radar and satellite products*, Draft final report (K5/1152) to the Water Research Commission for the period March 2000 to March 2003. Water Research Commission. Pretoria. 65-67.

Dixon, W.J. and Massey, F.J. (1983). *Introduction to Statistical Analysis, 4th Edition.* McGraw-Hill Book Company. Singapore.

Ehret, U. (2002). *Rainfall and Flood Nowcasting in Small Catchments using Weather Radar.* PhD Thesis. University of Stuttgart.

Franco, M., Sempere-Torres, D., Sanchez-Diezma, R. and Andrieu, H. (2002). A methodology to identify the vertical profile of reflectivity from radar scans and to estimate the rainrate at ground at different distances. *Proceedings of ERAD (2002).* 299-304.

Gordon, A.D. (1981). *Classification: Methods for the exploratory analysis of multivariate data.* Chapman and Hall. New York, USA.

Habib, E. and Krawjewski, W.F. (2002). Uncertainty Analysis of the TRMM Ground-Validation Radar-Rainfall Products: Application to the TEFLUN-B Field Campaign. *Journal of Applied Meteorology*, **41**, 558-572.

Hengl, T., Geuvelink, G.B.M. and Stein, A. (2003). Comparison of Kriging with external drift and regression-kriging. *Technical note, ITC*, Available online at http://www.itc.nl/library/Academic_output/

Interpreting Weather Radar Information (2005). Retrieved 18 February 2005, from the Innovation and Research Division of the South African Weather Service website: (http://metsys.weathersa.co.za/RadarInfo.htm)

Jordan, P., Seed, A. and Austin, G. (2000). Sampling errors in radar estimates of rainfall. *Journal of Geophysical Research*, **Vol. 105**, No. D2, 2247-2257.

Joss, J. and Lee, R. (1995). The Application of Radar-Gauge Comparisons to Operational Precipitation Profile Corrections. *Journal of Applied Meteorology*. 2612-2630.

Journel, A.G. and Huijbregts, Ch.J. (1978). *Mining Geostatistics*. Academic Press. London, UK.

Lack, S.A. and Fox, N.I. (2004). Errors in Surface Rainfall Rates Retrieved from Radar due to Wind Drift. *Sixth International Symposium on Hydrological Applications of Weather Radar.* Melbourne, Australia.

Lebel, T., Bastin, G., Obled, C. and Creutin, J.D. (1987). On the Accuracy of Areal Rainfall Estimation: A Case Study. *Water Resources Research*, **Vol. 23**, No. 11, 2123-2134.

Marshall, J.S. and Palmer, W.M. (1948). The Distribution of Raindrops with Size. *Journal of Meteorology*, **5**, 65-166.

Mather, G.K., Terblanche, D.E. and Steffens, F.E. (1997). *National Precipitation Research Programme: Final Report for the Period 1993 – 1996.* WRC report No 726/1/97.

Matheron, G. (1962). Traite de Geostatistique Appliquee, Tome I. *Memories du Bureau de Recherches Geologiques et Minieres*, **No. 14**. Editions Technip, Paris.

Mazzetti, C. (2004). *Report on Performance of the New Methodology: The Application of the Block Kriging and Bayesian combination technique*. MUSIC Deliverable 7.4, Contract No. EVK1-CT-2000-00058.

Mittermaier, M.P. (1999). *Investigating the Characteristics of the Radar Reflectivity Profile.* MSc Eng dissertation, Civil Engineering, University of KwaZulu-Natal, Durban, Howard College.

Mittermaier, M. P. (2003). *Investigating synergies between weather radar data and mesoscale model output.* PhD Thesis, Department of Meteorology, University of Reading.

Mittermaier, M. P. and Terblanche, D.E. (1997). Converting Weather Radar Data into Cartesian Space: A New Approach Using DISPLACE Averaging. *Water SA*, **Vol. 23**, No. 1, 46-50.

Moolman, W.H. (1985). A *Study for Tests for First Order Stationarity.* PhD thesis, University of Durban-Westville, Durban, South Africa.

Pegram, G.G.S. (2000). *Testing for Variability and Persistence in Time Serious.* Detecting trend and other changes in hydrological data, WMO-TD No. 1013. Kundzewicz, Z.W. and Robson, A. eds. Geneva, 107-112.

Pegram, G.G.S. and Clothier, A.N. (2001). High Resolution Space-Time Modelling of Rainfall: The "String of Beads" model. *J. Hydrology.* **V 241**. 26-41.

Pegram, G.G.S., Sinclair, D.S. and Wesson, S.M. (2005). *Daily Rainfall Mapping over South Africa: Modelling.* Final report (K5/1425) to the Water Research Commission. August 2005. Water Research Commission. Pretoria.

Pegram, G.G.S and Terblanche, D.E. (2001). Radar Hydrology Developments in South Africa. *Radar hydrology for real time flood forecasting: Proceedings of an advanced study course,* Griffith, R.J., Cluckie, I.D., Austin, G.L. and Han, D, eds., European Commission. Luxembourg, Belgium. 245-262.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). *Numerical Recipes in C – The Art of Scientific Computing.* 2nd Edition. Cambridge University Press. Cambridge.

Riedl, J. (1994). Examples of Improvements to Clutter Suppression in Current Operational Weather Radar Systems. *COST 75 Weather Radar Systems International Seminar.* Brussels, Belgium. 114-124

Sabyasachi, B., Gunst, R.F., Guertal, E.A. and Hartfield, M.I. (1997). The Effects of Influential Observations on Sample Semivariograms. *Journal of agriculture, biological, and environmental statistics*, **2**, 490-512.

Sanchez-Diezma, R., Sempere-Torres, D., Delrieu, G. and Zawadzki, I. (2001). An improved methodology for ground clutter substitution based on a pre-classification of precipitation types. *30th International Conference on Radar Meteorology*. Munich, Germany. 271-273.

Sanchez-Diezma, R., Zawadzki, I. and Sempere-Torres, D. (2000). Identification of the bright band through the analysis of volumetric radar data. *Journal of Geophysical Research,* **Vol. 105**, N0. D2, 2225-2236.

Sempere-Torres, D., Sanchez-Diezma, R., Zawadzki, I. and Creutin, J.D. (2000). Identification of Stratiform and Convective Areas Using Radar Data with Application to the Improvement of DSD Analysis and Z-R Relationships. *Phys. Chem. Earth*, **25**, 985-990.

Seed, A.W. (1992). Generation of a Spatially Distributed Daily Rainfall Database for Various Weather Modification Scenarios. WRC 373/1/92. Report to Water Research Commission. Pretoria, South Africa.

Seed, A.W. and Pegram, G.G.S. (2001). Using Kriging to Infill Gaps in Radar Data due to Ground Clutter in Real Time. *Proceedings Fifth International Symposium on Hydrological Applications of Weather Radar – Radar Hydrology*. Kyoto, Japan. 73-78.

Sinclair, D.S. (2001). *A Linear Catchment Model for Real Time Flood Forecasting*. MSc Eng dissertation, Civil Engineering, University of KwaZulu-Natal, Durban, Howard College.

Sinclair, D.S. and Pegram, G.G.S. (2003). Combining traditional and remote sensing techniques of rainfall measurement as a tool for Hydrology, Agriculture and Water Resource Management. *Proceedings 11th SA National Hydrology Symposium*, Port Elizabeth, South Africa.

Sonka, M. Hlavac, V. and Boyle, R. (1999). *Image Processing, Analysis, and Machine Vision*. 2nd Edition. PWS Publishing. Pacific Grove.

Steiner, M., Houze, R.A. and Yuter, S.E. (1978). Climatological Characterization of Three-Dimensional Storm Structure from Operational Radar and Rain Gauge Data. *Journal of Applied Meteorology*, **34**, 1978-2007.

Sugier, J., Parent du Chatelet, J., Roquain, P. and Smith, A. (2002). Detection and removal of clutter and anaprop in radar data using a statistical scheme based on echo fluctuation. *Proceedings of ERAD 2002.* 17-24.

Terblanche, D.E., Pegram, G.G.S. and Mittermaier, M.P. (2001). The development of weather radar as a research and operational tool for hydrology in South Africa. *Journal of Hydrology*, **241**, 3-25.

Todini, E. (2001). A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrology and Earth System Sciences*, **5(2)**, 187-199.

Tyson, P.D. and Preston-Whyte, R.A. (2000). *The Weather and Climate of Southern Africa, 2nd Edition*. Oxford University Press. Cape Town, South Africa.

Vezzani, G. (1994). Ground Clutter Suppression in Atmospheric Phenomena Observation. *COST 75 Weather Radar Systems International Seminar.* Brussels, Belgium. 281-295.

Visser P. (2003). *Spatial interpolation and Mapping of rainfall: 2. Radar and Satellite products*, Draft final report (K5/1152) to the Water Research Commission for the period March 2000 to March 2003. Water Research Commission. Pretoria. 10-19.

Wesson, S.M. and Pegram, G.G.S. (2004). Radar Rainfall Image Repair Techniques. *Hydrology and Earth System Sciences*, **8(2)**, 220-334.

Wilkinson, J.H. (1988). *The Algebraic Eigenvalue Problem*. Oxford Science Publication.

Wilson, J.W. and Brandes, E.A. (1979). Radar Measurement of Rainfall – A Summary. *American Meteorological Society*. **60 (9)**, 1048-1058.